

Strategien beim Lösen von rezeptiven Sprachtestaufgaben

Malgorzata Barras

Heimatort: Broc (FR)

2021

Dissertation zur Erlangung der Doktorwürde an der Philosophischen Fakultät
der Universität Freiburg (Schweiz)

Genehmigt von der Philosophischen Fakultät auf Antrag von
Prof. Thomas Studer (1. Gutachter), Prof. Karen Schramm (2. Gutachterin) und
Prof. Katrin Wisniewski (3. Gutachterin)

Freiburg, den 19.11.2021

Prof. Dominik Schoebi, Dekan

[T]he purpose of qualitative research is not to prove anything, but to generate ideas which are sufficient to make us think again about what is going on in the world.
(Holliday, 2015, S. 53)

Danksagung

Bei der vorliegenden Arbeit handelt es sich um ein Dissertationsprojekt, das mit dem Projekt „Innovative Formen der Beurteilung“ (IFB) des schweizerischen Kompetenzzentrums für Mehrsprachigkeit (KFM) assoziiert war. Bei der Planung, Durchführung, Analyse und Darstellung der vorliegenden Resultate wurde ich von vielen Menschen und Institutionen unterstützt, denen ich an dieser Stelle meinen aufrichtigen Dank aussprechen möchte.

Vor allem danke ich den 30 Schüler/innen, die an der Studie beteiligt waren, für ihre Zeit, Geduld und den motivierten Einsatz bei der Datenerhebung. Auch ihren Lehrpersonen und Schulverantwortlichen, die bei der Organisation der Datenerhebung mitwirkten, sowie ihren Eltern, die der Teilnahme ihrer Kinder am Projekt zustimmten, bin ich sehr dankbar.

Meinem Betreuer Prof. Thomas Studer danke ich für sein Vertrauen und seine stetige Unterstützung bei der Arbeit am vorliegenden Projekt. Die vielen anregenden Diskussionen und seine wohlwollend-kritischen Anmerkungen zu den ganz grossen und ganz kleinen Aspekten meiner Arbeit zwangen mich jeweils dazu, mir die einzelnen Schritte noch genauer zu überlegen und diese noch fundierter zu begründen.

Einen ganz besonderen Dank möchte ich dem KFM und vor allem dem IFB-Kernteam aussprechen, zu dem Peter Lenz, Katharina Karges und Thomas Aeppli zählen. Dem Projektleiter Peter Lenz bin ich aus unzähligen Gründen zu Dank verpflichtet, vor allem aber dafür, dass ich seit 2012 meine ersten Schritte als Forscherin unter seiner Leitung machen durfte und von seiner kritischen Lektüre meiner Texte in den vergangenen Jahren enorm profitieren konnte. Dankbar bin ich ihm auch dafür, dass er mir das spannende Feld der Testforschung näherbrachte und es mir ermöglichte, das vorliegende Dissertationsprojekt an das Projekt IFB anzuschliessen. Ebenfalls lang und bestimmt nicht vollständig ist die Liste der Sachen, die ich Katharina Karges zu verdanken habe: Sie half mir u.a. bei der Programmierung der IFB-Aufgaben für die vorliegende Untersuchung, versorgte mich mit wichtigen Informationen zum quantitativen Teil der IFB-Studie, half mir bei der tabellarischen und grafischen Darstellung der Vergleiche meiner Stichprobe mit der IFB-Stichprobe, las den gesamten Dissertationstext kritisch durch und unterstützte mich moralisch während unzähliger Kaffeepausen, die wir live und online zusammen verbrachten. Thomas Aeppli bin ich wiederum für seine organisatorischen Künste dankbar: Dank seiner perfekten Planung der IFB-Datenerhebung wusste ich immer, wohin die Reise geht und wer in welchem Zimmer auf mich wartet. Dem KFM danke ich für die Finanzierung der Transkriptionen von Verbalprotokollen, die im Rahmen der vorliegenden Studie erhoben wurden. Lisa Schenk, Désirée Scheidegger und Niki Ott bin ich für die vielen Stunden zu Dank verpflichtet, die sie in die Anfertigung dieser Transkriptionen investierten.

In den vergangenen Jahren konnte ich stets auf die grosse fachliche und persönliche Unterstützung seitens meiner Kolleg/innen vom KFM und von der Universität Freiburg zählen. Ein besonderer Dank geht dabei an Dr. Jan Vanhove, dem ich die grafische Darstellung der codierten Strategien verdanke, an Nina Hicks, die die Arbeit Korrektur las, und an Dr. Moritz Sommet, der mich bei der Dokumentation des Dissertationsprojekts unterstützte.

Viel Unterstützung konnte ich auch während zahlreicher Doktorandenkolloquien erfahren. Insbesondere möchte ich mich dabei bei Prof. Karen Schramm bedanken, die mein Projekt von Anfang an mit hilfreichen Inputs begleitete und mich auch davon überzeugte, neben dem *Stimulated Recall* auch das Laute Denken als Datenerhebungsmethode in der vorliegenden Studie einzusetzen.

Meiner Familie und meinen Freunden bin ich letztlich für die ununterbrochene Unterstützung, Geduld und erwünschte Ablenkung dankbar, auf die ich mich in den letzten Jahren immer verlassen konnte. Ein ganz persönlicher Dank geht schliesslich an meinen Mann Frédéric und an unseren Sohn Gabriel, der bei der Endphase dieses Projekts für einen ganz besonderen Motivationskick (teilweise wörtlich) sorgte.

Bern, im November 2021

ABSTRACT

Um die Validität der Nutzung von (Sprach-)Tests sicherzustellen ist es heutzutage wichtig, neben der quantitativen Evidenz zum psychometrischen Funktionieren eines Tests auch die Antwortprozesse (*response processes*) der Testteilnehmenden, über deren Kompetenz man mithilfe der Testergebnisse Aussagen machen will, zu untersuchen (Ercikan & Pellegrino, 2017a). In der Testforschung wird dafür im Rahmen der Erprobung von neu erstellten Testinstrumenten oft auf introspektive Datenerhebungsverfahren wie Lautes Denken oder *Stimulated Recall* zurückgegriffen (vgl. z.B. Arras, 2013; Lazaraton & Taylor, 2007). Diese ermöglichen wertvolle Einblicke in die kognitiven Prozesse und Strategien, die beim Aufgabenlösen eingesetzt werden und die bei der statistischen Auswertung von Testergebnissen normalerweise unsichtbar sind (Lumley & Brown, 2005).

Im vorliegenden Dissertationsprojekt wurden die introspektiven Verfahren angewandt, um zu beleuchten, 1) welche Strategien deutschsprachige Französischlernende auf der Sekundarstufe I anwenden, wenn sie computerbasierte Testaufgaben zum Lese- und Hörverstehen lösen, 2) ob sich Zusammenhänge zwischen den konstruktirrelevanten *test-wiseness*-Strategien (vgl. A. D. Cohen, 2014) und dem erfolgreichen bzw. nicht erfolgreichen Lösen der Testaufgaben erkennen lassen, sowie 3) ob die Einbettung der Testaufgaben in ein Szenario einen Einfluss auf den Einsatz von Strategien beim Lösen der Testaufgaben hatte. Die in der Studie eingesetzten Testaufgaben wurden im Rahmen des Schweizer Forschungsprojekts „Innovative Formen der Beurteilung“ (vgl. Karges, Lenz et al., 2021; KFM, 2020a) erstellt und in ein Szenario eingebettet, dies u.a. um die Lernenden stärker zu motivieren und einen authentischeren Lösungsprozess anzustossen (vgl. Sabatini et al., 2014b).

Die Einblicke in die Gedanken und Lösungswege der Schüler/innen wurden in zweistündigen individuellen Sitzungen mit 30 Schüler/innen mittels 1) Lautem Denken und 2) *Stimulated-Recall*-Interviews (inkl. mündlicher Zusammenfassungen der rezipierten Texte) gewonnen. Um ausserdem zu untersuchen, ob und welchen Effekt die Szenarioeinbettung der Testaufgaben auf die Lösungsprozesse der Schüler/innen hatte, wurde ca. die Hälfte der Lernenden mit szenariobasierten Aufgaben zu einem übergreifenden Thema konfrontiert, während die restlichen Lernenden ähnliche Aufgaben zu verschiedenen Szenarien ohne thematischen Zusammenhang lösten. Durch die Verwendung von Lautem Denken und *Stimulated-Recall*-Interviews sollten im Sinne der Methodentriangulation die Schwächen der einen Methode möglichst durch Stärken der anderen Methode kompensiert und gleichzeitig auch die gewonnenen Erkenntnisse vertieft und erweitert werden (vgl. Knorr & Schramm, 2016). Alle Sitzungen wurden mithilfe der Software OBS (*OBS Studio*, 2017) und der

Bildschirmkamera aufgenommen, transkribiert (GAT 2-Basistranskript) und nach den Prinzipien der inhaltlich strukturierenden und evaluativen qualitativen Inhaltsanalyse ausgewertet (Kuckartz, 2018). Bei der vertiefenden Analyse der Strategien wurden die Verbalprotokolle von 20 Testteilnehmenden zu ausgewählten Testitems berücksichtigt.

Die Resultate zeigen, dass die Schüler/innen beim Lösen der Testaufgaben auf viele verschiedene Strategien zurückgriffen. Obwohl die meisten beobachteten Strategien konstruktrelevant sind, führten sie nicht immer zu erfolgreichen Aufgabenlösungen – vor allem dann nicht, wenn die Sprachkompetenzen der Lernenden nicht ausreichten, um die Inputtexte richtig verstehen und interpretieren zu können. Die konstruktirrelevanten *test-wiseness*-Strategien konnten 1) vor allem bei den schwächsten Französischlernenden beobachtet werden, die von diesen Strategien Gebrauch machten, wenn sie die Aufgaben aufgrund ihrer lückenhaften Französischkompetenzen nicht lösen konnten, aber auch 2) bei den stärkeren Schüler/innen, die die Testaufgaben schnell lösen wollten und den Textinhalt nur oberflächlich zur Kenntnis nahmen. Der Einsatz von *test-wiseness*-Strategien führte bei der Wahl der Antwortoptionen meistens zu einem Misserfolg, was aus Sicht der Testentwickler/innen als positiv zu interpretieren ist. Schliesslich zeigte sich auch, dass die Einbettung der Aufgaben in ein übergreifendes Szenario von den Schüler/innen der Szenario-Gruppe kaum bewusst wahrgenommen wurde und sich folglich wahrscheinlich kaum auf ihre Vorgehensweise beim Lösen der Testaufgaben auswirkte. Zudem hielten die Schüler/innen der Off-Szenario-Gruppe eine entsprechende Einbettung nicht für nötig. Der Vergleich der Strategien, die von Szenario- und Off-Szenario-Schüler/innen eingesetzt wurden, zeigte, dass es zwischen den beiden Schüler/innen-Gruppen zwar Unterschiede im Strategieneinsatz gab, aber offen bleibt, welche Ursachen dies konkret hatte. Da in der vorliegenden Studie nur einige wenige Szenario- und Off-Szenario-Aufgaben genauer untersucht werden konnten, müsste das Potenzial von szenariobasierten Testaufgaben in weiteren Studien vertieft untersucht werden. Die Erkenntnisse aus der vorliegenden Studie zeigen ausserdem, dass mithilfe von introspektiven Verfahren spannende Einblicke in die Denkprozesse der Sprachlernenden gewonnen werden können, die nicht nur für Testvalidierungszwecke, sondern auch allgemein für die Optimierung des Fremdsprachenunterrichts genutzt werden können. So zeigte sich u.a., dass einige Schüler/innen bisher kaum Erfahrung mit standardisierten Sprachtestformaten hatten; dass die Testaufgaben bei den Schüler/innen u.a. deswegen sehr gut ankamen, weil sie in ihren Augen gut erklärt, alltagsnah und in Bezug auf die gewählten Textsorten sowie Inhalte der Inputtexte attraktiv waren; oder dass der motivierende Effekt der Testaufgaben durch die Testsituation eingeschränkt war.

INHALTSVERZEICHNIS

Abstract	V
Inhaltsverzeichnis	VII
Abkürzungsverzeichnis	XI
Abbildungsverzeichnis	XII
Tabellenverzeichnis	XIII
1 Einleitung.....	1
1.1 Kontext.....	1
1.2 Problemstellung, Forschungsgegenstand und methodologischer Zugriff	4
1.3 Aufbau der Studie.....	9
2 Konzeptueller Rahmen und Literaturüberblick.....	11
2.1 Rezeptive Kompetenzen in der L2 überprüfen und bewerten	11
2.1.1 Leseverstehen	14
2.1.1.1 Leseverstehen als Prozess	14
2.1.1.2 Leseverstehen in der L1 und in der L2	18
2.1.1.3 Modell des Leseverstehens nach Khalifa & Weir (2009).....	19
2.1.1.4 Leseverstehen überprüfen und bewerten	21
2.1.2 Hörverstehen	24
2.1.2.1 Hörverstehen als Prozess	24
2.1.2.2 Hörverstehen in der L2.....	27
2.1.2.3 Hörverstehen überprüfen und bewerten	28
2.1.3 Neuere Entwicklungen	30
2.1.3.1 Digitale Textsorten	30
2.1.3.2 <i>Scenario-Based Assessment</i>	31
2.2 Einsatz und Erforschung von Strategien bei der Bewältigung von (rezeptiven) Sprachtestaufgaben	34
2.2.1 Die Rolle der strategischen Kompetenz in Modellen der kommunikativen Kompetenz	35
2.2.2 L2-Strategien: Definitions- und Klassifizierungsversuche.....	39
2.2.3 Lernerstrategien und Teststrategien	45
2.2.3.1 Lernerstrategien	46
2.2.3.2 Teststrategien	46
2.2.4 Erforschung von Strategien im Rahmen der Testerprobungsphase	47
2.2.4.1 Testerprobung und Testvalidierung	48
2.2.4.2 (Neuere) Studien zur Erforschung von Teststrategien mittels introspektiver Verfahren	52
2.2.5 Zusammenfassung	70
3 Vorgehen bei der Untersuchung der Strategien der Testteilnehmenden.....	73
3.1 Forschungskontext: Datenerhebung und Testaufgaben im Projekt IFB	73

3.2	Forschungsfragen und Forschungsdesign.....	78
3.3	Introspektive Verfahren als Datenerhebungsmethoden	80
3.4	Datenerhebung	84
3.4.1	Pilotstudie	84
3.4.2	Eingesetzte IFB-Testaufgaben.....	86
3.4.3	Untersuchungsteilnehmende	90
3.4.3.1	Auswahl der Teilnehmenden für die qualitative Untersuchung	91
3.4.3.2	Kompetenzen der Schüler/innen der qualitativen Stichprobe im Vergleich zur IFB-Stichprobe für Französisch	96
3.4.4	Erhebung der introspektiven Daten.....	104
3.4.4.1	Lautes Denken	105
3.4.4.2	<i>Stimulated-Recall</i> -Interviews	109
3.5	Datenaufbereitung.....	111
3.6	Datengrundlage für die Analyse	113
3.7	Datenauswertung	115
3.7.1	Die inhaltlich strukturierende und evaluative qualitative Inhaltsanalyse nach Kuckartz (2018) und ihr Einsatz in der vorliegenden Studie	117
3.7.2	Vorgehensweise bei der Datenauswertung.....	119
3.7.3	Das Kategoriensystem.....	129
3.7.3.1	Codierung der allgemeinen Informationen zu den Items.....	130
3.7.3.2	Definition und Codierung der Strategien.....	133
3.7.3.3	Codierung weiterer thematischer Kategorien.....	149
3.7.4	Berücksichtigung der Gütekriterien für die interne Studiengüte bei der qualitativen Inhaltsanalyse	151
3.8	Zusammenfassung.....	156
4	Ergebnisse.....	158
4.1	Auswertung der untersuchten Testaufgaben	159
4.2	Zu Forschungsfrage 1: Strategien der Schüler/innen beim Lösen der Testaufgaben	164
4.2.1	Lernerstrategien.....	166
4.2.2	Teststrategien	182
4.2.2.1	Testmanagementstrategien.....	183
4.2.2.2	<i>test-wiseness</i> -Strategien	208
4.2.3	Individuelle Unterschiede beim Strategieneinsatz	223
4.3	Zu Forschungsfrage 2: Zusammenhänge zwischen konstruktirrelevanten Strategien und erfolgreichen bzw. nicht erfolgreichen Aufgabenlösungen	229
4.3.1	<i>test-wiseness</i> -Strategien und (nicht) erfolgreiche Lösungen der Szenario-Aufgaben	230
4.3.2	<i>test-wiseness</i> -Strategien und (nicht) erfolgreiche Lösungen der Off-Szenario-Aufgaben	233

4.4	Zu Forschungsfrage 3: Strategien beim Lösen der szenariobasierten und nicht szenariobasierten Testaufgaben	237
4.4.1	Wahrnehmung der Szenario- und Off-Szenario-Testaufgaben durch die Schüler/innen.....	238
4.4.2	Lernerstrategien der Szenario- und Off-Szenario-Schüler/innen	242
4.4.3	Teststrategien der Szenario- und Off-Szenario-Schüler/innen	246
4.4.3.1	Testmanagementstrategien der Szenario- und Off-Szenario-Schüler/innen	246
4.4.3.2	<i>test-wiseness</i> -Strategien der Szenario- und Off-Szenario-Schüler/innen....	249
4.4.4	Strategien, die von Szenario- und Off-Szenario Schüler/innen bei der Bearbeitung von äquivalenten Tasks eingesetzt wurden.....	253
4.5	Weitere ausgewählte Ergebnisse	257
4.5.1	Einschätzung des Lauten Denkens durch die Schüler/innen.....	257
4.5.2	Einsatz von digitalen Textsorten in einem Sprachtest	259
4.5.3	Sprachtestaufgaben in den Augen der Schüler/innen	261
5	Diskussion.....	265
5.1	Zusammenfassung und Interpretation der Ergebnisse	265
5.1.1	Strategien der Schüler/innen beim Lösen der Testaufgaben.....	265
5.1.2	Zusammenhänge zwischen konstruktirrelevanten Strategien und (nicht) erfolgreichen Aufgabenlösungen.....	269
5.1.3	Unterschiede beim Einsatz der Strategien beim Lösen der Szenario- und Off-Szenario-Aufgaben	271
5.2	Limitationen der Studie und Generalisierbarkeit der Resultate	274
5.3	Relevanz der Studie	277
5.3.1	Empirische und theoretische Erkenntnisse	277
5.3.1.1	Szenariobasierte Testaufgaben für <i>large scale assessments</i> in Fremdsprachen	277
5.3.1.2	Qualitative Evidenz im Prozess der Testvalidierung	278
5.3.1.3	Strategische Kompetenz als Teil des Testkonstrukts.....	280
5.3.2	Methodologische Erkenntnisse	280
5.3.2.1	Lautes Denken	280
5.3.2.2	<i>Stimulated Recall</i>	285
5.3.2.3	Lautes Denken und <i>Stimulated Recall</i> als einander ergänzende Datenerhebungsmethoden	288
5.3.2.4	Die inhaltlich strukturierende und evaluative qualitative Inhaltsanalyse als Datenauswertungsmethoden.....	289
5.3.3	Mögliche Implikationen der Studie für den Fremdsprachenunterricht	290
6	Fazit und Ausblick	293
7	Literaturverzeichnis	296
8	Anhang	312
8.1	Anhang 1: Infobrief an die Erziehungsberechtigten	315

8.2	Anhang 2: Analyisierte Szenario-Aufgaben	317
8.3	Anhang 3: Analyisierte Off-Szenario-Aufgaben	329
8.4	Anhang 4: Transkriptionsrichtlinien.....	339
8.5	Anhang 5: Kategorienleitfaden	342
8.5.1	Codierung der allgemeinen, nicht strategiebezogenen Informationen	344
8.5.2	Codierung der Strategien beim Lauten Denken (LD).....	346
8.5.3	Codierung der Strategien beim <i>Stimulated Recall</i> (SR).....	363
8.5.4	Codierung der weiteren thematischen Kategorien.....	385

ABKÜRZUNGSVERZEICHNIS

ETS	<i>Educational Testing Service</i>
GeR	<i>Gemeinsamer europäischer Referenzrahmen für Sprachen</i> (Europarat, 2001)
IFB	Projekt „Innovative Formen der Beurteilung“
IFM	Institut für Mehrsprachigkeit in Freiburg (Schweiz)
IRT	<i>Item Response Theory</i>
KFM	Kompetenzzentrum für Mehrsprachigkeit in Freiburg (Schweiz)
LD	Lautes Denken
OFF/Off	Schüler/innen der Off-Szenario-Gruppe
S	Schüler/in
SBA	<i>Scenario Based Assessment</i>
SR	<i>Stimulated Recall</i>
SuS	Schüler/innen
SZ	Schüler/innen der Szenario-Gruppe
TN	Testteilnehmende
ÜGK	„Überprüfung der Erreichung der Grundkompetenzen“ (EDK, 2021)

ABBILDUNGSVERZEICHNIS

Abbildung 1: <i>Model of reading</i> von Khalifa & Weir 2009, S. 43.	20
Abbildung 2: IFB-Studie, Schulzimmer kurz vor der Datenerhebung.	77
Abbildung 3: Einbettung des Szenarios „Stadttrip“.	86
Abbildung 4: Szenario „Stadttrip“, Task A, erste Bildschirmseite.	87
Abbildung 5: Szenario „Stadttrip“, Task B, erste Bildschirmseite.	88
Abbildung 6: Szenario „Stadttrip“, Task C, erste Bildschirmseite.	88
Abbildung 7: Szenario „Stadttrip“, Task D, erste Bildschirmseite.	89
Abbildung 8: „Herkömmliches“ Leseverstehen.	97
Abbildung 9: „Herkömmliches“ Hörverstehen.	98
Abbildung 10: <i>Grammaticality judgement</i>	99
Abbildung 11: Wortschatztest.	100
Abbildung 12: <i>Sight-word recognition</i>	100
Abbildung 13: Hörverstehen Segmentierung.	101
Abbildung 14: Nichtverbaler Intelligenztest mit Ravens Matrices.	101
Abbildung 15: Eine prototypische Datenerhebungssituation, kurz vor der Datenerhebung. Rechts: Sitzplatz des/der Lernenden, links: Sitzplatz der Forscherin.	104
Abbildung 16: Screenshot einer anonymisierten Videoaufnahme aus der Hauptuntersuchung.	105
Abbildung 17: Blatt mit der Aufforderung zum Lauten Denken.	108
Abbildung 18: Die visuelle Aufmachung der Hörverstehensaufgabe (Szenario „Stadttrip“, Task B).	202
Abbildung 19: Das Design der Aufgabe „Walkies“, Seite 1.	220
Abbildung 20: Gebrauch der sieben am häufigsten codierten Strategien pro Schüler/in.	225
Abbildung 21: Ausschnitt aus dem codierten Lautdenkprotokoll des Szenario-Schülers Toby.	227
Abbildung 22: Ausschnitt aus dem codierten Lautdenkprotokoll der Off-Szenario-Schülerin Loup.	228
Abbildung 23: Sieben am häufigsten codierte Lernerstrategien pro Szenario- (SZ) und Off-Szenario- Schüler/innen (OFF), ohne Strategie L10.	245
Abbildung 24: Sieben am häufigsten codierte Testmanagementstrategien pro Szenario- (SZ) und Off- Szenario-Schüler/innen (OFF).	248
Abbildung 25: Alle codierten <i>test-wiseness</i> -Strategien pro Szenario- (SZ) und Off-Szenario- Schüler/innen (OFF).	252
Abbildung 26: Sieben am häufigsten codierte Strategien pro Task (links: Szenario-Aufgaben, rechts: Off-Szenario-Aufgaben).	254
Abbildung 27: Bearbeitungsdauer der Tasks im Szenario „Stadttrip“ mit und ohne Lautes Denken.	283

TABELLENVERZEICHNIS

Tabelle 1: Vergleichender Überblick über die neueren Studien zur Überprüfung der Teststrategien mittels introspektiver Verfahren.	67
Tabelle 2: Taskspezifikationen Szenario „Stadttrip“	89
Tabelle 3: Taskspezifikationen Off-Szenario.	90
Tabelle 4: Übersicht über die Teilnehmenden an der qualitativen Studie.	94
Tabelle 5: Rang der Schüler/innen aus der qualitativen Studie (n=30) in den „herkömmlichen“ Lese- und Hörverstehensaufgaben und in Teilkomponententests im Vergleich zur IFB- Teilstichprobe für Französisch (n=331).	103
Tabelle 6: Schüler/innen, die für die vertiefende Analyse ausgewählt wurden.	114
Tabelle 7: Tasks, die für die vertiefende Analyse ausgewählt wurden.	115
Tabelle 8: Codierung der Antworten als „richtig“ oder „falsch“	132
Tabelle 9: Codierte Lernerstrategien.	136
Tabelle 10: Codierte Testmanagementstrategien.	141
Tabelle 11: Codierte <i>test-wiseness</i> -Strategien.	147
Tabelle 12: Übersicht über richtige und falsche Antworten der Szenario-Schüler/innen zu den Tasks aus den vertiefenden Analysen (inkl. empirische Itemschwierigkeit aus der quantitativen Hauptstudie).	161
Tabelle 13: Übersicht über richtige und falsche Antworten der Off-Szenario-Schüler/innen zu den Tasks aus den vertiefenden Analysen (inkl. empirische Itemschwierigkeit aus der quantitativen Hauptstudie).	162
Tabelle 14: Häufigkeit der codierten Lernerstrategien inkl. Gewichtungen.	167
Tabelle 15: Häufigkeit der codierten Testmanagementstrategien inkl. Gewichtungen.	184
Tabelle 16: Häufigkeit der <i>test-wiseness</i> -Strategien inkl. Gewichtungen.	209
Tabelle 17: Items, bei denen Szenario-Schüler/innen <i>test-wiseness</i> -Strategien einsetzten (gemäss Codierungen im Lautdenkprotokoll und/oder <i>Stimulated-Recall</i> -Interview).	231
Tabelle 18: Items, bei denen Off-Szenario-Schüler/innen <i>test-wiseness</i> -Strategien einsetzten (gemäss Codierungen im Lautdenkprotokoll und/oder <i>Stimulated-Recall</i> -Interview).	235
Tabelle 19: Häufigkeit der Lernerstrategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“ und „25“.	243
Tabelle 20: Häufigkeit der Testmanagementstrategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“.	247
Tabelle 21: Häufigkeit der <i>test-wiseness</i> -Strategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“.	250

1 EINLEITUNG

In diesem Kapitel wird zuerst der Kontext der vorliegenden Untersuchung näher beleuchtet. Auch werden hier die Problemstellung, der Forschungsgegenstand und der methodologische Zugriff umrissen, wobei die Darstellung der Forschungsfragen erst in Kap. 3.2 erfolgt, damit die Herleitung ihrer Formulierung aufgrund der Erkenntnisse aus den Kap. 1 und 2 besser nachvollzogen werden kann. Zuletzt wird der Aufbau der Studie skizziert.

1.1 KONTEXT

Wie gut beherrschen die Schüler/innen Fremdsprachen¹ nach einigen Jahren Unterricht? Oder anders, präziser gefragt: Erreichen die Schüler/innen am Ende einer bestimmten Lernphase die Kompetenzen², die von nationalen Institutionen für den schulischen Fremdsprachenunterricht festgelegt werden? Um darauf eine Antwort zu finden, werden Leistungsmessungen (*large scale assessments*) durchgeführt, deren Ziel es ist,

to monitor student progress through standardized test administration, scoring, and reporting, to collect uniform baseline information from a large group of students across geographical areas, to provide diagnostic information to all stakeholders (teachers, students, parents, school, administrators, etc.), and for state level accountability purposes (Kunnan, 2008, S. 135).

In der Schweiz wurde zu diesem Zweck das Projekt „Überprüfung der Erreichung der Grundkompetenzen“ (ÜGK) lanciert (vgl. EDK, 2021), in dessen Rahmen 2017 erstmals die Grundkompetenzen der Fremdsprachenlernenden am Ende der Primarschule (6. Klasse) überprüft³ wurden. Dafür lösten die Schüler/innen computerbasierte Tests zum Hören und

¹ Der Begriff „Fremdsprache“ wird in dieser Arbeit im Kontext des schulischen Fremdsprachenunterrichts gebraucht.

² Der Kompetenzbegriff, der sich laut Vollmer (2010, S. 32) mittlerweile (auch) in der Fremdsprachendidaktik etablierte (vgl. auch Studer, 2010, S. 1264), stammt ursprünglich von Weinert. Dieser versteht unter Kompetenzen „die bei Individuen verfügbaren oder von ihnen erlernbaren kognitiven Fähigkeiten und Fertigkeiten, bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27–28). In diesem Sinne wird der Kompetenzbegriff auch in dieser Arbeit verwendet.

³ Die Begriffe „Beurteilung“, „Bewertung“ und „Überprüfung (von Kompetenzen)“ sowie ihre Ableitungen werden im Folgenden ähnlich wie in Arras (vgl. 2007, S. 13) im Sinne des englischen Begriffs „assessment“ (vgl. Council of Europe, 2001; Bachman & Palmer, 2010, S. 19–22) synonym verwendet, da ihre begriffliche Unterscheidung, die u.a. in der deutschen Übersetzung des Gemeinsamen Europäischen Referenzrahmens (Europarat, 2001, S. 172, Fussnote 6) gemacht wird, für die vorliegende Untersuchung unerheblich ist. Im Folgenden werden auch die Begriffe „Test“ und „Prüfung“ synonym verwendet. Gemeint ist damit immer die formale und standardisierte (d.h. unter gleichen Bedingungen ablaufende) Überprüfung der Sprachkompetenzen der Testteilnehmenden.

Lesen in ihrer ersten schulischen Fremdsprache, also je nach Schulsprache und Lehrplanregion in Französisch, Deutsch oder Englisch. Im Frühjahr 2020 war die Überprüfung der Grundkompetenzen am Ende der obligatorischen Schulzeit (Sekundarstufe I) in beiden schulischen Fremdsprachen geplant, d.h. je nach Schulsprache und Lehrplanregion in Englisch, Französisch oder Deutsch.⁴

Für die Erstellung der ÜGK-Lese- und Hörverstehensaufgaben, die am Ende der Sekundarstufe I eingesetzt werden sollten, war in den Jahren 2018-2019 das schweizerische Kompetenzzentrum für Mehrsprachigkeit (KFM) mitverantwortlich. Mit Blick auf diese Aufgabenentwicklung wurde am KFM in den Jahren 2016-2020 das Projekt „Innovative Formen der Beurteilung“ (IFB) durchgeführt (Karges, Lenz et al., 2021; KFM, 2020a). Das Projekt IFB (n=631) verscrieb sich der Grundlagenforschung im Bereich des Sprachentestens: Die Erfahrungen und Erkenntnisse aus diesem Projekt sollten wichtige Hinweise für die Konzipierung von computerbasierten, in Szenarien (Sabatini et al., 2014b) eingebettete Aufgaben zur Überprüfung und Beurteilung von rezeptiven Fremdsprachenkompetenzen liefern (vgl. auch Kap. 2.1.3.2). Davon sollten insbesondere zukünftige Testentwicklungsprojekte für Leistungsmessungen profitieren. Um dieses Ziel zu erreichen, orientierte sich IFB in Bezug auf wichtige inhaltliche Parameter an Vorgaben für die ÜGK: Die Tests sollten sich an den Kann-Beschreibungen der geltenden Lehrpläne der obligatorischen Schule am Ende der Sekundarstufe I und am Gemeinsamen europäischen Referenzrahmen für Sprachen (Europarat, 2001, im Folgenden GeR) orientieren, rezeptive Kompetenzen überprüfen und auf Tablets ohne Tastatur bearbeitet werden können.

Im Rahmen von IFB wurden innovative computerbasierte Testaufgaben zur Überprüfung von rezeptiven Sprachkompetenzen in schulischen Fremdsprachen entwickelt und erprobt. Das Merkmal „innovativ“ bezieht sich hauptsächlich auf zwei Aspekte: die Verwendung (Simulation) von digitalen Textsorten und die Einbettung der einzelnen Aufgaben in Szenarien, die mehrere Lese- und Höraufgaben in einem übergreifenden Handlungsrahmen vereinen. Der Einsatz von digitalen Textsorten wie Smartphone-Chats oder Podcasts trägt der Tatsache Rechnung, dass Kommunikation im alltäglichen Sprachgebrauch immer mehr über elektronische Kanäle abgewickelt wird. Digitale Textsorten erschienen zudem authentischer, da die Aufgaben am Computer gelöst werden sollten.

Die Zielgruppe waren Schüler/innen aus der Deutschschweiz am Ende der Sekundarstufe I. Diese Schüler/innen sollten gemäss der nationalen Bildungsstandards (Grundkompetenzen)

⁴ Italienisch und Romanisch als schulische Fremdsprachen wurden wegen der geringen Anzahl der Lernenden nicht getestet. Diese zweite Leistungsmessung musste im Zusammenhang mit der Covid-19-Pandemie auf 2023 verschoben werden (vgl. EDK, 2021).

in ihren beiden schulischen Fremdsprachen Französisch und Englisch am Ende der obligatorischen Schulzeit, also nach mindestens 5 bzw. 7 Lernjahren im Lese- und Hörverstehen mindestens das Niveau A2.2 erreicht haben (EDK, 2011). Ausgangspunkt für die Entwicklung der Lese- und Hörverstehensaufgaben im Projekt IFB waren die Beschreibungen der Grundkompetenzen in den nationalen Bildungsstandards (EDK, 2011) sowie die Kompetenzbeschreibungen des GeR und verwandter Dokumente. Aus diesen wurden wahrscheinliche Sprachverwendungsszenarien abgeleitet. Im Sinne des *Scenario-Based Assessment* (vgl. Kap. 2.1.3.2) wurden diese dann in nah-authentische Handlungsaufgaben (Tasks)⁵ mit jeweils mehreren Testitems⁶ unterteilt. Die gleichen Aufgaben wurden in zwei möglichst übereinstimmenden Fassungen umgesetzt: für Französisch und für Englisch als Fremdsprache.⁷

Die neuen szenariobasierten IFB-Aufgaben wurden mittels qualitativer und quantitativer Forschungsmethoden untersucht. Im qualitativen Teil, der den Gegenstand der vorliegenden Dissertationsstudie bildet, wurden mittels introspektiver Verfahren (Lautes Denken und *Stimulated Recall*) vor allem die Strategien der Schüler/innen erforscht (vgl. u.a. A. D. Cohen, 2014). Der quantitative Teil umfasste sowohl psychometrische Analysen der Resultate der szenariobasierten Lese- und Hörverstehenstests als auch Analysen des Zusammenhangs zwischen den Ergebnissen der kompetenzorientierten Tests und vermutlich korrelierenden Variablen, die miterhoben wurden. Die Kombination von qualitativen und quantitativen Forschungsmethoden sollte Einsichten in das Funktionieren und die Anforderungen dieser Aufgaben erlauben. Zudem sollten Aussagen zur Eignung der verschiedenen Erhebungsinstrumente ermöglicht werden. Die Studie liefert auch Hinweise zu den Englisch- und Französischkompetenzen der Deutschschweizer Schüler/innen am Ende der obligatorischen Schulzeit (vgl. Karges, Lenz et al., 2021). Die Testergebnisse sind jedoch aufgrund der Zusammensetzung der Stichprobe (*convenience sample*) nicht als repräsentativ für eine bestimmte Population von Schüler/innen zu betrachten.

⁵ Unter einer Aufgabe bzw. einem Task wird in dieser Arbeit nach Bachman und Palmer „an activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular setting“ (2010, S. 59–60) verstanden.

⁶ Ein Testitem ist „jedes Einzelelement eines Tests, das getrennt bewertet wird. Beispiele sind: Lücken in einem Cloze-Test, Multiple-Choice-Frage mit drei oder vier Auswahlantworten, ein Satz, der grammatikalisch umzuformulieren ist, oder eine Frage, auf die ein Satz als Antwort erwartet wird“ (ALTE, 1998, S. 107).

⁷ Diese Aufgaben können unter <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000003403> eingesehen werden.

1.2 PROBLEMSTELLUNG, FORSCHUNGSGEGENSTAND UND METHODOLOGISCHER ZUGRIFF

Neu erstellte Testaufgaben werden vor ihrem Einsatz in standardisierten Beurteilungsanlässen normalerweise während der sogenannten Testerprobungsphase (*pre-operational testing*, vgl. Kenyon & MacGregor, 2012) gründlich geprüft. Die Phase der Testerprobung, auf die detaillierter in Kap. 2.2.4 eingegangen wird, sollte insbesondere bei der Konzipierung und Erstellung von neuen Aufgaben- und Itemtypen nicht aussen vorgelassen werden (vgl. Kenyon & MacGregor, 2012; Reed, 2014). Dabei soll z.B. untersucht werden, ob die Aufgaben psychometrisch wie gewünscht funktionieren, ob ihre Schwierigkeit angemessen ist, ob die Testteilnehmenden die Anweisungen und Aufgabenstellungen verstehen oder ob sich das Test- und Aufgabendesign für die Zielgruppe eignet. Es soll also aussagekräftige qualitative und quantitative Evidenz für die Validität der Testnutzung gesammelt werden, welche nach Messick definiert wird als „ein integriertes bewertendes Urteil über das Ausmaß, in dem die Angemessenheit und die Güte von Interpretationen und Maßnahmen auf Basis von Testwerten oder anderen diagnostischen Verfahren durch empirische Belege und theoretische Argumente gestützt sind“ (Messick, 1989, S. 13; deutsche Übersetzung von Hartig in Hartig et al., 2012, S. 144).

Wie bereits erwähnt, wurden im Projekt IFB neuartige computerbasierte Testaufgaben zur Überprüfung von Hör- und Leseverstehenskompetenzen in schulischen Fremdsprachen entwickelt und erprobt. Das Ziel des Projekts war es, unter Einbezug von qualitativen und quantitativen Methoden empirische Erkenntnisse zum Funktionieren (oder auch Nicht-Funktionieren) und zur Validität (Messick, 1989) von neu konzipierten, innovativen und prototypischen Aufgaben- und Itemtypen zu gewinnen. In diesem Sinne könnte auch das gesamte Projekt IFB als ein umfassendes *pre-operational testing* (Kenyon & MacGregor, 2012) der prototypischen Aufgaben- und Itemtypen betrachtet werden. Daraus folgt auch, dass die Testaufgaben nicht für den unmittelbaren Einsatz in einer Leistungsmessung vorgesehen waren.

Die Beurteilung von Sprachkompetenzen ist mannigfaltig und wird von zahlreichen, vielfältigen Faktoren sowie deren Zusammenspiel beeinflusst. Diese Faktoren können auf den Beurteilungskontext, auf die Beurteilungsinstrumente und auf die Testteilnehmenden zurückgeführt werden (vgl. u.a. Bachman & Palmer, 2010). Im Prozess der Testvalidierung, d.h. während Evidenz für die Validität einer Testnutzung gesammelt wird (vgl. Kap. 2.2.4), werden diese Faktoren und ihre Interaktion unter die Lupe genommen. Im Fokus der Testvalidierung stehen seit der einflussreichen Validitätsdefinition von Messick (1989) nicht

nur (wie bis dahin) der Test und seine Eigenschaften, die traditionell mit quantitativen Methoden untersucht werden, sondern u.a. auch die kognitiven Prozesse der Testteilnehmenden und ihr Antwortverhalten beim Lösen der Testaufgaben (*response processes*) (vgl. Chapelle, 2012; Ercikan & Pellegrino, 2017a). Das Interesse an den kognitiven Prozessen, die der Leistung in einem Sprachtest zugrunde liegen, entstand laut Purpura aus Bedenken hinsichtlich der Validität einer Testnutzung, die damit zusammenhängen, inwieweit sich die Testteilnehmenden mental in einer vom Testkonstrukt⁸ beabsichtigten Weise mit den Testaufgaben beschäftigen (2014, S. 1452). Der Zusammenhang zwischen dem Testkonstrukt und den kognitiven Prozessen der Testteilnehmenden während der Testbewältigung ist jedoch entscheidend für die bedeutungsvolle Interpretation und Verwendung der Informationen aus der Beurteilung (ebd., S. 1452-1453). Für die Untersuchung der kognitiven Prozesse und des Antwortverhaltens der Testteilnehmenden (aber auch des Beurteilungskontextes, z.B. des Verhaltens der Beurteilenden) werden seit den 1990er Jahren vermehrt und ergänzend zu quantitativen Methoden auch qualitative Forschungsmethoden eingesetzt (vgl. Arras, 2013; Lumley & Brown, 2005).

Vor dem oben beschriebenen Hintergrund ist die vorliegende Studie wie folgt einzuordnen: Sie ist qualitativ orientiert, explorativ-interpretativ ausgerichtet und wurde als eigenständiger Projektteil, aber ergänzend zum vorwiegend quantitativen Projekt IFB durchgeführt, in dem Erkenntnisse zur Validität von neuen, szenariobasierten Aufgaben- und Itemtypen gewonnen werden sollten. Das gesamte Projekt IFB ist somit als eine *mixed-methods*-Studie zu verstehen und das Verhältnis der qualitativen und quantitativen Teilstudien des Projekts IFB lässt sich nach Kuckartz (2014) als „paralleles Design“⁹ bezeichnen, d.h. „das Projekt besteht aus einer qualitativen und einer quantitativen Teilstudie, die parallel zueinander angeordnet sind, d.h. gleichzeitig durchgeführt werden“ (ebd., S. 71). Im Fokus der qualitativen Untersuchung stehen die Herangehensweisen der Testteilnehmenden an die IFB-Testaufgaben, die aus einer emischen Perspektive, d.h. durch einen analytischen und interpretativen Blick auf die Innenansicht der Testteilnehmenden durch die Forscherin herausgearbeitet werden (vgl. Schramm, 2016, S. 51). Das gesamte Projekt IFB ist im Bereich der Sprachtestforschung angesiedelt und gehört somit dem rationalistischen Forschungsparadigma an (Harsch, 2016, S. 205; vgl. auch L. Cohen et al., 2018).

⁸ Mit einem Testkonstrukt ist die „theoretische Definition der Fertigkeit oder Kompetenz, die gemessen werden soll“ gemeint (Eberharter et al., 2018, S. 59).

⁹ In der englischsprachigen Literatur wird diese Form des *mixed-methods*-Designs als *convergent design*, *concurrent design* oder *parallel design* bezeichnet (Kuckartz, 2014, S. 72).

Die Forschenden im Sprachtestbereich bedienen sich unterschiedlicher Ansätze, um die Zusammenhänge zwischen kognitiven Prozessen und der Beurteilung der Sprachkompetenzen zu untersuchen. Diese Ansätze unterteilt Purpura (2014) in *the factorial approach*, *the "Learner Strategies" approach*, *the strategic competence approach*, *the cognitive processing approach*, und *the social approach*. In der vorliegenden Studie wurde der von A. D. Cohen (u.a. 2006, 2011b, 2014) vorgeschlagene konzeptuelle Rahmen übernommen¹⁰, den Purpura (2014) als *the "Learner Strategies" approach* bezeichnet. Dieser *bottom-up*-Ansatz ist an internen mentalen Prozessen der Testteilnehmenden beim Lösen der Testaufgaben interessiert und der Zugriff auf Informationen über diese Prozesse geschieht „by asking learners to verbalize the "strategies" (i.e., the thoughts and behaviors) they reported using to respond to test questions, or by asking them direct questions about their strategies“ (Purpura, 2014, S. 1455). Die mithilfe von introspektiven Verfahren (siehe Kap. 3.3) erhobenen Daten werden untersucht, um festzustellen, ob die von den Testteilnehmenden eingesetzten Strategien für das zu messende Testkonstrukt relevant oder irrelevant sind (vgl. Purpura, 2014, S. 1455). A. D. Cohen (2011b, S. 305) definiert Strategien, die Testteilnehmende beim Lösen der Sprachtestaufgaben einsetzen, als (mehr oder weniger) bewusst gewählte Prozesse zur Handhabung der sprachbezogenen Aspekte und der Anforderungen der Testaufgaben. Von diesem Verständnis der Strategien wird auch in der vorliegenden Studie ausgegangen.

Die Strategien der Testteilnehmenden, auf die in Kap. 2.2.3 ausführlicher eingegangen wird, unterteilt A. D. Cohen (2014) in 1) Lernerstrategien (*language learner strategies*), und 2) Teststrategien (*test taking strategies*), die sich wiederum in die 2a) Testmanagementstrategien (*test-management strategies*) und 2b) *test-wiseness*-Strategien¹¹ gliedern lassen. Lernerstrategien (z.B. Ableitung der Bedeutung eines unbekannten Wortes aus dem Kontext) sind im Gegensatz zu Teststrategien nicht beschränkt auf eine Testsituation: Es sind Strategien, die von den Sprachlernenden zur Bewältigung jeglicher kommunikativen Aufgaben in der L2¹² benutzt werden und die als für das Testkonstrukt relevant zu betrachten sind. Testmanagementstrategien (z.B. ein systematischer Umgang mit den Antwortoptionen in Multiple-Choice-Aufgaben) helfen den Lernenden, die Testaufgaben sinnvoll anzugehen und zu lösen und sind daher ebenfalls konstruktrelevant. Als irrelevant für das Testkonstrukt sind hingegen *test-wiseness*-Strategien

¹⁰ Wie u.a. in den Studien von A. D. Cohen & Upton (2006), L. Gu & So (2017) und Suvorov (2018).

¹¹ Die deutschen Übersetzungen werden hier nach Wisniewski (2016) verwendet.

¹² „L2“ wird in dieser Arbeit im Sinne von „nicht Erstsprache (L1)“ als Oberbegriff verwendet, wenn beim besprochenen Sachverhalt die Unterscheidung zwischen der im Unterricht gelehrt und gelernten Fremdsprache und der Zweitsprache (mit der die im Zielland erworbene und im Alltag der Lernenden kommunikativ unmittelbar relevante Sprache gemeint ist) unerheblich ist.

einzustufen, d.h. jene Strategien, mit denen die Testteilnehmenden (aus welchen Gründen auch immer) den Test zu „überlisten“ versuchen: Sie nutzen dafür beispielsweise Randinformationen wie das Wissen über das Funktionieren der Testformate oder wählen in einem Multiple-Choice-Lesetest eine bestimmte Antwortoption, weil sie anders als die restlichen Antwortoptionen aussieht (vgl. u.a. A. D. Cohen, 2011b, 2014).

Die Ergebnisse der Untersuchungen zu Strategien von Testteilnehmenden werden als Evidenz für die Validität eines Tests genutzt oder dienen als Grundlage für dessen Überarbeitung (Purpura, 2014, S. 1455; vgl. auch Harsch, 2016, S. 208). So müssten beispielsweise bestimmte IFB-Aufgaben- und -Itemformate überarbeitet (und vor ihrem eventuellen Einsatz in Leistungsmessungen nochmals erprobt) werden, wenn sich beispielsweise zeigen sollte, dass sie Strategien begünstigen oder hervorrufen, die aus Sicht des Testkonstrukts als irrelevant einzustufen sind. Ein Beispiel dafür wäre ein nicht optimal konstruiertes Multiple-Choice-Item, bei dem die Testteilnehmenden nur aufgrund ihres Weltwissens und auch ohne Lektüre des Inputtextes die richtige Antwort problemlos finden könnten. Ebenso müssten z.B. Aufgaben überarbeitet werden, bei denen die Testteilnehmenden die Frage(n) zum Inputtext nicht verstehen oder nicht wissen, wie sie ihre Antworten am Computer eingeben sollen.

Hauptziel der vorliegenden Untersuchung ist es, Erkenntnisse darüber zu gewinnen, wie die Testteilnehmenden beim Lösen der IFB-Aufgaben vorgehen, mit welchem Erfolg sie dabei auf konstruktirrelevante *test-wiseness*-Strategien zurückgreifen und welchen Einfluss (wenn überhaupt) die Szenarioeinbettung der Aufgaben auf die Lösungsprozesse der Lernenden hat. Um Licht auf diese Problematik zu werfen, wurde der von A. D. Cohen (u.a. 2006, 2011b, 2014) vorgeschlagene und von Purpura (2014) als *the “Learner Strategies” approach* benannte Ansatz gewählt: Dank der Verbalisierungen der Testteilnehmenden sollen wichtige und mit anderen Mitteln unerreichbare Einsichten in ihre Gedanken gewonnen werden. Die im *learner-strategies*-Ansatz explizit vorgenommene Unterscheidung in konstruktrelevante und konstruktirrelevante Strategien wurde darüber hinaus (in Übereinstimmung mit A. D. Cohen & Upton, 2006; L. Gu & So, 2017; Suvorov, 2018) als hilfreich erachtet, um das gemessene Testkonstrukt besser zu verstehen und zu eruieren, welche Testaufgaben eventuell anfällig für den Gebrauch konstruktirrelevanter, also unerwünschter *test-wiseness*-Strategien sind.

An der vorliegenden Studie nahmen 30 Deutschschweizer Schüler/innen aus der IFB-Stichprobe teil, die in individuellen, zweistündigen Sitzungen ausgewählte Französischaufgaben lösten. Mit der Untersuchung der Aufgaben zum Französischen als Fremdsprache soll ein kleiner Beitrag zur Erforschung von Sprachtestaufgaben in einer

anderen schulischen Fremdsprache als Englisch geleistet werden. Zudem genießt Französisch in der Schweiz einen besonderen Status als eine der vier Nationalsprachen. Ursprünglich war geplant, in der qualitativen Studie zusätzlich auch die englischsprachigen Versionen der IFB-Aufgaben unter die Lupe zu nehmen. Die Pilotierung zeigte aber, dass sich die englischen Inputtexte in den IFB-Aufgaben (Zielniveau: A2/B1) für die Schüler/innen als sehr einfach erwiesen (vgl. Kap. 3.4.1). Dies führte dazu, dass sie beim Lösen der Englischaufgaben wenig zu verbalisieren hatten, was folglich auch in wenig interessanten Daten für die Analyse resultierte. Es wurde daher entschieden, in dieser Studie nur die französische Version der Aufgaben zu untersuchen.

Um den Einfluss der Szenarioeinbettung auf die Lösungswege der Lernenden besser verstehen zu können, bearbeiteten 17 Schüler/innen Aufgaben, die in ein übergreifendes Szenario eingebettet waren, während die restlichen 13 eine ähnliche Anzahl Aufgaben lösten, die aber aus verschiedenen Szenarien stammten und daher nicht thematisch verbunden waren (im Folgenden als „Off-Szenario-Aufgaben“ bezeichnet). Um einen möglichst umfassenden Einblick in die Gedanken der Schüler/innen zu bekommen, wurden im Sinne der Triangulation (vgl. Knorr & Schramm, 2016) zwei introspektive Datenerhebungsmethoden kombiniert: Lautes Denken (vgl. u.a. Bowles, 2010; Heine & Schramm, 2007) und *Stimulated Recall* (u.a. Gass & Mackey, 2017; Knorr & Schramm, 2012). Die Schüler/innen wurden am Anfang der Datenerhebungssession in die Methode des Lauten Denkens eingeführt (das Training dauerte ca. 20 Minuten) und gebeten, ihre Gedanken während der Aufgabenbearbeitung zu verbalisieren (Dauer: ca. 45 Minuten). Im Anschluss an die Bearbeitung der Aufgaben wurden mit den Schüler/innen ca. einstündige *Stimulated-Recall*-Interviews durchgeführt, in denen die soeben gelösten Aufgaben besprochen wurden. Als Stimulus dienten dabei Bildschirmseiten der bearbeiteten Aufgaben, in denen die Lösungen der Schüler/innen sichtbar waren. Die Lernenden wurden in den Interviews auch gebeten, die Inhalte der Inputtexte in ihrer Schulsprache bzw. in ihrem Dialekt zusammenzufassen. Ziel war es, Falsch- oder Missverständenes bei der Datenanalyse berücksichtigen zu können. In einer ähnlichen Studie (Karges, Barras et al., 2021) wurde zu einem früheren Zeitpunkt festgestellt, dass Lernende mit einer schwach ausgeprägten Fremdsprachenkompetenz beim Lesen der fremdsprachigen Texte oft etwas anderes verstanden, als im Text zu lesen war (vgl. Barras, 2018).

Die Datensitzungen wurden mithilfe der Software OBS (*OBS Studio*, 2017) und der Bildschirmdkamera aufgenommen. Dies geschah mit der mündlichen Zustimmung der Lernenden und schriftlicher Zustimmung ihrer Erziehungsberechtigten. Auf den Aufnahmen sind die Gesichter der Schüler/innen zu sehen sowie alles, was bei der Bearbeitung der

Aufgaben auf dem Bildschirm passierte. Alle Datensessionen wurden transkribiert (GAT 2-Basistranskript). Aufgrund der grossen Datenmenge und Validitätsprobleme einiger Datensätze (z.B. im Fall von Schüler/innen, die beim Lauten Denken nicht oder wenig verbalisierten) wurden für die vertiefende Analyse der Strategien Transkriptausschnitte von 20 Schüler/innen (10 „Szenario-Schüler/innen“ und 10 „Off-Szenario-Schüler/innen“) zu bestimmten Items ausgewählt.

Die Analyse wurde nach den Prinzipien der inhaltlich strukturierenden und evaluativen qualitativen Inhaltsanalyse (Kuckartz, 2018) durchgeführt. Das für die Analyse entwickelte Kategoriensystem orientierte sich an der von A. D. Cohen (u.a. 2014) vorgeschlagenen Klassifizierung der Strategien in Lerner-, Testmanagement- und *test-wiseness*-Strategien. Für die Auswertung der *Stimulated-Recall*-Interviews wurde ein paralleles Kategoriensystem eingesetzt. Die Auswertung der *Stimulated-Recall*-Daten hatte zum Ziel, die in den Lautdenkprotokollen vorgenommenen Codierungen zu validieren und ggf. um weitere relevante Informationen zu ergänzen, die von den Schüler/innen beim Lauten Denken nicht verbalisiert wurden.

Im Zentrum dieser Arbeit standen, wie erwähnt, die Strategien der Französischlernenden beim Lösen der neuen computerbasierten Testaufgaben. Den umfangreichen Verbalprotokollen der Schüler/innen wären aber auch weitere interessante Erkenntnisse zu entnehmen, z.B. ganz konkrete Hinweise zur Benutzerfreundlichkeit (*usability*) des Tests und seiner Teile, zu der von Schüler/innen empfundenen Schwierigkeit der Aufgaben, zu den Wünschen der Lernenden in Bezug auf den Fremdsprachenunterricht und die Beurteilungsanlässe, oder zu den Einstellungen der Lernenden gegenüber der französischen Sprache und dem Französischunterricht. Auf alle diese wichtigen und interessanten Punkte kann im Rahmen dieser Arbeit aus Platzgründen nicht ausführlich eingegangen werden. In Kap. 4.5 werden jedoch kurz ausgewählte Erkenntnisse dargestellt, die sich zwar nicht direkt auf die Forschungsfragen beziehen, aber u.a. für Testentwickler/innen und Fremdsprachenlehrpersonen dennoch aufschlussreich sein könnten.

1.3 AUFBAU DER STUDIE

Die vorliegende Arbeit ist in drei Teile gegliedert. Der erste Teil (Kap. 2) ist den relevanten Erkenntnissen aus der Forschungs- und Fachliteratur, d.h. den theoretischen Grundlagen der Studie, gewidmet. In diesem Teil wird auf die Problematik der Überprüfung der Lese- und Hörverstehenskompetenzen der Fremdsprachenlernenden aus der Perspektive der Sprachtestforschung eingegangen. Auch Strategien sowie ihr Einsatz bei der Bewältigung von

Sprachtestaufgaben werden thematisiert. Des Weiteren werden die Ziele der Erforschung von Strategien der Testteilnehmenden ausführlicher dargestellt.

Im zweiten Teil (Kap. 3) wird die Vorgehensweise bei der Erhebung der introspektiven Daten im Detail beschrieben und begründet. In diesem Teil wird auch geschildert, wie die Daten aufbereitet, für die Analyse ausgewählt und schliesslich ausgewertet wurden.

Der dritte Teil (Kap. 4 und 5) ist der Auswertung und Diskussion der empirischen Studie gewidmet. In Kap. 4 werden die Ergebnisse dargestellt, die im Hinblick auf die Beantwortung der Forschungsfragen erarbeitet wurden. Hier wird auch über weitere ausgewählte Ergebnisse zur Einschätzung des Lauten Denkens durch die Schüler/innen, zum Einsatz von digitalen Textsorten in einem Sprachtest sowie zur Sicht der Schüler/innen auf die Sprachtestaufgaben berichtet. In Kap. 5 werden die Resultate schliesslich zusammengefasst und in Bezug auf die theoretischen Grundlagen der Arbeit diskutiert.

2 KONZEPTUELLER RAHMEN UND LITERATURÜBERBLICK

In diesem Kapitel werden die wichtigsten Erkenntnisse aus der Forschungs- und Fachliteratur zusammengefasst, die die theoretischen Grundlagen der vorliegenden Untersuchung bilden. Das erste Unterkapitel (2.1) ist der kommunikativen Beurteilung der Lese- und Hörverstehenskompetenzen der L2-Lernenden (vor allem aus dem Blickwinkel der Sprachtestforschung) gewidmet. Dabei ist zu beachten, dass die Überprüfung und Bewertung von Sprachkompetenzen zwar aus verschiedenen Gründen durchgeführt werden kann (vgl. u.a. Bachman & Palmer, 2010; Grabe & Jiang, 2014; Anthony Green, 2014; Harsch, 2016; Hughes, 2003), der Fokus der vorliegenden Untersuchung und somit auch dieses Literaturüberblicks jedoch auf dem *large-scale assessment* (vgl. J. Fox, 2014; Kunnan, 2008) im Kontext des schulischen Fremdsprachenunterrichts liegt.

In einem weiteren Unterkapitel (2.2) wird auf die Strategien von Sprachlernenden eingegangen. Von besonderem Interesse sind hier Strategien, auf die die Lernenden bei der Bewältigung von rezeptiven Sprachtestaufgaben zurückgreifen bzw. potenziell zurückgreifen können. Diese Strategien werden häufig im Rahmen der Testerprobung mittels Introspektion erforscht und die Ergebnisse dieser Untersuchungen spielen eine wichtige Rolle im Prozess der Testvalidierung, wie in einem weiteren Teil dieses Unterkapitels aufgezeigt wird. Zuletzt werden fünf ausgewählte Studien zusammengefasst (A. D. Cohen & Upton, 2006; L. Gu & So, 2017; Nikolov, 2006; Rossa, 2012; Suvorov, 2018), die bei der Planung der Datenerhebung und -analyse in der vorliegenden Untersuchung von besonderer Bedeutung waren.

2.1 REZEPTIVE KOMPETENZEN IN DER L2 ÜBERPRÜFEN UND BEWERTEN

Die Überprüfung und Bewertung der Lese- und Hörverstehenskompetenzen der L2-Lernenden hat zum Ziel, Erkenntnisse darüber zu gewinnen, wie gut die Testteilnehmenden ihre (fremd-)sprachlichen Fähigkeiten einsetzen können, um relevante schriftliche und mündliche Inputtexte zu verstehen, mit denen sie sich im Alltag und/oder als Teil eines Sprachlernprogramms auseinandersetzen müssen (Anthony Green, 2014, S. 97). Lese- und Hörverstehen geschehen intentional und zielgerichtet und werden als „höchstkomplexe und simultan ablaufende kognitive Vorgänge in den Köpfen der Sprachlernenden“ bezeichnet (Stadler, 2018, S. 69). Eine grosse Schwierigkeit bei der Überprüfung von Hör- und Leseverstehenskompetenzen liegt darin, dass Verstehensprozesse und ihre Ergebnisse nicht direkt beobachtbar sind. Sie müssen daher erst aus bestimmten Verständnissnachweisen der Testteilnehmenden, d.h. aus den „Produkten“ ihres Verstehens (vgl. u.a. Alderson et al., 2015, S. 72–73; Koda, 2012, S. 272)

inferiert und interpretiert werden. Anthony Green (2014, S. 97) bringt es folgendermassen auf den Punkt:

Simply watching someone read a report or listen to the news on the radio can generally tell us very little about how much they are able to understand. Assessment designers can only find evidence of a language learner's receptive language processing by getting them to do something else (in addition to reading or listening) in order to demonstrate how well they have understood. This 'something else' might involve the assessee performing actions, drawing, speaking or writing.

Um angemessene Aussagen über die Lese- und Hörverstehenskompetenzen der Testteilnehmenden machen zu können, müssen also dafür geeignete Aufgaben eingesetzt werden. Diese Aufgaben sollen die Testteilnehmenden nicht nur zum Lese- oder Hörverstehen veranlassen, sondern auch ein Verhalten elizitieren, das den erfolgreichen Einsatz der vom Test gesuchten Fähigkeiten demonstriert (vgl. Hughes, 2003, S. 136; für vertiefende Informationen zum komplexen Prozess der Testaufgabenentwicklung siehe u.a. Bachman & Palmer, 1996, 2010; und für praktische Hinweise zur Entwicklung von Lese- und Hörverstehensaufgaben vgl. z.B. Anthony Green, 2014; Hughes, 2003).

Die Beschaffung aussagekräftiger Evidenz über die Lese- und Hörverstehenskompetenzen der Testteilnehmenden beinhaltet laut Anthony Green (2014, S. 98) folgende Schritte:

- defining the nature of the knowledge, skills or abilities to be measured;
- choosing and specifying the kinds of texts or recordings that will be appropriate for use as input; and
- deciding on the kinds of tasks that assessees with these abilities should be able to perform when they encounter these texts or recordings.

Beim ersten Punkt handelt es sich um den „starting point for test construction“ (Buck, 2001, S. 1), also um das Festlegen des Testkonstrukts. Buck listet drei mögliche Ausgangspunkte für die Konstruktdefinition auf: 1) die Definition der Kompetenzen, die die Testteilnehmenden haben sollten, 2) die Definition der Aufgaben, die die Testteilnehmenden bewältigen können sollten, und 3) eine Kombination der beiden Ansätze; wobei dieser dritte Punkt für die meisten Sprachtests von grösster praktischer Relevanz sei (2001, S. 102–111; vgl. auch Grotjahn, 2012, S. 73–74).

Sind erste Schritte zu einem Testkonstrukt hin erfolgt, so gilt es eine Reihe weiterer Aspekte zu beachten, die für eine Testdurchführung relevant sind. Diese wichtigen Eckpunkte führen Bachman und Palmer (1996, 2010) in ihrem *framework of language tasks characteristics* auf.

Dazu gehören 1) Setting (darunter u.a. die Eigenschaften der Testteilnehmenden), 2) Aufgabenkontext (*assessment rubric*¹³), 3) Input, 4) erwartetes Antwortverhalten sowie 5) Wechselbeziehung zwischen dem Input und dem erwarteten Antwortverhalten (für vertiefende Informationen siehe Bachman & Palmer, 2010, S. 59–84). Die Charakteristika all dieser Variablen können sich unterschiedlich auf die Schwierigkeit der Testaufgaben auswirken, die u.a. dadurch bestimmt wird, ob die Aufgaben zeigen, was getestet werden soll und wie sie bewertet werden (für Erkenntnisse zu Itemschwierigkeit und/oder Effekten der Testformate in Lese- bzw. Hörverstehenstests vgl. u.a. Freedle & Kostin, 1993, 1996; In'nami & Koizumi, 2009; Kirsch, 2001; Kostin, 2004).

Wurden die Testcharakteristika festgelegt, die Testaufgaben und -items auf dieser Basis erstellt und von den Testteilnehmenden gelöst, gilt es, die Testergebnisse valide zu interpretieren. Auf die Validität und Validierung von Sprachtests wird in Kap. 2.2.4 näher eingegangen. An dieser Stelle müssen jedoch zwei grundlegende Gefahren für die Validität von Interpretationen der Testergebnisse erwähnt werden, auf die Messick (1989) hinweist, d.h. auf: 1) die konstruktirrelevante Varianz und 2) die Unterrepräsentation des zu messenden Konstrukts. Konstruktirrelevante Varianz liegt vor, wenn bestimmte Eigenschaften der Testaufgaben oder der Testteilnehmenden, die nichts mit dem eigentlichen Testkonstrukt zu tun haben, die Bewältigung der Testaufgaben schwieriger oder einfacher machen. Weil die Ergebnisse von Lese- und Hörverstehensprozessen nicht direkt beobachtbar sind und mithilfe von „etwas anderem“ erfasst werden müssen, entsteht konstruktirrelevante Varianz zum Beispiel durch nicht oder schlecht verständliche Testinstruktionen oder durch die (evtl. nicht vorhandenen) Fähigkeiten der Testteilnehmenden, so zu handeln, zu zeichnen, zu sprechen oder zu schreiben, wie es in der Testaufgabe vorgesehen wurde (vgl. Anthony Green, 2014, S. 97). Von einer Unterrepräsentation des Konstrukts spricht man hingegen, wenn in einem Test wichtige Elemente des zu messenden Konstrukts zu kurz kommen. Für die Überprüfung sprachlicher Kompetenzen sollten daher Testaufgaben eingesetzt werden, die zu möglichst wenig konstruktirrelevanter Varianz führen und das zu messende Konstrukt möglichst umfassend abdecken. Aus diesem Grund sollte im Rahmen der Testerprobung untersucht werden, ob die Testaufgaben für die Messung des intendierten Konstrukts geeignet sind und wie gewünscht funktionieren (mehr dazu in Kap. 2.2.4).

¹³ Bachman und Palmer (2010, S. 69) verstehen darunter den Kontext, in dem die Sprach(test)aufgaben ausgeführt werden: "The rubric includes those characteristics that provide the purpose and structure for particular tasks and that indicate how language users or test takers are to proceed in accomplishing the tasks. The characteristics of rubric include the following: 1) the instructions, 2) the structure of the assessment: how the tasks are organized in the assessment, 3) the duration of the assessment as a whole and of the individual tasks, and 4) how a record will be produced from the test taker's response."

In dieser Studie wurden neu erstellte und prototypische Testaufgaben zum Überprüfen des Lese- und Hörverstehens untersucht, die potenziell für ein Bildungsmonitoring eingesetzt werden könnten. Im Folgenden werden daher zunächst die wichtigsten Erkenntnisse zum Leseverstehen und zur Überprüfung und Bewertung dieser Kompetenz zusammengefasst. Ein weiteres, parallel aufgebautes Unterkapitel ist dem Hörverstehen und seiner Überprüfung und Bewertung gewidmet. Da vieles, was über das Leseverstehen und seine Beurteilung bekannt ist, auch für das Hörverstehen gilt (obwohl es natürlich zwischen den beiden Prozessen auch fundamentale Unterschiede gibt), fällt dieses Unterkapitel entsprechend etwas kürzer aus. Zum Schluss wird auf neuere, für die vorliegende Studie wichtige Entwicklungen im Bereich der Überprüfung der rezeptiven Kompetenzen eingegangen. Dazu gehören der Einsatz von digitalen Textsorten in Sprachtests und der sogenannte szenariobasierte Ansatz (u.a. Sabatini et al., 2014a), d.h. die Einbettung von Testaufgaben in ein übergreifendes Szenario.

2.1.1 LESEVERSTEHEN

In diesem Teil wird zuerst dargestellt, was in der aktuellen Diskussion unter dem Prozess des Leseverstehens verstanden wird (Kap. 2.1.1.1). Dies wird in erster Linie aus der Perspektive der Sprachtestforschung unternommen. Die Erkenntnisse zum Leseverstehen basieren vor allem auf Arbeiten, die diesen Prozess aus der L1-Perspektive erforschen. Auf die Unterschiede zwischen dem Lesen in der L1 und in der L2 wird in Kap. 2.1.1.2 näher eingegangen. Das nächste Unterkapitel (2.1.1.3) ist dem – im Sprachtestbereich mittlerweile weitläufig anerkannten – Modell des Leseverstehens von Khalifa und Weir (2009) gewidmet, das auch die Basis des Lese- und Hörverstehenskonstrukts im Projekt IFB bildete. Zuletzt wird die Problematik der Überprüfung und Bewertung des Leseverstehens in einer L2 näher beleuchtet (Kap. 2.1.1.4). Für eine umfassendere Darstellung des L2-Leseverstehens und seiner Überprüfung muss an dieser Stelle auf weiterführende Fachliteratur verwiesen werden (z.B. Alderson, 2000; Alderson et al., 2015; Grabe, 2009; Grabe & Stoller, 2011; Koda, 2005).

2.1.1.1 LESEVERSTEHEN ALS PROZESS

Leseverstehen ist eine komplexe Aktivität (vgl. u.a. Alderson, 2000; Alderson et al., 2015; Adolf et al., 2011; Grabe, 2009; Koda, 2012; Kong, 2019; Schramm, 2001), die von vielen Faktoren und parallel ablaufenden Prozessen beeinflusst wird, die kurz als „a combination of text input, appropriate cognitive processes, and the information that we already know“ (Grabe, 2009, S. 74) zusammengefasst werden können. Flüssiges Lesen ist laut Grabe und Stoller (2011, S. 11) 1) ein schneller Prozess, 2) ein effizienter Prozess, 3) ein interaktiver Prozess, 4) ein strategischer Prozess, 5) ein flexibler Prozess, 6) ein evaluativer Prozess, 7) ein

gezielter Prozess, 8) ein Verstehensprozess, 9) ein Lernprozess, und 10) ein linguistischer Prozess. Das Endziel des Lesens ist es, auf Basis von visuell codierten linguistischen Informationen Textbedeutungen zu konstruieren (Koda, 2012, S. 227). Als solches umfasst das Lesen drei Hauptschritte, die aufeinander aufbauend erlernt werden: „(a) decoding (extracting linguistic information from printed words), (b) text-information building (assembling locally constructed information into larger text units), and (c) reader-model construction (integrating background knowledge and assembled text meanings)“ (ebd.). Koda (ebd.) weist dabei auf den generellen Konsens hin, dass es sich beim Textverständnis nicht um einen „all or nothing pursuit“, sondern „a matter of degree“ handelt.

Der Prozess des Leseverstehens wurde mittlerweile in zahlreichen Arbeiten umfassend untersucht und veranschaulicht (vgl. z.B. Alderson, 2000; Alderson et al., 2015; Grabe, 2009; Koda, 2005; Kong, 2019; Shiotsu, 2010). Am gründlichsten wurde dabei das Lesen in der L1, und zwar vor allem in Englisch als L1 erforscht (Alderson et al., 2015, S. 69). Auf der Grundlage dieser Studien wurde das Lesekonstrukt von verschiedenen Autor/innen ausführlich und – wie Grabe und Jiang (2014, S. 187) bemerken – grösstenteils auch überlappend beschrieben. So sind sich die Expert/innen darüber einig, dass beim Prozess des Leseverstehens die Sprachkenntnisse, Wissensressourcen und allgemeinen kognitiven Fähigkeiten der Lesenden zum Einsatz kommen (Grabe & Jiang, 2014, S. 188). Aufgrund einer umfassenden Analyse von Resultaten aus der Leseforschung aus L1- und L2-Kontexten erstellten Grabe und Jiang (ebd.) eine Liste von Faktoren, die einen starken Einfluss auf den Leseprozess ausüben. Diese Faktoren treten in verschiedenen Kombinationen auf und erklären die individuell ausgeprägten Lesefähigkeiten (*reading abilities*) von Leser/innen, die zu verschiedenen Zwecken lesen und auch verschiedenen Leseleistungsniveaus zugeordnet werden können (ebd.). Zu diesen Faktoren gehören:

1. efficient **word recognition** processes (phonological, orthographic, morphological, and semantic processing);
2. a large **recognition vocabulary** (vocabulary knowledge);
3. efficient **grammatical parsing** skills (grammar knowledge under time constraints);
4. the ability to **formulate** the **main ideas** of a text (formulate and combine appropriate semantic propositions);
5. the ability to **engage** in a range of **strategic processes** while reading more challenging texts (including goal setting, academic inferencing, monitoring);
6. the ability to **recognize discourse structuring and genre patterns**, and use this knowledge to support comprehension;
7. the ability to use **background knowledge** appropriately;

8. the ability to **interpret text meaning critically** in line with reading purposes;
9. the efficient use of **working memory** abilities;
10. the efficient use of **reading fluency skills**;
11. extensive amounts of **exposure to L2 print** (massive experience with L2 reading);
12. the ability to **engage in reading**, to expend effort, to persist in reading without distraction, and achieve some level of success with reading (reading motivation) (Grabe & Jiang, 2014, S. 188, Hervorhebung MB).

Diese kognitiven Prozesse werden in der Fachliteratur in Prozesse auf unteren und höheren Verarbeitungsebenen unterteilt (*low-level* und *high-level processes*). Diese Bezeichnungen sind metaphorisch zu verstehen und sollen nicht suggerieren, dass die Prozesse auf den unteren Verarbeitungsebenen einfacher als diejenigen auf den höheren Verarbeitungsebenen sind – der Unterschied zwischen ihnen ist, dass die Prozesse auf den unteren Verarbeitungsebenen stark automatisiert und damit unbewusst ablaufen können (vgl. u.a. Alderson et al., 2015; Grabe, 2000, 2009; Grabe & Stoller, 2011; Koda, 2005).

Bei den Prozessen auf den unteren Verarbeitungsebenen handelt es sich also um die grundlegenden „mechanischen“ Fertigkeiten, d.h. um „word recognition, syntactic parsing (using grammatical information), and semantic-proposition encoding (building clause-level meaning from word meanings and grammatical information)“ (Grabe, 2009, S. 22). Diese Prozesse werden als Teil des Arbeitsgedächtnisses (*working memory*)¹⁴ ausgeführt und ihr reibungsloser Ablauf ist unabdingbar für den Lesefluss sowie das Ermöglichen des Textverständnisses (ebd.). Das Erkennen der Wörter spielt dabei eine besonders wichtige Rolle (ebd.): Studien, die den Prozentsatz des für das Verstehen schriftlicher Texte notwendigen Vokabulars untersuchten, zeigten, dass die Lesenden dafür mit ca. 98% der Wörter im gelesenen Text vertraut sein müssen (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011, 2017).

Das Verständnis eines Textes ist aber viel mehr als das Verständnis der einzelnen Wörter und/oder der Syntax einzelner Sätze: Die Prozesse auf den höheren Verarbeitungsebenen führen dazu, dass der Text als Ganzes verstanden wird. Gemeint sind damit nicht nur die Entnahme von expliziten Informationen aus dem Text (sog. Textbasis, Grabe, 2009, S. 40-43), sondern auch die Inferenzen und Schlussfolgerungen der Lesenden, die aufgrund der Interaktion zwischen dem Text und ihrem eigenen Hintergrundwissen entstehen (sog.

¹⁴ Das Arbeitsgedächtnis ist als (metaphorischer) Rahmen zu verstehen, „in which cognitive processing and knowledge resources are integrated for comprehension“ (Grabe, 2009, S. 22, für mehr Informationen dazu vgl. ebd., S. 32-36).

Situationsmodell, Grabe, 2009, S. 43–46; vgl. auch Alderson et al., 2015; Kintsch, 1998; Perfetti, 1999).

Die Prozesse auf den unteren und höheren Verarbeitungsebenen laufen simultan ab und interagieren zu manchen Zeitpunkten auch miteinander (Grabe, 2009, S. 21–22). Sie werden oft je nach Verarbeitungsrichtung als *bottom-up*-Prozesse (d.h. aufsteigende Prozesse) und *top-down*-Prozesse (d.h. absteigende Prozesse) bezeichnet (ebd., S. 55). Mittlerweile ist man sich darüber einig, dass für den Leseverstehensprozess – entgegen einiger früherer Konzeptualisierungen – sowohl die Prozesse auf den unteren wie auch auf den höheren Verarbeitungsebenen nötig sind und dass die früher postulierte Unterteilung der Lesenden in *top-down readers* und *bottom-up readers* längst überholt ist (ebd., S. 21, 55; vgl. auch u.a. Alderson, 2000; Khalifa & Weir, 2009).

Die konkrete Anwendung der oben aufgelisteten Faktoren beim Lesen eines Textes hängt aber nicht nur von den Kompetenzen, der Motivation und/oder den eingesetzten Strategien der Lesenden ab, sondern wird auch von der zu bewältigenden Aufgabe, der Zielsetzung, sowie den Texteigenschaften beeinflusst (Grabe & Jiang, 2014, S. 188; vgl. auch Bachman & Palmer, 1996, 2010; Alderson, 2000; Grabe, 2009). Die geschriebenen Texte können nämlich von unterschiedlichen Lesenden auf unterschiedliche Arten und zu verschiedenen Zwecken rezipiert werden (vgl. Urquhart & Weir, 1998). In bestimmten Situationen und beim Rezipieren bestimmter Textsorten wie z.B. beim Orientieren am Flughafen oder beim Sichten von Ergebnissen einer Suchmaschine macht es normalerweise wenig Sinn, alle schriftlich vorhandenen Informationen ganz genau zur Kenntnis zu nehmen: Hier ist vielmehr eine schnelle, selektive und effiziente Vorgehensweise gefragt (vgl. Anthony Green, 2014, S. 99). Diese wird als *expeditious reading* bezeichnet und umfasst *scanning* (Identifizieren eines bestimmten Elements auf der lokalen Ebene, z.B. eines Wortes, einer Phrase oder einer Zahl), *skimming* (schnelles und auf der globalen Ebene ablaufendes Verschaffen eines Überblicks über das Hauptthema des Textes, auch *reading to obtain the gist*) und *search reading* (Lesen, um vorbestimmte Informationen zu gewissen Themen zu lokalisieren, was sowohl auf der lokalen wie auf der globalen Ebene ablaufen kann) (Urquhart & Weir, 1998; Khalifa & Weir, 2009). In anderen Situationen, wie beim Lesen zum Lernen oder beim Lesen eines Handbuches, ist hingegen *careful reading* gefragt, also ein sorgfältiges, nicht selektives und generell langsam und linear ablaufendes Lesen, das sowohl auf der lokalen Ebene (Wort-/Satzebene) wie auch auf der globalen Ebene (Textebene) ablaufen kann (vgl. Urquhart & Weir, 1998; Khalifa & Weir, 2009). Sehr oft werden auch verschiedene Herangehensweisen miteinander kombiniert – so wird z.B. bei der Lektüre eines Zeitungsartikels zuerst *skimming* eingesetzt, um zu bestimmen, ob in einem nächsten Schritt eine sorgfältige Lektüre

angebracht ist oder eher nicht (vgl. Urquhart & Weir, 1998; Khalifa & Weir, 2009; Anthony Green, 2014).

2.1.1.2 LESEVERSTEHEN IN DER L1 UND IN DER L2

Das Leseverstehen in der L2 (worunter hier nach Koda (2012, S. 229) die Situation verstanden wird, in der sich die Lernenden das Lesen in einer L2 aneignen, nachdem sie in ihrer L1 mit dem Lesen (und Schreiben) vertraut wurden) unterscheidet sich deutlich vom Leseverstehen in der L1 (vgl. u.a. Alderson, 2000; Alderson et al., 2015; Grabe, 2009; Grabe & Stoller, 2011; Koda, 2005). Gemäss Koda (2007, S. 1) handelt es sich beim Lesen in der L2 um einen sprachensübergreifenden und noch komplexeren Prozess als beim Lesen in der Erstsprache, weil dabei zwei Sprachen (L1 und L2) zum Einsatz kommen, die miteinander interagieren:

The dual-language involvement implies continual interactions between the two languages as well as incessant adjustments in accommodating the disparate demands each language imposes. For this reason, L2 reading is crosslinguistic and, thus, inherently more complex than L1 reading.

Die Unterschiede zwischen dem Lesen in der L1 und in der L2 machen sich bei L2-Lernenden auf allen Leistungsstufen bemerkbar, wobei sie vor allem bei Anfänger/innen und schwächeren L2-Leser/innen offensichtlich sind, weil ihnen das Verständnis der Bedeutung von L2-Wörtern Mühe bereitet (Grabe, 2009, S. 130).

Laut Grabe und Stoller (2011, S. 34) lassen sich die Unterschiede zwischen dem Lesen in der L1 und in der L2 in a) linguistische und verarbeitungsbezogene Unterschiede (*linguistic and processing differences*), b) individuelle und erfahrungsbezogene Unterschiede (*individual and experiential differences*) und c) soziokulturelle und institutionelle Unterschiede (*sociocultural and institutional differences*) unterteilen. Die linguistischen und verarbeitungsbezogenen Unterschiede zwischen den beiden Leseprozessen betreffen die jeweils unterschiedlich ausgeprägten Wortschatz-, Grammatik- und Diskurskenntnisse der Sprachbenutzer/innen am Anfang des Lesekompetenzerwerbs in der L2 im Vergleich zum Anfang des Lesekompetenzerwerbs in ihrer L1; das normalerweise in der L2 stärker ausgeprägte metalinguistische und metakognitive Bewusstsein; die sprachliche Distanz zwischen der jeweiligen L1 und L2; die L2-Kenntnisse der Lernenden; die Auswirkungen des Sprachtransfers sowie die interaktiven Einflüsse der beiden Sprachen auf den Leseprozess. Zu den individuellen und erfahrungsbezogenen Unterschieden zwischen den L1 und L2-Leser/innen gehören die Lesefähigkeiten der L2-Lernenden in ihrer L1; die Motivation und Einstellungen der Lernenden zum Lesen in der L2, die Menge der Kontaktmöglichkeiten mit

geschriebenen L1- und L2-Texten, die Genres der von L2-Lernenden rezipierten Texte sowie die sprachlichen Ressourcen wie Wörterbücher, Glossare und Erklärungen zu kulturellen Hintergründen, die den Lesenden in ihrer L1 und L2 zur Verfügung stehen. Der Erwerb von Lesekompetenz in der L2 wird darüber hinaus auch von einer Reihe ausserunterrichtlicher soziokultureller und institutioneller Aspekte beeinflusst, die je nach Ausgangs- und Zielsprache sehr unterschiedlich ausfallen und sich auf den L2-Leseprozess auswirken können. Dazu gehören beispielsweise der Stellenwert und die Bedeutung von schriftlichen Texten in der privaten und gesellschaftlichen Umgebung der L2-Lernenden, die Organisationsformen von Diskursen und Texten in der L1 und L2 oder die Erwartungen seitens der L1- und L2-Bildungseinrichtungen, die sich in Curricula, Lehrwerken, Prüfungen oder auch den Rollen der Lehrpersonen und Lernenden manifestieren (Grabe & Stoller, 2011, S. 36–57; vgl. auch Alderson, 2000; Grabe, 2009).

2.1.1.3 MODELL DES LESEVERSTEHENS NACH KHALIFA & WEIR (2009)

Mittlerweile wurden viele verschiedene Modelle zur Systematisierung und Beschreibung der Prozesse und Komponenten vorgeschlagen, die beim Leseverstehen zum Einsatz kommen (für einen umfassenden Überblick dazu siehe Grabe, 2009). Das Leseverstehensmodell, das im Sprachtestbereich eine hohe Akzeptanz genießt (vgl. z.B. Brunfaut & McCray, 2015; Brunfaut, 2016b; Kong, 2019; Hasselgreen & Caudwell, 2016; Bax, 2013) wurde von Khalifa und Weir (2009) entwickelt. In Anlehnung an dieses Modell kann auch die Modellierung des fremdsprachlichen Hörverstehens vorgenommen werden (Grotjahn, 2012, S. 75). Das Modell von Khalifa und Weir, das im Folgenden näher dargestellt wird, lag auch dem Lese- und Hörverstehenskonstrukt im Projekt IFB zugrunde.

Khalifa und Weir entwickelten ihr Modell zur Beschreibung des Leseprozesses in der L1 in Anlehnung an Forschung aus der Kognitionspsychologie. Der in ihrem Modell veranschaulichte Leseprozess in der L1 kann als Ziel betrachtet werden, auf das L2-Leser/innen schrittweise hinarbeiten (Khalifa & Weir 2009, S. 43). Das Modell besteht aus drei Hauptteilen, die in Abbildung 1 von links nach rechts dargestellt werden. Zu diesen Hauptteilen gehören: 1) eine metakognitive Komponente¹⁵ (*metacognitive activity*), 2) eine zentrale Verarbeitungskomponente (*central processing core*) und 3) eine Wissenskomponente (*knowledge base*). Diese drei Hauptteile schliessen ihrerseits jeweils weitere Unterprozesse mit ein. Khalifa und Weir ist es gelungen, in ihrem *model of reading*

¹⁵ Deutsche Termini werden hier nach Grotjahn (2012, S. 75) verwendet.

sowohl die kognitiven und metakognitiven Prozesse wie auch das Sprach- und Hintergrundwissen der Leser/innen darzustellen.

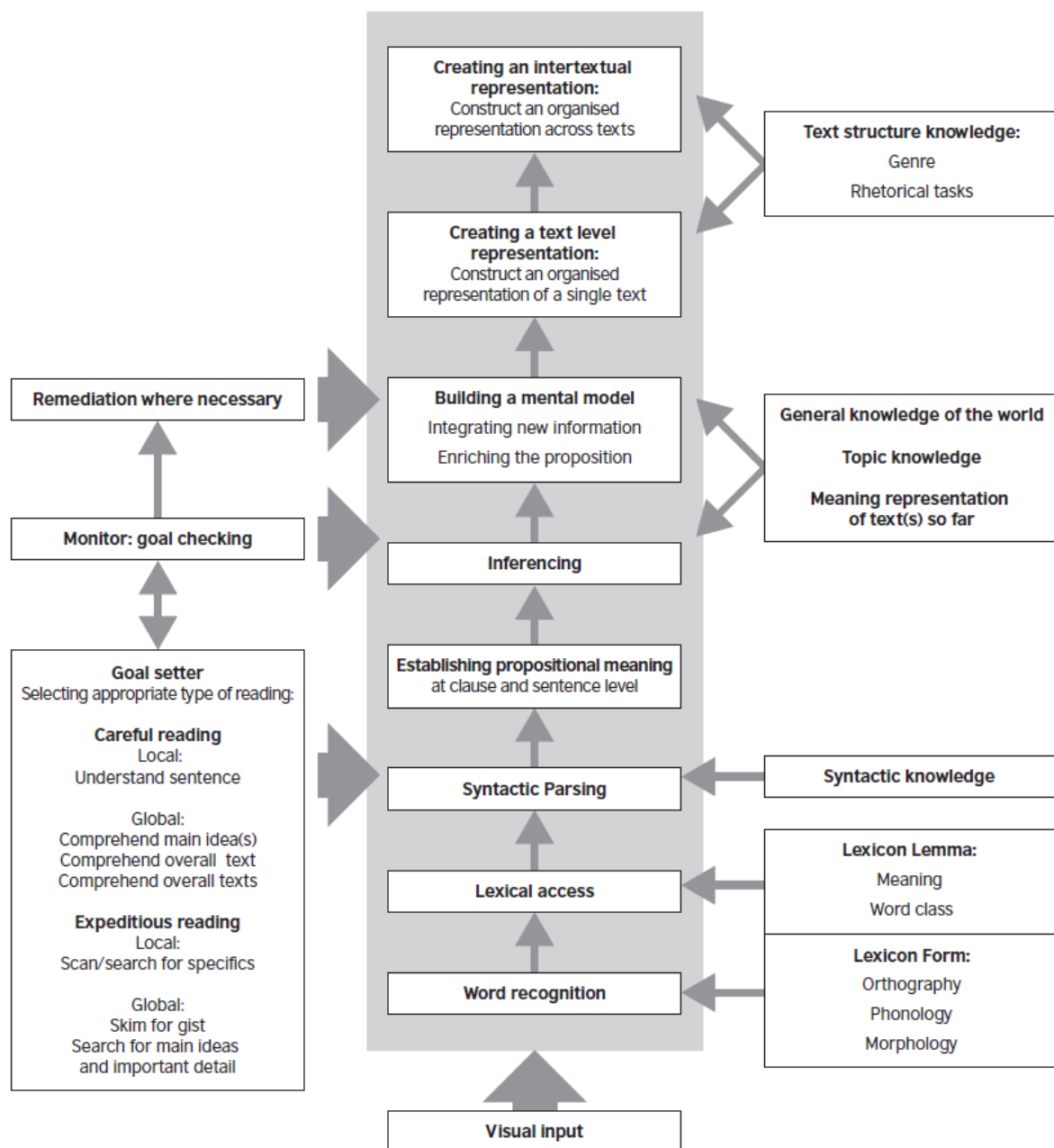


Abbildung 1: *Model of reading* von Khalifa & Weir 2009, S. 43.

Die metakognitive Komponente (dargestellt in der linken Spalte) umfasst in diesem Modell die Zielsetzung für die Leseaktivität (*goal setter*), die Überwachung der Leseaktivität (*monitor*) sowie – falls nötig – das Ergreifen von korrigierenden Massnahmen (*remediating*). Die Zielsetzung für die Leseaktivität ist laut Khalifa und Weir (2009) bestimmend, weil diese Entscheidung die jeweilige Gewichtung der Prozesse in der zentralen Verarbeitungskomponente des Modells beeinflusst. Das heisst, bei der Zielsetzung entscheiden die Leser/innen über den oder die *type(s) of reading* (vgl. auch Urquhart & Weir,

1998), die sie einsetzen wollen, um die jeweilige Aufgabe zu lösen. Die Leser/innen legen also fest, ob sie den Text sorgfältig lesen wollen, um alle Informationen zu erfassen (*careful reading*); oder ob sie ihn eher zügig überfliegen, um nur auf bestimmte Informationen aus dem Text zuzugreifen (*expeditious reading*). Auch entschieden sie, ob das Lesen auf der lokalen, d.h. wort- oder satzbezogenen Ebene stattfindet, oder eher auf der globalen Ebene, die auf den ganzen Text ausgerichtet ist. Weiter kontrollieren die Leser/innen während des Lesens, ob die Aktivität mit den gesetzten Zielen übereinstimmt und korrigieren ihr Verhalten bei Bedarf.

Die zentrale Verarbeitungskomponente (mittlere Spalte) besteht aus acht aufeinander aufbauenden kognitiven Prozessen, deren Zusammenspiel zum Textverständnis führt (vgl. Kap. 2.1.1.1.). Dazu gehören die Prozesse auf den unteren Verarbeitungsebenen, die stark automatisiert werden können (d.h. *word recognition*, *lexical access*, *syntactic parsing* und *establishing propositional meaning*), wie auch die Prozesse auf den höheren Verarbeitungsebenen (d.h. *inferencing*, *building a mental model*, *creating a text level representation* und *creating an intertextual representation*) (Khalifa & Weir, 2009; vgl. auch Grabe, 2009).

Zur Wissenskomponente (*knowledge base*, rechte Spalte) gehören laut Khalifa und Weir (2009) die verschiedenen Wissensquellen, die den Lesenden beim Kontakt mit schriftlichen Texten zur Verfügung stehen und die gemäss diesem Modell auf verschiedene Aspekte des zentralen Verarbeitungskerns verweisen. Diese Komponente „is drawn upon by the central processing core in line with the intended purpose and the performance conditions established by the task“ (ebd., S. 47). Sie besteht aus den Kenntnissen der Form und der Bedeutung der Wörter sowie der Syntax, dem allgemeinen und themenspezifischen Weltwissen, der bisher etablierten Bedeutung des Textes und den Kenntnissen der Textstruktur, also z.B. des Genres.

2.1.1.4 LESEVERSTEHEN ÜBERPRÜFEN UND BEWERTEN

Wie bereits unter 2.1 angemerkt ist es bei der Erstellung von Sprachtests zentral, das Testkonstrukt zu definieren, d.h. festzulegen, welche Kenntnisse und Fähigkeiten der Testteilnehmenden mit welchen Testaufgaben zu messen sind. Eine der grössten Herausforderungen beim Testen von Leseverstehenskompetenz ergibt sich laut Grabe und Jiang (2014, S. 195) aus der Komplexität des Leseverstehenskonstrukts. Da das Verstehen von geschriebener Sprache vielschichtig ist und viele Teilkomponenten umfasst (vgl. Kap. 2.1.1.1 bis 2.1.1.3), ist es ein schwieriges Unterfangen, dieses breite Konstrukt in standardisierten Tests angemessen abzudecken (Grabe, 2009, S. 357; Grabe & Jiang, 2014,

S. 195). Hinzu kommen weitere kontextbezogene Einschränkungen: „In standardized assessment contexts, practices that might expand the reading assessment construct are constrained by concerns of validity, reliability, time, cost, usability, and consequence, which limit the types of reading assessment tasks that can be used“ (Grabe & Jiang, 2014, S. 195-196). In standardisierten Leseverstehenstests wird das Konstrukt auf vielfältige Weise beschrieben, d.h. in Bezug auf die verschiedenen Lesezwecke, auf repräsentative Leseaufgaben oder auf die kognitiven Prozesse, die das Verstehen unterstützen (ebd., S. 188–189). Laut Grabe und Jiang (ebd., S. 197) ist jedoch davon auszugehen, dass das Lesekonstrukt in allen bekannten standardisierten Tests unterrepräsentiert ist.

Mit den Resultaten der standardisierten Testaufgaben werden, wie ebenfalls bereits erwähnt, die Produkte (und nicht die Prozesse) des Leseverstehens erfasst. Bachman und Palmer (2010, S. 78) unterscheiden zwei Antworttypen (*type of response*), die durch den Einsatz von Testaufgaben elizitiert werden können: *selected response* und *constructed response*. Bei *selected-response*-Aufgaben, zu denen Multiple-Choice- oder Matching-Aufgaben (d.h. Zuordnungsaufgaben) gehören, wird die Antwort auf die Frage aus mindestens zwei vorgeschlagenen Optionen ausgewählt. Bei den *constructed-response*-Aufgaben wie Kurzantwortfragen (*short answer questions*) müssen die Testteilnehmenden ihre Antworten erzeugen bzw. konstruieren. Alle diese Antwortformate weisen bestimmte Vor- und Nachteile auf (vgl. u.a. Alderson, 2000; Anthony Green, 2014), die es bei der Testkonstruktion jeweils zu berücksichtigen und abzuwägen gilt. Auch ist dabei zu beachten, dass sich die gewählten Antwortformate unterschiedlich auf das zu untersuchende Konstrukt auswirken (vgl. u.a. Jamieson et al., 2008; Lenz et al., 2019; Ozuru et al., 2013; Rupp et al., 2006; Shohamy, 1984).

In standardisierten Leseverstehenstests (und auch in standardisierten Hörverstehenstests) werden oft *selected-response*-Aufgaben verwendet, wie sie auch im Projekt IFB zum Einsatz kamen. Zwar ist die Produktion solcher Aufgaben „a very skilled and time-consuming business“ (Alderson, 2000, S. 212), sie können aber bei der Testdurchführung einfach und kostengünstig eingesetzt und bewertet werden (vgl. Stevenson, 2015, S. 325). Kritik an diesem Antwortformat bezieht sich u.a. auf die Möglichkeit, die richtige Antwort aus einem „falschen Grund“ zu wählen (d.h. ohne den Text verstanden zu haben) oder umgekehrt die falsche Antwort trotz eines ausreichenden Textverständnisses anzukreuzen (vgl. Alderson, 2000, S. 211–212). Die Validität von *selected-response*-Aufgaben kann aber durch das Beachten bestimmter Regeln bei ihrer Konstruktion erhöht werden (vgl. z.B. Anthony Green, 2014, S. 112–114). Daneben ist es wichtig, alle Testaufgaben im Voraus zu erproben und bei Bedarf zu überarbeiten (vgl. Kap. 2.2.4.1).

Zu den Faktoren, die zur konstruktirrelevanten Varianz eines L2-Leseverstehenstests (oder auch eines Hörverstehenstests) beitragen können, gehört u.a. die Wahl der Sprache, in der die Instruktionen, Fragen zu den Inputtexten und die Antwortoptionen zu formulieren sind. Werden diese in der L2 verfasst, besteht die Möglichkeit „that some learners pursue the task without understanding what they are being asked to do“ (Koda, 2012, S. 230). Diese Vermutung wird durch die Befunde des Projekts „Task Lab“ gestützt, ein Vorgänger des Projekts IFB, in dem die Aufgaben zur Erfassung der Lesekompetenzen von jungen Französischlernenden untersucht wurden (Karges, Barras et al., 2021; vgl. auch KFM, 2020b). Aufgrund der klaren Ergebnisse der qualitativen Daten aus „Task Lab“ wurde von den Autor/innen (jedenfalls für Fremdsprachenlernende auf tiefem Sprachkompetenzniveau und in lokalen Kontexten) die Formulierung der Fragen und Antworten in der Schulsprache empfohlen: Dank dem Einsatz der Schulsprache konnte nämlich zuverlässiger gemessen werden, was gemessen werden sollte, d.h. das Verständnis der fremdsprachlichen Texte und nicht etwa der Instruktionen oder Antwortoptionen (Barras et al., 2016).

In kommunikativen Sprachtests werden Aufgaben eingesetzt, die soweit wie möglich den Sprachgebrauch der Testteilnehmenden im „realen Leben“ nachahmen wollen. Eine in diesem Kontext oft diskutierte Frage betrifft die Authentizität der Inputtexte. Die im Sprachtestbereich weit verbreitete Auffassung von Authentizität orientiert sich an der Definition von Bachman und Palmer (1996, S. 23), die darunter „the degree of correspondence of the characteristics of a given language test task to the features of a TLU¹⁶ [Target Language Use] task“ verstehen. Im Fall des schulischen Fremdsprachenunterrichts wird die relevante *target language use domain* allerdings oft mit „the classroom simulating life outside“ (Hasselgreen & Caudwell, 2016, S. 53) umschrieben.

Wenn aber die Testteilnehmenden einen Text während einer (standardisierten) Prüfung lesen, sind sie sich durchaus bewusst, dass sie für Prüfungszwecke lesen, und nicht beispielsweise um neue Wörter oder Inhalte zu lernen (vgl. u.a. A. D. Cohen & Upton, 2006; Rupp et al., 2006). Grabe und Jiang (2014, S. 196) kommen daher zu dem Schluss, dass es wohl nicht realistisch sei, zu verlangen, dass die Textrezeption während eines Tests die Erfahrungen in der „realen Welt“ exakt widerspiegelt (vgl. auch z.B. Taylor, 2013, S. 25). Es sei aber wichtig, die Testteilnehmenden mit realistischen Texten, Aufgaben und Kontexten zu konfrontieren, da dies einen positiven Washback-Effekt auf den Sprachunterricht haben kann (Grabe & Jiang, 2014, S. 196): Werden in Sprachtests realweltliche Aufgaben gestellt, ist

¹⁶ Die sog. *target language use domain* definieren Bachman und Palmer als „situation or context in which the test taker will be using the language outside of the test itself“ (1996, S. 18).

davon auszugehen, dass im Rahmen der Vorbereitung auf solche Sprachtests die realweltlichen Aufgaben auch im Sprachunterricht vermehrt eingesetzt werden.

2.1.2 HÖRVERSTEHEN

In diesem Unterkapitel wird das Hörverstehen ¹⁷ näher beleuchtet, wie zuvor beim Leseverstehen ebenfalls mit Fokus auf die Sprachtestforschung. Da es sich sowohl beim Hörverstehen als auch beim Leseverstehen um Verstehensprozesse handelt, lässt sich vieles von dem, was oben bereits über das Leseverstehen zusammengefasst wurde, auch auf das Hörverstehen übertragen. In den folgenden Unterkapiteln werden neben den Ähnlichkeiten vor allem auch die Unterschiede zwischen den beiden Prozessen aufgezeigt. Für eine umfassendere Diskussion muss hier aber, wie schon beim Leseverstehen, auf die einschlägige Literatur verwiesen werden (u.a. Buck, 2001; Field, 2008, 2019; Rost, 2011; Vandergrift & Goh, 2012).

2.1.2.1 HÖRVERSTEHEN ALS PROZESS

Ähnlich wie das Leseverstehen ist auch das Hörverstehen ein sehr komplexer Prozess (vgl. u.a. Buck, 2001; Field, 2008), dem jedoch lange wenig Beachtung geschenkt wurde, sowohl in der Forschung als auch im Kontext der Überprüfung von L2-Kompetenzen (vgl. u.a. Buck, 2001; Grotjahn, 2012; Vandergrift, 2007, 2015). Dies überrascht angesichts der Tatsache, dass das Hörverstehen in kommunikativen Situationen etwa die Hälfte der Zeit in Anspruch nimmt (Ockey, 2012, S. 212; Wagner, 2014, S. 47) und dass es sowohl empirische wie auch theoretische Anhaltspunkte dafür gibt, dass der L2-Lernprozess durch die frühe Förderung des Hörverstehens erheblich unterstützt werden kann (Buck, 2018, S. XI). Dass das Hörverstehen bisher am wenigsten verstanden und erforscht wurde, liegt laut Vandergrift an der Schwierigkeit des Untersuchungszugangs: „The covert nature of the process and ephemeral nature of the input make the perceptual and interpretation processes involved in listening comprehension difficult to access“ (2015, S. 299). In der letzten Zeit kann jedoch eine erfreuliche Zunahme an Studien und praxisorientierten Beiträgen zum L2-Hörverstehen, seiner Vermittlung und Beurteilung beobachtet werden, auch wenn diesbezüglich noch lange nicht alle Fragen geklärt werden konnten (Vandergrift, 2015, S. 299; vgl. auch u.a. Vandergrift & Goh, 2012; Grotjahn, 2012; Shafer, 2018; Holzknecht, 2019; Rossa, 2012; Field, 2019; R. Green, 2017; Ockey & Wagner, 2018a; Geranpayeh & Taylor, 2013; Vandergrift & Baker, 2015, 2018; Rost, 2011; Brunfaut, 2016a; Holzknecht et al., 2017).

¹⁷ Das Hör-Sehverstehen (vgl. Porsch et al., 2010) ist für die hier beschriebene Untersuchung unerheblich und bleibt daher aus Platzgründen ausgeklammert.

Das Hörverstehen wird im Bereich der Sprachtestforschung nach Buck (2001, S. 114) verstanden als:

the ability

- to process extended samples of realistic spoken language, automatically and in real time;
- to understand the linguistic information that is unequivocally included in the text; and,
- to make whatever inferences are unambiguously implicated by the content of the passage.

Dieser Prozess wird also als ein Zusammenspiel physiologischer und kognitiver Prozesse auf unterschiedlichen Verarbeitungsebenen sowie kontextueller Einflussfaktoren konzeptualisiert (Vandergrift, 2015, S. 299). Wie auch beim Leseverstehen lassen sich die Ergebnisse des Hörverstehens nicht direkt beobachten, sondern müssen aus den Reaktionen der Hörer/innen abgeleitet werden.

Das Hörverstehen kann laut Grotjahn (2012, S. 75) in Anlehnung an das in Kap. 2.1.1.3 vorgestellte Modell des Leseverstehens von Khalifa und Weir (2009) modelliert werden (für andere Modellierungsversuche vgl. u.a. Field, 2008, 2013, 2019; Rost, 2011; Vandergrift, 2011; Vandergrift & Goh, 2012). Wie beim Leseverstehen können beim Hörverstehen die aufsteigenden *bottom-up*- und die absteigenden *top-down*-Prozesse unterschieden werden. Bei den *bottom-up*-Prozessen werden kleinere Elemente zu umfassenderen Einheiten zusammengefügt (z.B. lautliche Segmente zu Silben und Wörtern), um das Verständnis einer Äußerung zu ermöglichen; bei den *top-down*-Prozessen wird hingegen die Verarbeitung auf den unteren Stufen durch die hierarchiehöheren Komponenten (z.B. kontextuelle Hinweise oder Kenntnisse des Themas) beeinflusst (Field, 2008, S. 132-133; Grotjahn, 2012, S. 75; Vandergrift, 2011, S. 456). Das (fremdsprachliche) Hörverstehen wird dabei als „ein interdependentes Wechselspiel von aufsteigender und absteigender Verarbeitung“ angesehen (Grotjahn, 2012, S. 75; vgl. auch Field, 2008, S. 132-137): Für ein erfolgreiches Hörverstehen wird ein geschicktes Zusammenspiel der beiden Prozesse vorausgesetzt. Ob die Hörer/innen jedoch dabei mehr die aufsteigenden oder die absteigenden Prozesse nutzen, hängt vom Zweck des Zuhörens bzw. von der Aufgabe, von den Eigenschaften der Lernenden (wie z.B. ihre Sprachkompetenz oder die Kapazität ihres Arbeitsgedächtnisses) und vom Kontext ab (Vandergrift, 2011, S. 456). Analog zu den *types of reading* wird in der Fachliteratur u.a. zwischen dem detaillierten, selektiven, globalen, orientierenden, kursorischen oder totalen Hören unterschieden; auch wird in Testspezifikationen oft zwischen Hörzielen bzw. Höraktivitäten zum Verstehen von

Kernaussagen, Hauptaussagen und relevanten Details unterschieden (Grotjahn, 2012, S. 81; vgl. auch Anthony Green, 2014, S. 103).

Neben vielen Ähnlichkeiten lassen sich auch grundlegende Unterschiede zwischen dem Hör- und Leseverstehen identifizieren. Weil das Hörverstehen in Echtzeit passiert (Buck, 2001, S. 114) und der auditive Eindruck als flüchtig zu bezeichnen ist, kann falsch Verstandenes beim weiteren Hören schwieriger zu korrigieren sein (Grotjahn, 2012, S. 76). Auch neigen die Hörenden dazu, sich bei der Rezeption längerer Texte relativ früh auf die wesentlichen Informationen und weniger auf die Details zu konzentrieren (ebd.). Weitere entscheidende Unterschiede zwischen dem Lese- und Hörverstehen listet Field (2008, S. 27–28) wie folgt auf:

A reader has the advantage of a standardised spelling system. By contrast, a listener is exposed to speech sounds which vary considerably from one utterance to another, and from one speaker to another, and which even blend into each other. A reader benefits from blank spaces between each of the words in a written text. There are no such regular gaps separating words in connected speech; and the listener largely has to decide for herself where one word ends and the next begins. [...] the signal which the listener has to deal with requires an entirely different kind of processing to that demanded by reading. [...] Of course, it cannot be denied that the two skills have certain meaning building elements in common. They both draw upon the same comprehension processes (extracting ideas, relating the ideas to what has gone before, interpreting what the speaker/reader has left unsaid, making connections to world knowledge). But this resemblance should not be overstated. Because of the temporary nature of the speech signal, a listener has to carry forward in her memory all the ideas that have been expressed so far if she wishes to build a complete account of a conversation. By contrast, [...] a reader can always look back. So even at the level of comprehension the processes are distinct.

Das Spezielle am Hörverstehen ist, wie schon im obigen Zitat angedeutet, dass die Geschwindigkeit der Sprachverarbeitung (anders als beim Leseverstehen, Sprechen oder Schreiben) nicht von den Sprachnutzer/innen bestimmt wird (Buck, 2018, S. XI). Die Hörer/innen müssen die Sprache normalerweise in der Geschwindigkeit verarbeiten, in der sie von der/dem Sprechenden produziert wird, was auf der Seite der Rezipient/innen eine Automatisierung der Sprachverarbeitung voraussetzt (ebd.). Neben guten sprachlichen Basiskenntnissen werden dafür auch weitere spezifische Fähigkeiten gebraucht, die weder beim Lesen noch beim Schreiben oder Sprechen zum Einsatz kommen (ebd.). Dazu gehören laut Buck (ebd., S. XII) 1) Kenntnisse der Phonetik, Phonologie und Intonation; 2) die Fähigkeit, die in spontanen mündlichen Äusserungen oft anzutreffenden „Fehler“ und

Füllelemente von den bedeutungsvollen Inhalten zu trennen; und 3) die Fähigkeit, mit gesprochener Sprache umzugehen, die sich von ihrer geschriebenen Variante u.a. in Bezug auf Wortschatz, Grammatik, Satzlänge oder Diskursstruktur erheblich unterscheiden kann. Neben diesen sprachbezogenen Fähigkeiten und Kenntnissen müssen die Hörenden ausserdem, wie bereits erwähnt, auch den Kontext richtig interpretieren und ihr allgemeines Wissen in geeigneter Art und Weise nutzen können (vgl. ebd., S. XII-XIII).

2.1.2.2 HÖRVERSTEHEN IN DER L2

Die in das Hörverstehen involvierten Prozesse laufen in der L1 simultan, automatisch und oft auch ohne ihre bewusste Wahrnehmung ab (vgl. R. Green, 2017, S. 3). Ganz anders sieht es beim Hörverstehen in der L2 aus, vor allem dann, wenn sich der oder die Hörende erst am Anfang seines/ihres Lernprozesses befindet:

Listeners learning a new language [...] have limited linguistic knowledge; therefore, less of what they hear can be automatically processed. Depending on their level of proficiency or the difficulty of the text or task, these listeners may need to consciously process some of the input and, given the limitations of working memory, comprehension either turns out to be incomplete or breaks down (Vandergrift, 2015, S. 300).

L2-Hörer/innen können versuchen, ihre Verständnislücken auszugleichen, indem sie auf metakognitive und kognitive Strategien, kontextbezogene Hinweise und andere relevante Informationen zurückgreifen (ebd.). Gerade bei Hörer/innen im Anfängerstadium kann jedoch die Kompensation grösserer Defizite in der Sprachverarbeitung auf den unteren Verarbeitungsebenen durch übermässigen Verlass auf *top-down*-Prozesse zu einem inadäquaten Hörverständnis führen (Grotjahn, 2012, S. 76; vgl. auch Buck, 2001, S. 50–51).

Field (2019, S. 21–22) identifiziert insgesamt vier Faktoren, die zur Gesamtkompetenz von L2-Hörer/innen beitragen. Dazu gehören neben den bereits oben angesprochenen 1) Sprachkenntnissen (d.h. Kenntnisse der Wörter und die Fähigkeit, ihre gesprochene Form zu erkennen, sowie Grammatikkenntnisse und die Fähigkeit, die syntaktischen Strukturen von gesprochenen Äusserungen während des Hörverstehens zu erfassen) und 2) der strategischen Kompetenz von L2-Hörer/innen (d.h. ihrer Fähigkeit, mit Verstehenslücken umzugehen) auch 3) kulturelles Wissen (d.h. die Fähigkeit der Hörenden, die Intentionen der Sprechenden unter Berücksichtigung der Pragmatik, Stilmerkmale und Intonation der L2 zu erkennen) sowie 4) die „Expertise“ der L2-Lernenden (d.h. der Automatisierungsgrad ihres Hörverstehensprozesses in der Zielsprache und ihre Fähigkeit, wiederkehrende sprachliche Chunks zu erkennen) (vgl. ebd.).

2.1.2.3 HÖRVERSTEHEN ÜBERPRÜFEN UND BEWERTEN

Entscheidungen, die während der Entwicklung eines Hörverstehenstests zu treffen sind, hängen – wie bei der Überprüfung aller anderen sprachlichen Kompetenzen – vom Testzweck, von den Eigenschaften der Testteilnehmenden und von der Definition des zu messenden Testkonstrukts ab (Ockey, 2012, S. 213; mit Verweis auf Bachman & Palmer, 2010). Bei der Festlegung der Testaufgaben und ihrer Spezifikationen ist auch im Falle des Hörverstehens zu beachten, dass und wie sich die Eigenschaften der Testaufgaben und des Kontextes auf das auswirken, was gemessen wird (vgl. u.a. Brunfaut, 2016a; Buck, 2001; Buck & Tatsuoka, 1998; Anthony Green, 2014; Holzknacht, 2019; Ockey, 2012; Révész & Brunfaut, 2013; Taylor, 2013; Wagner, 2014). Die Schwierigkeit von Hörverstehensaufgaben wird u.a. beeinflusst durch Akzent, Dialekt und Sprechtempo der sprechenden Person(en); die Anzahl der Präsentationen des Inputtextes (normalerweise einmal oder zweimal); die Art und Weise, in der die Items präsentiert werden (vor, während oder nach dem Hören); oder auch die Begleitung des Hörtextes durch einen Videostimulus (vgl. z.B. Brunfaut, 2016a für einen umfassenden Überblick).

Auch wenn das Hörverstehen als Verstehensprozess viele wichtige Merkmale mit dem Leseverstehen teilt, ist die Entwicklung und Durchführung von standardisierten Hörverstehenstests dennoch komplizierter und technisch sowie zeitlich aufwändiger als im Falle von Leseverstehenstests (Buck, 2001, S. 31–32). Deshalb soll laut Buck überlegt werden, ob und wann sich die Mühe überhaupt lohnt: Geht es darum, das allgemeine Sprachverständnis der Testteilnehmenden zu beurteilen, kann dies durch die Überprüfung von Leseverstehenskompetenzen einfacher gemacht werden (ausser natürlich, Lesen gehört nicht zu den Kompetenzen der Getesteten). Die Überprüfung des Hörverstehens ist hingegen angezeigt, wenn man an den Kenntnissen, Fertigkeiten und Fähigkeiten interessiert ist, die für das Hörverstehen charakteristisch sind (ebd., S. 32; für weitere Diskussion dazu vgl. auch Field, 2019; Grotjahn, 2012; Ockey & Wagner, 2018a). Als wichtigen Grund für die Überprüfung der Hörverstehenskompetenz nennt Buck ausserdem den positiven Washback-Effekt auf den Unterricht, denn was getestet wird, muss zuerst gelehrt werden (vgl. auch Wagner, 2014, S. 59–60). Im Falle des Hörverstehens sollte es sich aber um die Überprüfung der Kenntnisse, Fähigkeiten und Fertigkeiten handeln, „which are unique to real-world listening, in order to encourage teachers to teach learners to comprehend realistic spoken language in a realistic manner“ (Buck 2001, S. 32).

In den letzten Jahrzehnten wurde das Hörverstehen neben dem Leseverstehen, Schreiben und Sprechen hauptsächlich als isolierte Zielfertigkeit überprüft (vgl. Brunfaut, 2016a, S. 100), u.a. aus Gründen der Reliabilität und Validität (vgl. Wagner, 2014, S. 50–51). Die Ergebnisse

des Hörverstehensprozesses lassen sich allerdings, wie erwähnt, nicht direkt beobachten. Damit das Verstandene sichtbar wird, beinhalten Hörverstehensaufgaben daher immer auch etwas Lesen, Schreiben und/oder Sprechen, wobei Testentwickler/innen die daraus entstehende konstruktirrelevante Varianz möglichst klein zu halten versuchen (Brunfaut, 2016a, S. 100). So wird beispielsweise empfohlen, die Testaufgaben für junge Lernende in lokalen Kontexten in der L1 zu formulieren, weil so die Items besser verstanden werden können und die Lernenden auch weniger versucht seien, bestimmte Strategien zum „Austricksen“ des Tests anzuwenden (Field, 2019, S. 108).

Beim Testen des Hörverstehens als isolierte Fertigkeit hören die Testteilnehmenden normalerweise einen authentischen, semiauthentischen oder speziell für die Testzwecke kreierten Inputtext (vgl. Ockey & Wagner, 2018b, S. 5) und beantworten Fragen dazu. Als Antwortformate werden in diesem Fall oft 1) *selected-response*-Aufgaben wie Multiple-Choice oder Matching mit schriftlich bzw. visuell präsentierten Antwortoptionen oder 2) *constructed-response*-Aufgaben wie Notizen, *gap-filling*-Tasks, Satzvervollständigungsaufgaben oder Kurzantwortfragen verwendet, wobei die Antworten auf die *constructed-response*-Aufgaben in schriftlicher oder mündlicher Form vorgesehen werden können (Brunfaut, 2016a, S. 100). Es wird aber verschiedentlich hervorgehoben, dass das Hörverstehen im realen Leben selten isoliert stattfindet: „[I]n most real-life listening domains, listeners must listen to and process oral information, and then immediately do something with that information“ (Wagner, 2014, S. 51; siehe auch Ockey & Wagner, 2018b). Um den Zielsprachengebrauch ausserhalb der Testsituation exakter abzubilden, werden in manchen Tests sog. integrierte Testaufgaben eingesetzt, deren erfolgreiche Bewältigung neben dem Hörverstehen auch vom Sprechen und/oder Schreiben abhängig ist; was jedoch seinerseits wiederum andere Herausforderungen bezüglich Praktikabilität, Reliabilität und Validität mit sich bringt (vgl. Brunfaut, 2016a; Ockey & Wagner, 2018a; Rukthong & Brunfaut, 2020; Wagner, 2014). Field (2019, S. 79–80) empfiehlt deshalb, in einem Test mehrere Aufgabenformate anzuwenden, damit 1) die Testteilnehmenden nicht routinemässig auf bestimmte *test-wiseness*-Strategien (siehe Kap. 2.2.3) zurückgreifen, 2) Testformate, die sich auf kognitive Fähigkeiten stützen, die nichts mit dem Hörverstehen zu tun haben (z.B. eine Entscheidungsfindung oder das Behalten von mehreren Optionen im Gedächtnis), nicht bestimmte Testteilnehmende bevorzugen oder benachteiligen, und auch damit 3) das gewählte Format auf das jeweils anvisierte Kompetenzniveau abgestimmt werden kann, wodurch das Problem der begrenzten L2-Kenntnisse auf den unteren Niveaus und die damit verbundenen Schwierigkeiten bei der Rezeption der schriftlichen Items (oder Produktion der Antworten) angegangen werden kann.

Trotz der zunehmenden Anzahl an Forschungsarbeiten bleiben viele Fragen bezüglich der reliablen und validen Überprüfung des L2-Hörverstehens immer noch offen (Buck, 2018, S. XVI; Wagner, 2014, S. 59). Vor allem praktische Aspekte stellen nach wie vor eine Herausforderung dar, wie Buck (2018, S. XVI) anmerkt: „The theory is pretty well understood, but as a field, we just do not know enough to implement that in a practical manner“.

2.1.3 NEUERE ENTWICKLUNGEN

In diesem Teil werden zwei neuere Entwicklungen aus dem Bereich der (Sprach-)Testforschung vorgestellt, die für die Überprüfung der Lese- und Hörverstehenskompetenzen der L2-Lernenden von Bedeutung sind und in der vorliegenden Arbeit eine zentrale Rolle spielen. Es handelt sich dabei um den Einsatz von digitalen Textsorten in den Testaufgaben sowie um das ursprünglich für die Überprüfung von L1-Lesekompetenzen entwickelte *Scenario-Based Assessment*.

2.1.3.1 DIGITALE TEXTSORTEN

Durch die Verbreitung von Computern, Tablets und Smartphones gewannen das Lese- und Hörverstehen neue Dimensionen. Flüchtige Hörtexte und gedruckte Lesetexte werden im Alltag durch neue, digitale (d.h. primär für das Lesen oder Hören auf digitalen Geräten vorgesehene) Textsorten wie Podcasts, Chats, Hypertext oder Tweets ergänzt und in manchen Fällen sogar ersetzt. Dies hat Auswirkungen u.a. auf den (Fremdsprachen-)Unterricht und damit auch auf die Beurteilung der Sprachkompetenzen der Lernenden (vgl. z.B. Ockey, 2012, S. 215; OECD, 2019; Taylor, 2013, S. 23–24; Wagner, 2014, S. 60).

Die Wichtigkeit digitaler Textsorten im realweltlichen Lesen hat auch Auswirkungen auf die Beurteilung von Leseverstehenskompetenzen und wird entsprechend in modernen Leseverstehenskonstrukten berücksichtigt. So beinhaltet die in der internationalen Schulleistungsstudie PISA verwendete Definition von *reading literacy* seit 2009 Verweise auf digitale Textsorten und wird laufend aktualisiert und erweitert, um die Fähigkeiten besser zu erfassen, die für das Lesen und die Interaktion mit geschriebenen, zunehmend digitalen Texten erforderlich sind (OECD, 2019). Darüber hinaus wurde in der letzten grossen PISA-Studie die Bandbreite der verwendeten Texte u.a. um digitale Textsorten erweitert. Neben sogenannten „statischen“ Texten (z.B. PDF-Dateien, die zum Ausdrucken vorgesehen waren, aber auf einem Computerbildschirm angezeigt werden), gehören dazu „dynamische“ Texte mit innovativen Merkmalen, die die Interaktionsmöglichkeiten der Lesenden mit dem Text erhöhen (wie z.B. eingebettete Hyperlinks, die die Leser/innen zu anderen Abschnitten, Seiten

oder Websites führen; erweiterte Suchfunktionen; oder soziale Interaktion wie bei E-Mail, Foren oder Instant Messaging-Diensten (vgl. ebd.)).

Auch in den Lese- und Hörverstehensaufgaben, die im Forschungsprojekt IFB entwickelt und in der vorliegenden Studie untersucht wurden, wurden Textsorten eingesetzt, denen die Schüler/innen in ihrem Alltag normalerweise in digitaler Form begegnen. Dies wurde im Sinne eines Testkonstrukts vorgenommen, welches lebensnahe Sprachhandlungen abbilden möchte. Die computerbasierten IFB-Leseverstehensaufgaben umfassten sowohl „statische“ Texte (z.B. das Programm eines Comicfestivals, das auch ausgedruckt werden könnte) wie auch „dynamische“ Texte (z.B. Suchmaschinenergebnisse oder ein Fragment eines Chats). Letztere wurden in den IFB-Aufgaben jedoch aus technischen Gründen immer in einer statischen Form dargestellt, d.h. in Form eines Screenshots und auf jeweils einem Bildschirm, sodass das Scrollen oder Anklicken der Hyperlinks nicht nötig bzw. möglich war. Die Hörverstehensaufgaben umfassten neben Podcasts auch Sprachnachrichten, die normalerweise in einem Chat z.B. an Schulkolleg/innen geschickt werden (mehr dazu in Kap. 3). Quantitative und qualitative Evidenz aus dem Projekt IFB zeigte u.a., dass die Schüler/innen mit den eingesetzten Textsorten vertraut waren und auf die digitalen Hör- und Lesetexte positiv reagierten, weil 1) sie solche Texte aus ihrem Alltag kannten, und 2) sie in ihrem Fremdsprachenunterricht mit digitalen Lese- und Hörtexten nie oder selten arbeiteten und sich dies in Zukunft wünschen würden (für vertiefende Informationen siehe Barras et al., 2020).

2.1.3.2 SCENARIO-BASED ASSESSMENT

Angesichts der erwähnten Vielfalt der involvierten Faktoren beim Lese- und Hörverstehen überrascht es nicht, dass Testautor/innen oft kritisiert werden, ihre Aufgaben würden der Komplexität dieser Prozesse nicht gerecht werden. Im Fall von Leseverstehenstests betrifft diese Kritik (die sich auch leicht auf das Hörverstehen übertragen lässt) insbesondere:

the failure to incorporate theoretical models of reading into the assessment design, the use of artificial and narrow purposes for reading, an over-reliance on multiple-choice format, the omission of digital texts and multimedia, weak links to instruction and the lack of diagnostic information, too narrow a focus on the product of comprehension rather than the process of how it unfolds over time, and the failure to control for individual differences such as student motivation and background knowledge (Sabatini et al., 2013, S. 1, mit Verweis auf weitere Quellen).

Insbesondere wird verschiedentlich hervorgehoben, dass in „traditionellen“ standardisierten Tests zum Leseverstehen Aufgaben und Texte eingesetzt werden, die nicht der Bandbreite

von Texten und Aufgaben entsprechen, denen die Schüler/innen im 21. Jahrhundert ausgesetzt sind (Partnership for 21st century Skills, 2004, 2008; Pellegrino et al., 2001; Rupp et al., 2006; zit. in Sabatini et al., 2013, S. 695).

Im Rahmen der *Reading-for-Understanding*-Initiative wurde vom Educational Testing Service (ETS) (ETS, 2019) ein neuartiges Beurteilungsdesign entwickelt – das sogenannte *Scenario-Based Assessment* (SBA). Mithilfe von SBA sollen das Leseverstehenskonstrukt breiter gefasst und die Testaufgaben zum Leseverstehen besser auf die theoretischen Grundlagen abgestützt werden (ETS Research, 2019). Die *Reading-for-Understanding*-Initiative und ihre Beurteilungsinstrumente wurden für Schüler/innen aus den USA erarbeitet, die in sogenannten *pre-K-12*-Klassen (Kindergarten bis 12. Schuljahr) lernen. Die Initiative hatte zum Ziel, die Leseverstehenskompetenzen der Schüler/innen in ihrer Schulsprache zu fördern (ebd.).

Um ein authentisches Leseziel anzubieten, werden die SBA-Testaufgaben nicht isoliert dargestellt, sondern – wie bereits der Name verrät – um ein zentrales, übergreifendes Leseziel bzw. Szenario herum organisiert (vgl. O'Reilly & Sabatini, 2013; Sabatini et al., 2013). Ein solches Szenario kann beispielsweise die Koordination der gemeinsamen Arbeit mit Schulkolleg/innen bei der Vorbereitung einer Präsentation oder einer Prüfung sein. Beim Bearbeiten der Aufgaben, die zu einem Szenario gehören, werden die Testteilnehmenden mit unterschiedlichen digitalen Textsorten wie z.B. Blogs, Websites und traditionellen Textsorten konfrontiert. Sie müssen auch eine Reihe von Teilaufgaben bearbeiten, um ihr Endziel zu erreichen, z.B. Quellen bewerten, wichtige oder relevante Ideen identifizieren, oder auch Entscheidungen treffen. Die SBA-Aktivitäten sollen dabei denjenigen Leseaktivitäten ähneln, die man auch in der Schule, im Beruf oder in der Freizeit antrifft (vgl. ebd.). SBA kombiniert also „a cluster of techniques for delivering a set of tasks and items that provide a platform for measuring the kinds of demanding reading skills“, und soll so „the instructional relevance of the assessment“ erhöhen (Shore et al., 2017, S. 234).

SBA, wie es vom ETS-Team für das Leseverstehen konzipiert wurde (vgl. ETS Research, 2019; O'Reilly & Sabatini, 2013; Sabatini et al., 2013, 2014a; Shore et al., 2017), unterscheidet sich von anderen Arten von Beurteilungsinstrumenten dadurch, dass die Testaufgaben unter Berücksichtigung der folgenden Prinzipien und Techniken erstellt werden:

- sie bieten ein authentisches Leseziel (*purpose for reading*);
- sie stellen das Lesen in einen Kontext und die zusammenhängenden Aktivitäten können von stärker gelenkten zu weniger gelenkten übergehen;

- die Items erfordern in der Regel die Integration und Einschätzung von Informationen aus verschiedenen Quellen; und
- in vielen Fällen beinhalten die Tasks zusätzliche Hilfestellungen, z.B. in Form von grafischen Elementen zur Analyse von Textstrukturen oder in Form von Richtlinien, wie z.B. Tipps zum Verfassen einer Zusammenfassung, die genutzt werden können, um einen solchen Text zu erstellen (Shore et al., 2017, S. 234–235; vgl. auch O'Reilly & Sabatini, 2013).

SBA kann laut Shore et al. (2017) ausserdem auch Items enthalten, die den sozialen Aspekten einer authentischen Sprachverwendungssituation Rechnung tragen. Dazu gehören z.B. Interaktionen des Testteilnehmenden mit Peers oder mit der Lehrperson, die dazu dienen, im Test gelesene Informationen besser zu verstehen oder bestimmte Aspekte zu klären. Dadurch soll das Spektrum der Interaktionen, Perspektiven und Informationen, denen die Teilnehmenden im Test begegnen, möglichst breit gehalten werden, was auch der im 21. Jahrhundert benötigten Lesekompetenz besser entsprechen soll (ebd.).

Dank der Ähnlichkeit mit authentischen Sprachverwendungssituationen soll SBA diejenigen kognitiven Prozesse besser erfassen, die tatsächlich bei der Auseinandersetzung mit authentischen Lesetexten zum Einsatz kommen (Sabatini et al., 2013, S. 29). Dadurch soll die Motivation der Testteilnehmenden erhöht werden, was wiederum dazu beitragen soll, dass diese eher bereit sind, ihr Bestes zu geben, um bei dem Test gut abzuschneiden (ebd., S. 29-30; vgl. auch O'Reilly & Sabatini, 2013). Dadurch soll der Test ihre wahre Fähigkeit besser widerspiegeln, was einer Verbesserung der Validität der Testnutzung gleichkommt (Braun et al., 2011; zit. in Sabatini et al., 2013, S. 30).

Die Ergebnisse der bisherigen SBA-Implementierungsversuche des ETS-Teams zeigten, dass die szenariobasierten Aufgaben (deren Erstellung übrigens nicht ohne Herausforderungen ist, vgl. Sabatini et al., 2014a) psychometrisch adäquat funktionieren und an Computern in realen Schulsituationen durchführbar sind (Shore et al., 2017, S. 239; siehe auch Sabatini et al., 2014a). Auch sind die Testresultate reliabel und im Sinne der *concurrent validity* mit anderen Tests zum Leseverstehen vergleichbar (Sabatini et al., 2014a). Ausserdem zeigte sich, dass bestimmte Merkmale von SBA-Aufgaben hilfreich sein können, um im diagnostischen Sinne mehr über bestimmte Aspekte der Kompetenz der Schüler/innen zu erfahren (Shore et al., 2017, S. 239). So konnte beispielsweise durch die Unterteilung einer komplexen Aufgabe in kleinere Schritte besser erfasst werden, was die Schüler/innen gut können und womit sie noch Mühe haben (ebd., S. 239–240). Shore et al. (ebd., S. 240) schlussfolgern daraus vorsichtig:

While more research is needed to uncover the potential value of using SBA, there is preliminary evidence to suggest that it can both help elicit complex thinking and help identify what parts of a more complex task students can or cannot do.

Der vom ETS-Team für den L1-Kontext entwickelte szenariobasierte Ansatz wirkt vielversprechend, nicht zuletzt wegen des positiven Washback-Effekts auf den Unterricht, der mit diesen Testaufgaben angestrebt wird. Aus diesem Grund wurde im Forschungsprojekt IFB der Versuch unternommen, das Konzept der szenariobasierten Beurteilung auf die Überprüfung von Verstehenskompetenzen in den Fremdsprachen zu übertragen. Die im Projekt IFB potenziell für den Einsatz im *large scale assessment* entwickelten Lese- und Hörverstehensaufgaben wurden in übergreifende Szenarien eingebettet. Durch die Nähe zu realweltlichen Aufgabenstellungen sowie die Verwendung von digitalen Textsorten, die den Lernenden aus ihrem Alltag bekannt sein sollten, sollte ein möglichst authentisches Lesen und Hören in der Fremdsprache simuliert werden (mehr Informationen dazu in Kap. 3).

2.2 EINSATZ UND ERFORSCHUNG VON STRATEGIEN BEI DER BEWÄLTIGUNG VON (REZEPTIVEN) SPRACHTESTAUFGABEN

Wenn neue Testaufgaben für standardisierte Beurteilungsanlässe erstellt werden, werden sie vor ihrem Einsatz normalerweise gründlich erprobt. Während der Testerprobung werden oft neben den psychometrischen Eigenschaften der Testaufgaben auch – wie in dieser Studie – die Strategien der Testteilnehmenden untersucht (vgl. u.a. A. D. Cohen, 2011b). Dieses Kapitel ist daher den Strategien gewidmet, die beim Lösen der Sprachtestaufgaben eingesetzt werden können. Dazu wird zunächst die Rolle der strategischen Kompetenz in den Modellen der kommunikativen Kompetenz geschildert (2.2.1). Anschliessend wird in Kap. 2.2.2 auf die gegenwärtigen Definitions- und Klassifizierungsversuche von L2-Strategien eingegangen, da die Forschung auf dem Gebiet der *sprachtestbezogenen* Strategien stark von den Erkenntnissen, Instrumenten und Forschungsmethoden aus dem allgemeineren Gebiet der Sprachlernstrategien beeinflusst wurde (Phakiti, 2016, S. 76). Es folgt eine Darstellung der von A. D. Cohen (u.a. 2007b, 2011b, 2012, 2014) vorgeschlagenen und für den empirischen Teil dieser Arbeit relevanten Klassifizierung von sprachtestbezogenen Strategien (Kap. 2.2.3). Zum Schluss (Kap. 2.2.4) wird näher auf die Erforschung von testbezogenen Strategien mittels Introspektion im Rahmen der Validierung von Sprachtestaufgaben eingegangen, bevor fünf neuere Validierungsstudien zu rezeptiven Testaufgaben vorgestellt werden, die u.a. mithilfe introspektiver Verfahren die Strategien der Sprachlernenden untersuchten und deshalb für die Erarbeitung des Designs und/oder der Analyseschritte der vorliegenden Studie wichtig waren. Es handelt sich dabei um die Untersuchungen von A. D. Cohen und Upton (2006), Nikolov (2006), Rossa (2012), L. Gu und So (2017) und Suvorov (2018).

2.2.1 DIE ROLLE DER STRATEGISCHEN KOMPETENZ IN MODELLEN DER KOMMUNIKATIVEN KOMPETENZ

Die strategische Kompetenz spielt eine wichtige Rolle in Modellen der kommunikativen Kompetenz, die in der Fremdsprachen- und Sprachtestforschung zentral sind. Sie wurde gemäss A. D. Cohen (2014, S. 893) wohl zum ersten Mal im Modell von Canale und Swain (1980) theoretisch und als integraler Bestandteil der kommunikativen Kompetenz betrachtet.

Canale und Swain (1980) unterteilen kommunikative Kompetenz in 1) die grammatikalische Kompetenz (d.h. die Kenntnisse des Wortschatzes, der Syntax, der Morphologie, der Satzsemantik (*sentence-grammar semantics*) und der Phonologie); 2) die soziolinguistische Kompetenz (d.h. die Fähigkeit, die Sprache im soziolinguistischen Kontext angemessen zu verwenden); und 3) die strategische Kompetenz. Die strategische Kompetenz betrachten sie als verbale und nonverbale kommunikative Strategien, „that may be called into action to compensate for breakdowns in communication due to performance variables or to insufficient competence“ (Canale & Swain, 1980, S. 30)¹⁸. Canale (1983) ergänzte diese Differenzierung später um die Dimension der Diskurskompetenz (die im Modell von Canale und Swain noch als Teil der soziolinguistischen Kompetenz dargestellt war), verstanden als die Fähigkeit der Lernenden, Äusserungen in verschiedenen Genres (wie z.B. mündliche und schriftliche Erzählung, argumentatives Essay oder wissenschaftlicher Bericht) so zu verknüpfen, dass sie einen zusammenhängenden und kohärenten Text bilden (vgl. ebd., S. 9). Er differenzierte auch die strategische Kompetenz weiter: Gemäss der neuen Auffassung wird auf die verbalen und nonverbalen kommunikativen Strategien nicht nur aus kompensatorischen Gründen zurückgegriffen, sondern auch zur effektiveren Gestaltung der Kommunikation (vgl. ebd., S. 10–11).

Auf die Ansätze von Canale und Swain (1980) und Canale (1983), bei denen es sich übrigens nicht um Modellierungen im eigentlichen Sinne handelt, weil sie keine Interaktionen der Komponenten darstellen (Canale, 1983, S. 12; vgl. auch Wisniewski, 2016, S. 641), nahmen u.a. Bachman (1990) und Bachman und Palmer (1996, 2010) Bezug. Ihre Überlegungen sind für die vorliegende Studie insofern von Bedeutung, als dass sie der strategischen Kompetenz auch bei der Bewältigung von Sprachtestaufgaben eine wichtige Rolle zuschreiben (Bachman, 1990, S. 104–106; Bachman & Palmer, 1996, S. 70–75, 2010, S. 48–55). Ausserdem ist im Sprachtestbereich das Modell der kommunikativen L2-Kompetenz von Bachman und Palmer (1996, 2010) weit verbreitet (vgl. Wisniewski, 2016, S. 643).

¹⁸ Dieser defizitorientierte Ansatz war in den 1980er Jahren gemäss Wisniewski (2016, S. 641) nicht untypisch.

Bachman (1990) beschreibt die strategische Kompetenz als eine von drei Komponenten der kommunikativen Sprachfähigkeit (*communicative language ability*), und zwar neben der Sprachkompetenz und den psychophysiologischen Mechanismen. Für Bachman handelt es sich bei der strategischen Kompetenz um „the mental capacity for implementing the components of language competence in contextualized communicative language use“ (ebd., S. 84). Die strategische Kompetenz bietet also die Möglichkeit, die Sprachkompetenzen der Sprachbenutzer/innen in einem bestimmten Kommunikationskontext mit ihrem Weltwissen in Beziehung zu setzen (ebd.). Bachman sieht die strategische Kompetenz als einen wichtigen Teil jeglichen kommunikativen Sprachgebrauchs und unterteilt sie in eine Bewertungs-, Planungs- und Ausführungskomponente (*assessment, planning, execution component*) (ebd., S. 100–103). Dank der Bewertungskomponente sind die Sprachbenutzer/innen in der Lage zu bestimmen, 1) welche Informationen für die Realisierung eines bestimmten Kommunikationsziels in einem bestimmten Kontext wichtig sind; 2) mit welchen zur Verfügung stehenden Sprachkompetenzen sowie 3) mit welchen Fähigkeiten (*abilities*) und mit welchem Wissen, das mit dem Gesprächspartner geteilt wird, diese Informationen am effektivsten zum Erreichen des kommunikativen Ziels eingesetzt werden; und schliesslich 4) nach dem Kommunikationsversuch zu bestimmen, ob das Kommunikationsziel auch erreicht wurde. Die Planungskomponente ermöglicht Sprachbenutzer/innen, auf relevante (z.B. grammatikalische oder soziolinguistische) Elemente ihrer Sprachkompetenz zurückzugreifen und darauf basierend einen Plan für die Ausführung und Erreichung des kommunikativen Ziels zu formulieren. Die Ausführungskomponente schliesslich ist für das Zurückgreifen auf die relevanten psychophysiologischen Mechanismen zuständig, die für die kontext- und zielangemessene Ausführung des kommunikativen Plans relevant sind (ebd.).

Laut Bachman (ebd., S. 104–106) kann sich die strategische Kompetenz auch auf die Leistung in einem Sprachtest auswirken. So können beispielsweise Sprachlernende, die in zwei verschiedenen Sprachtests gleich gute Resultate erzielen, bei einem dritten, ähnlichen Test unterschiedlich bewertet werden, was bei einer näheren Betrachtung auf die von ihnen eingesetzten (bzw. eben nicht eingesetzten) Strategien zurückgeführt werden kann. Auch können gewisse Testtasks „almost by design“ strategische Kompetenz messen, z.B. Items, die eine Inferenz verlangen (ebd., S. 105). Bestimmte Tasks sind zudem „particularly susceptible to the effects of test takers’ strategic competence, in that they can successfully complete such tasks by utilizing their strategic competence to compensate for deficiencies in other competencies“ (ebd., S. 105). Bachman weist auch darauf hin, dass die Art der Bewertung den Effekt der strategischen Kompetenz auf die Testleistung beeinflussen kann: Wird beispielsweise eine komplexe Testleistung nur global aufgrund des am Ende stehenden

Produktes bewertet, kann der potenzielle Beitrag der strategischen Kompetenz zum Endergebnis höher sein (ebd.).

Auch wenn Bachman die strategische Kompetenz vor allem im Kontext des Sprachgebrauchs (*language performance*) beschreibt und die strategische Kompetenz – wie bereits erwähnt – als eine von drei Komponenten der kommunikativen Sprachfähigkeit konzeptualisiert, gibt er gleichzeitig zu, dass er die strategische Kompetenz nicht nur als einen Aspekt der Sprachkompetenz betrachtet, sondern mehr als eine allgemeine Fähigkeit (*general ability*):

However, rather than considering strategic competence solely an aspect of language competence, I consider it more as a general ability, which enables an individual to make the most effective use of available abilities in carrying out a given task, whether that task be related to communicative language use or to non-verbal tasks such as creating a musical composition, painting, or solving mathematical equations (Bachman, 1990, S. 106).

Diese allgemeine Fähigkeit erlaubt es, die verfügbaren Fähigkeiten bei der Ausführung von verbalen, aber auch nonverbalen Aufgaben möglichst effektiv zu nutzen (ebd.). Sie kann auch im Zusammenhang mit Intelligenz betrachtet werden (ebd.), was wiederum weitere Fragen aufwirft, auf die sich Bachman aber nicht weiter einlässt. Er verweist jedoch darauf, dass die Bestimmung der Auswirkungen verschiedener Fähigkeiten auf die Testleistung im Rahmen der empirischen Konstruktvalidierung untersucht werden kann (ebd., S. 106–107). Laut Wisniewski ist es in Bachmans' Modell trotzdem unklar, „ob eine seiner zentralen Komponenten überhaupt messbar bzw. von anderen kognitiven Konstrukten abgrenzbar ist und tatsächlich als *Kompetenz* begriffen werden kann“ (2016, S. 643, Hervorhebung im Original).

Bachman und Palmer (2010)¹⁹ unterscheiden in ihrem Modell zwischen den Ebenen der Planung und der Ausführung einer kommunikativen Sprachhandlung und siedeln die strategische Kompetenz auf der Ebene der Planung an. Strategische Kompetenz definieren sie als „higher-order *metacognitive* strategies that provide a management function in language use, as well as in other cognitive activities“ (ebd., S. 48, Hervorhebung M.B.). Metakognitive Strategien sind also zusammen mit Sprachkenntnissen und dem thematischen Wissen an einem Umsetzungsplan für das kommunikative Ziel oder für die Bewältigung einer Sprachgebrauchsaufgabe (*language use task*) bzw. einer Prüfungsaufgabe (*assessment*

¹⁹ Da die Ausführungen in Bachman und Palmer (1996) in der überarbeiteten Ausgabe ihres Buches (2010) von den Autoren aktualisiert wurden, wird im Weiteren lediglich auf Bachman und Palmer (2010) Bezug genommen.

task) beteiligt (ebd., S. 52). Bachman und Palmer unterscheiden dabei nicht wie etwa A. D. Cohen (1998) zwischen bewusst bzw. kontrolliert eingesetzten Strategien und unbewusst oder automatisch ablaufenden Prozessen, sondern vertreten die Meinung, dass metakognitive Strategien sowohl kontrolliert als auch automatisch eingesetzt werden können und verwenden daher in ihrem Buch für beide Fälle die Bezeichnung *strategy* (Bachman & Palmer, 2010, S. 57–58). Metakognitive Strategien kommen laut Bachman und Palmer in drei allgemeinen Bereichen zum Einsatz: Im Bereich der Zielsetzung (*goal setting*), der Einschätzung der sprachlichen Situationen (*appraisal*) und der Planung von Äusserungen (*planning*) (ebd., S. 49, deutsche Begriffe nach Wisniewski 2016, S. 643). Für die Ausführungsebene werden hingegen laut Bachman und Palmer *kognitive* Strategien eingesetzt, d.h. was Sprachbenutzer/innen anwenden, „when they execute plans so as to realize these in language use, either in comprehending information in the discourse, or in co-constructing discourse with another interlocutor“ (2010, S. 43). Auf kognitive Kompetenzen wird im Modell von Bachman und Palmer nicht weiter eingegangen, da sie nicht als Teil der strategischen Kompetenz gesehen werden. Wisniewski kritisiert, dass die Autoren zwar die strategische Kompetenz in das Modell der *communicative language ability* einbetten und ihre Funktionsweise mit Beispielen illustrieren, aber die Kompetenz „allein als Ensemble ihrer Komponenten“ beschreiben (2016, S. 643). Wisniewski weist ausserdem darauf hin, dass die Teststrategien in keinem der oben dargestellten Modelle thematisiert werden und dass diese Modelle „die Ausarbeitung einer vertikalen Kompetenzdimension“ vermeiden (ebd., S. 643), womit die Unterteilung der strategischen Kompetenz in (vertikale) Niveaustufen gemeint ist (vgl. ebd., S. 638).

Im GeR (Europarat, 2001), einem Dokument, das seit seiner Publikation ununterbrochen einen grossen Einfluss auf Sprachpolitik, Curricula, Sprachunterricht und Sprachtests in und ausserhalb Europas ausübt, wird strategische Kompetenz als zentral betrachtet und an mehreren Stellen erwähnt (vgl. u.a. Kap. 2.1.5, 4.4, 4.5, 7.2. im GeR). Allerdings wird dabei weder von einer strategischen Kompetenz im eigentlichen Sinne gesprochen noch eine Modellierung angestrebt (Wisniewski, 2016, S. 643). Bei der Ausführung von kommunikativen Aufgaben handelt es sich laut GeR um komplexe Vorgänge und diese Komplexität erfordert „das strategische Zusammenspiel einer ganzen Reihe von Lernerkompetenzen und aufgabenbezogener Faktoren“, wobei die allgemeinen und kommunikativen Strategien „ein ganz wesentliches Bindeglied zwischen den verschiedenen (angeborenen oder erworbenen) Kompetenzen eines Lernenden und der erfolgreichen Bewältigung einer Aufgabe dar[stellen]“ (Europarat, 2001, S. 155). Strategien wird also die Rolle von „Gelenkstellen“ zwischen den Kompetenzen der Lernenden und den kommunikativen Aktivitäten, bei denen sie ihre Kompetenzen einsetzen, zugeschrieben (ebd., S. 38). Der GeR beinhaltet ausserdem

Beispielskalen zu einzelnen Strategien, z.B. eine Skala „Planen“ zu den Produktionsstrategien (ebd., S. 70), eine Skala „Hinweise identifizieren/erschliessen“ (ebd., S. 78) zu den Rezeptionsstrategien oder eine Skala „[u]m Klärung bitten“ zu den Interaktionsstrategien (ebd., S. 89). Laut Wisniewski ist allerdings nicht klar, inwieweit die Deskriptoren in diesen Skalen „theoretisch haltbar und empirisch überprüfbar“ sind (2016, S. 644), da die GeR-Deskriptoren weder theoriebasiert erstellt noch an empirischer Lernaltersprache validiert wurden. Sie stellen vielmehr einen Bewertungskonsens von Sprachlehrpersonen dar, die am sog. Schweizer Projekt zur Skalierung der GeR-Skalen beteiligt waren (ebd.).

In allen oben genannten Modellen und im GeR wird der strategischen Kompetenz eine grosse Wichtigkeit zugeschrieben. Diese breite Anerkennung der strategischen Kompetenz beim Erwerb und Gebrauch von Fremdsprache(n) ging jedoch bisher nicht mit einer umfassenden Modellierung einher, die auch interdisziplinär akzeptiert wäre, wie Wisniewski (2016, S. 637) hervorhebt.

2.2.2 L2-STRATEGIEN: DEFINITIONS- UND KLASSIFIZIERUNGSVERSUCHE

Den Strategien, die beim Spracherwerb und bei der Sprachverwendung zum Einsatz kommen bzw. kommen können, wird in der Fremdsprachenerwerbsforschung seit dem Aufkommen des Kognitivismus und seit den Publikationen von Rubin (1975) und Stern (1975) über die „good language learners“ grosse Aufmerksamkeit geschenkt (u.a. O'Malley & Chamot, 1990; Oxford, 1990; Wenden, 1991; vgl. Griffiths, 2018, S. 50). Trotz der über 40-jährigen Forschungstradition ist man sich jedoch auch heutzutage noch nicht ganz darüber einig, wie Strategien überhaupt zu definieren sind (vgl. u.a. A. D. Cohen, 2007a, 2011b; Dörnyei, 2005; Y. Gu, 2012; Oxford, 2017; Rose, 2012; Wisniewski, 2016). Diese Uneinigkeit auf konzeptueller und definitorischer Ebene führte dazu, dass Ellis bereits in den 90er Jahren das Konzept der Strategie als „somewhat fuzzy“ bezeichnete (1994, S. 529). Sie war auch der Grund, warum Dörnyei (2005) Strategien zwar als ein hilfreiches Instrumentarium für aktives, autonomes und bewusstes Lernen bezeichnete, aber die Nützlichkeit des unklaren Konstrukts der Lernstrategien für die Forschung infrage stellte. Als Ersatz schlug er daher für Forschungszwecke das in seinen Augen vielfältigere Konstrukt von *self-regulation*²⁰ bzw. *self-regulated learning* vor. Dieses Konstrukt sei dynamischer, prozessorientierter und werde im Bereich der *educational psychology* bereits seit den 1990er Jahre anstelle der Lernstrategien erforscht (Dörnyei, 2005, S. 195). Der Vorschlag von Dörnyei löste in der

²⁰ Dörnyei (2005, S. 191) schreibt dazu: „[S]elf-regulation refers to the degree to which individuals are active participants in their own learning; it is a more dynamic concept than learning strategy, highlighting the learners' own “strategic efforts to manage their own achievement through specific beliefs and processes” (Zimmerman & Risemberg, 1997 p. 105)“.

L2-Strategieforschung eine angeregte und z.T. leidenschaftlich geführte Diskussion aus (z.B. Oxford, 2017; Rose, 2012; vgl. auch die Stellungnahme und die teilweise revidierte Position von Dörnyei in Dörnyei & Ryan, 2015).

A. D. Cohen (2011b, S. 48) zeigte sich in der zweiten und überarbeiteten Auflage seines zuerst 1998 erschienenen Buches „Strategies in learning and using a second language“ besorgt über die Versuche, Sprachlernerstrategien (*language learner strategies*) zugunsten von Selbstregulierung herunterzuspielen:

A concern that I have with regard to self-regulation enthusiasts is that they may be throwing the baby out with the bath water by giving such short shrift to the role of language learner strategies since I feel that a judicious use of them can enhance even the bumpiest of L2 learning experiences.

In seiner eigenen Arbeitsdefinition von Sprachlernerstrategien beschreibt A. D. Cohen diese als „thoughts and actions, consciously chosen and operationalized by language learners, to assist them in carrying out a multiplicity of tasks from the very onset of learning to the most advanced levels of target-language performance“ (2011b, S. 7).

Auch Y. Gu (2012) kritisiert die Idee von Dörnyei, auf den bereits breit etablierten Begriff „Strategien“ zu verzichten und bezeichnet diese als „not a healthy sign“ (ebd., S. 330). Y. Gu argumentiert, dass die Definitionsspitzzfindigkeiten zum einen nichts zum Erkenntnisfortschritt beitragen und zum anderen auch den Lehrpersonen und Lernenden in der Praxis nicht viel bringen würden. Mit Verweis auf weitere Stimmen fügt er weiter hinzu (ebd., S. 331):

A few scholars have argued against the proposed solution (e.g. Gao, 2007; Rose, 2012), reasoning that conceptual fuzziness should not be a problem serious enough to overthrow forty years of research on language learning strategies. The argument is clear and straightforward: if not being able to agree on the definition of a Planet until 2006 does not in any way discredit the scientific nature of astronomy, or necessitate the removal of the concept of “planet” altogether, why should we throw away a whole line of research on language learning strategies? In fact, the proposed alternative term “selfregulation” or even a more general and key term “learning” fall into the same fuzziness trap [...]. This indicates to me that the find-another-term solution is not viable.

Y. Gu machte den Gegenvorschlag, das Konzept der Lernstrategien nicht aus einer kategorialen, sondern aus einer prototypischen Perspektive anzugehen. Lernstrategien liessen sich nämlich nicht klar definieren, „because strategies differ from each other in terms

of „family resemblance” or „graded degrees of membership”“ (2012, S. 330). Der prototypische Kern einer Strategie sei ein dynamischer Prozess, der zum Ziel habe, ein Problem zu lösen. Er umfasse die selektive Aufmerksamkeit, die Analyse der Aufgabe, die Auswahl der Entscheidungen, die Ausführung des Plans, das Monitoring des Fortschritts und/oder die eventuelle Änderung des Plans sowie auch die Evaluation der Ergebnisse. Eine Strategie bewege sich somit im Hinblick auf ihre Prototypikalität entlang der Dimensionen der Intentionalität, der Selbstinitiierung, der metakognitiven Regulierung, der Sequenzialität der Aktivitäten, der Bündelung (*chunking*) der Aktivitäten und der Automatisität bei der Ausführung der Strategie. Als praktische Lösung schlägt Y. Gu daher vor, einen prototypischen Kern zu finden und die Dimensionen der Variation aufzuzeigen. Eine prototypische Strategie ist für ihn „a purposeful, self-initiated, and intentionally selected, monitored, and evaluated sequential series of activities“ (ebd., S. 342), wobei „the skillful execution of a strategy is done automatically“ (ebd., S. 348). Die Prototypikalität einer Strategie hänge daher davon ab, wie viele Dimensionen vorhanden seien und wie nahe jede Dimension am prototypischen Kern läge. Bei der Konzeptualisierung von Strategien solle aber nicht vergessen gehen, dass sie einen Schlüsselbegriff darstellen, der die Beziehungen zwischen Person, Situation und Leistung erklärt, fügt Y. Gu (ebd., S. 343) in Anlehnung an Biggs (1984) hinzu. Welche Strategien mit welchem Ergebnis von den Lernenden angewandt würden, hänge nämlich zwar schon von den Lernenden und ihren Einstellungen, ihrer Motivation usw. ab, aber auch die jeweilige Aufgabe (u.a. ihre Schwierigkeit oder Komplexität) und die Situation bzw. Umgebung hätten darauf einen Einfluss. Lernstrategien können daher nicht einfach als allgemeine oder spezifische strategische Kompetenz und kontrollierte oder automatische strategische Leistung betrachtet werden, die unabhängig vom Lernenden und dem Kontext vorkommen (Y. Gu, 2012, S. 347).

Oxford beschäftigt sich bereits seit den 1990er Jahren intensiv mit *learner strategies* und äusserte sich, ähnlich wie u.a. Y. Gu (2012), sehr kritisch zu Dörnyei's (2005) Idee, anstelle von *language learner strategies* die Selbstregulierung (*self-regulation*) zu beforschen (vgl. u.a. Oxford, 2017, S.10-11). In der überarbeiteten Ausgabe ihres Buches „Teaching and researching language learning strategies“²¹ (2017) setzte sie sich u.a. zum Ziel, etwas Ordnung in das definitorische Chaos rund um Strategien zu bringen. Sie führte dafür eine systematische Inhaltsanalyse von 33 Definitionen der *learning strategies* und verwandter Termini wie *learner strategies* oder *strategic* durch, die seit 1975 publiziert worden waren. Darauf basierend schlug sie ihre eigene Definition der L2-Lernstrategien vor, die die Kern- bzw. Prototypenmerkmale enthält und als Grundlage für das von ihr vorgeschlagene

²¹ Die erste Ausgabe erschien 2011.

Strategic-Self-Regulation-Modell dient (ebd., S. 9). Oxford (ebd., S. 48) definiert Strategien neu wie folgt:

L2 learning strategies are complex, dynamic thoughts and actions, selected and used by learners with some degree of consciousness in specific contexts in order to regulate multiple aspects of themselves (such as cognitive, emotional, and social) for the purpose of (a) accomplishing language tasks; (b) improving language performance or use; and/or (c) enhancing long-term proficiency. Strategies are mentally guided but may also have physical and therefore observable manifestations. Learners often use strategies flexibly and creatively; combine them in various ways, such as strategy clusters or strategy chains; and orchestrate them to meet learning needs. Strategies are teachable. Learners in their contexts decide which strategies to use. Appropriateness of strategies depends on multiple personal and contextual factors.

Diese umfassende Definition ist als Aufruf zur Diskussion und zum allgemeinen Konsens im Bereich der Strategieforschung zu verstehen (ebd., S. 1 und 9), und zwar „not necessarily on every point but in general“ (ebd., S. 48).

Auch Griffiths (2018) verweist auf die Schwierigkeiten, Sprachlernstrategien (*language learning strategies*) zu definieren. Sie vergleicht die Definitionsversuche bildlich mit den Bemühungen, einen Kraken in eine Schachtel zu zwingen, während sich seine Teile immer wieder unbemerkt herausschlängeln und die ganzen Bestrebungen zu sabotieren drohen (ebd., S. 7). Sie argumentiert aber, dass es mittlerweile möglich sei, die wesentlichen Merkmale von Strategien zu identifizieren und diese in eine praktikable Definition einfließen zu lassen. Griffiths gibt zu, dass sich auch ihre eigene Definition im Laufe der Zeit weiterentwickelte und schlägt in ihrem Buch eine gegenwärtige (und, wie sie selbst anmerkt, hoffentlich endgültige) Definition vor. Diese lautet:

Language learning strategies are:

- actions (the learner has to DO something);
- chosen by learners (as distinct from being imposed by someone else, e.g. the teacher);
- for the purpose of (they are goal-oriented);
- learning language (as distinct from e.g. communicating) (ebd., S. 19).

Die Autorin weist gleichzeitig darauf hin, dass heutzutage immer noch kein Konsens u.a. darüber herrscht, ob Strategien nur mentale oder auch physische Aktivitäten (wie z.B. das Notieren neuer Wörter) umfassen (ebd., S. 9). Sie versteht daher ihre knapp gefasste

Definition gemäss dem Vorschlag von Y. Gu (2012, S. 330) als einen „prototypischen Kern“ von Aktionen, die als Strategien gefasst werden können, wobei „Dimensionen der Variation“ je nach konkreten Zielsetzungen und Perspektiven von bestimmten Studien immer noch möglich seien. Es sei allerdings wichtig, dass die Forschenden die Zielsetzungen ihrer jeweiligen Studien sorgfältig darstellen und theoretisch begründen, und dass sie die Definition, von der sie ausgehen, entsprechend ausarbeiten und erklären (Griffiths, 2018, S. 20).

An dieser Stelle soll noch ergänzend auf den einleuchtenden Hinweis von Grabe und Stoller (2011, S. 10) hingewiesen werden: Sie empfehlen, Strategien für Definitionszwecke am besten als Fähigkeiten zu betrachten, „that are potentially open to conscious reflection, and reflect a reader’s intention to address a problem or a specific goal while reading (see Anderson, 2009)“. Diese Empfehlung wurde zwar im Zusammenhang mit dem Leseverstehen formuliert, kann aber auch auf Strategien übertragen werden, die im Zusammenhang mit anderen Fertigkeiten vorkommen.

In der L2-Strategieforschung wurden neben den zahlreichen Definitionsversuchen auch viele unterschiedliche Klassifizierungsversuche unternommen. So wurden Strategien beispielsweise unterteilt in „strategies for language learning vs. language use, strategies by language skill area, and strategies according to function (namely, metacognitive, cognitive, affective, or social)“ (A. D. Cohen, 2011b, S. 8). A. D. Cohen (u.a. 2011b) schlug ausserdem für die Zwecke der Testforschung den Begriff Teststrategien vor (*test taking strategies*, vgl. Kap. 2.2.3).

Oxford (1990) führte eine bekannte Klassifizierung von Strategien in *memory strategies*, *cognitive strategies*, *compensation strategies*, *metacognitive strategies*, *affective strategies* und *social strategies* ein, die auch ihrem Fragebogen *Strategy Inventory for Language Learning* zugrunde lag. Schon damals wies sie allerdings darauf hin, dass die Strategiekategorien überlappen und nicht immer klar voneinander abgrenzbar seien – ein Argument, dass in ihrem neusten Buch wiederholt zum Ausdruck kommt:

The roles or functions a given strategy plays depend on the task, the physical context, and the learner’s internal context (self-efficacy beliefs, mindsets, autonomy, agency, and numerous other factors). No strategy classification, typology, or taxonomy actually directly reflects how strategies operate, because strategies are complex and have multiple roles. We can identify what we think is the “main” category of a given strategy, but the strategy can also serve other purposes besides those implied by that category name (Oxford, 2017, S. 141).

Gemäss A. D. Cohen und Wang (2018, S. 169) ist mittlerweile die Klassifizierung der Strategien nach ihrer Funktion (d.h. nach ihrem Zweck oder ihrer Rolle in einer bestimmten Situation) am breitesten anzutreffen. Gemäss dieser Auffassung wird unterschieden zwischen 1) metakognitiven Strategien (z.B. Planung der Anwendung einer Strategie, Monitoring des Verlaufs oder Evaluierung des Erfolgs), 2) kognitiven Strategien (z.B. gedankliche Auseinandersetzung mit dem Sprachmaterial), 3) sozialen Strategien (z.B. Interaktion mit anderen Sprachbenutzer/innen) und 4) affektiven Strategien (z.B. Umwandlung der positiven oder negativen Reaktionen in strategisches Handeln) (ebd.).

Ausserdem wird verschiedentlich darauf hingewiesen, dass Strategien oft in Clustern bzw. in Ketten auftreten können (u.a. Oxford, 2017, S. 41–42). So kann beispielsweise eine Lernende mehrere Strategien gleichzeitig einsetzen, um die Bedeutung des Gesagten im Gespräch zu ermitteln: Sie kann innerhalb von Sekunden auf ihr Hintergrundwissen zurückgreifen und sprachliche Hinweise, Körpersprache, Tonhöhe und Lautstärke eines Gesprächspartners analysieren. Diese Strategien kommen gleichzeitig vor und können als Cluster betrachtet werden. Als Strategiekette bezeichnet Oxford eine Abfolge von Lernstrategien, die die Lernenden verwenden, um eine Aufgabe leichter zu bewältigen. Eine beispielhafte Strategiekette zum Lesen eines Artikels in der L2 könnte wie folgt aussehen: 1) den Titel des Artikels betrachten, 2) relevantes Hintergrundwissen abrufen, 3) bestimmte Informationen aus dem Kontext beim Lesen erraten, 4) den Erfolg der Strategien bewerten (vgl. ebd.). A. D. Cohen (2011a, S. 682–683) weist darauf hin, dass das Auftreten von Strategien in Clustern oder in Ketten ihre Erforschung erschweren kann:

[I]t may be difficult for researchers to isolate the impact of a single strategy because its actual impact is cumulative, and is based on the effect of other strategies as well. So, while it may be more elegant to list out the strategy types (metacognitive, cognitive, social, and affective) for definitional purposes, the reality is strategies are actually deployed in complex, interacting ways such that at a given moment it may be a challenge to determine the type of strategy that is being utilized.

Um (zumindest auf der theoretischen Ebene) etwas Ordnung in die verschiedenen Klassifizierungsversuche zu bringen, schlägt Griffiths (2018, S. 62 ff.) vor, die *language learning strategies* grundsätzlich in zwei Hauptkategorien einzuteilen: in die metakognitiven bzw. regulatorischen Strategien und in die kognitiven bzw. direkten Strategien, die sich jeweils weiter unterteilen lassen (ebd., vgl. Schema auf S. 63). Unter kognitiven Strategien versteht sie diejenigen Strategien, die die Lernenden wählen, um sich direkt mit dem zu lernenden Material zum Wissens- oder Verständnisaufbau auseinanderzusetzen. Die metakognitiven Strategien werden hingegen eingesetzt, um die Interaktion der Lernenden mit dem

Sprachmaterial zu regulieren: Sie sind für die Steuerung, Aufsicht oder Kontrolle des Lernprozesses zuständig und operieren auf einer indirekten Ebene oberhalb kognitiver Strategien. Auch Griffiths ist sich jedoch bewusst, dass es in der Praxis oft nicht einfach ist, eine bestimmte Strategie ausschliesslich der einen oder anderen Kategorie zuzuordnen (ebd., S. 63). Aus diesem Grund rät sie Forschenden, die Klassifizierung von Strategien in einem Forschungsprojekt gezielt vorzunehmen und im Hinblick auf die beforschte Zielgruppe, Situation sowie Projektziele zu begründen. Die Nützlichkeit der bestehenden Klassifizierungsversuche sollte dabei sorgfältig geprüft und wenn nötig angepasst werden. Auch bei der Interpretation der Resultate sollte der Untersuchungskontext mitberücksichtigt werden (ebd., S. 64).

2.2.3 LERNERSTRATEGIEN UND TESTSTRATEGIEN

Nicht nur aus der Perspektive der Sprachlehr- und Sprachlernforschung, sondern auch aus dem Blickwinkel der Sprachtestforschung und -entwicklung sind die Strategien, die die Sprachlernenden bei der Bewältigung von (Test-)Aufgaben einsetzen, von grosser Bedeutung (vgl. Kap. 2.2.4). A. D. Cohen definiert diese Art von Strategien als „the consciously-selected processes that the respondents use for dealing with both the language issues and the item-response demands in the test-taking tasks at hand“ (2011b, S. 305). Der Grad der bewussten Aufmerksamkeit bei der Auswahl der Strategien soll dabei als ein Kontinuum verstanden werden, das mit der allgemeinen Wahrnehmung (*general awareness*) beginnt und mit der starken Fokussierung (*highly focused attention*) endet (ebd.).

Gemäss A. D. Cohen (2011b) können Teststrategien (*test-taking strategies*), also die bei der Bewältigung von Sprachtestaufgaben zum Einsatz kommenden Strategien, in drei Typen unterteilt werden. Zu diesen gehören Lernerstrategien (*language learner strategies*), Testmanagementstrategien (*test-management strategies*) und *test-wiseness*-Strategien. Bei den Lernerstrategien handelt es sich, wie schon ihre Bezeichnung verrät, nicht um testbezogene Strategien, sondern um allgemeine Strategien für den Umgang mit geschriebenen bzw. gesprochenen L2-Texten, auf die die Sprachlernenden auch unabhängig vom Testkontext zurückgreifen können (ebd.). In seinem Aufsatz von 2014 präzisiert A. D. Cohen daher, dass nur die Testmanagementstrategien und *test-wiseness*-Strategien zu den eigentlichen Teststrategien (*test taking strategies*) zählen (2014, S. 894). Aus der Sicht der Sprachtestforschung sind Lernerstrategien und Testmanagementstrategien konstruktrelevant, weil sie zur gewünschten Varianz beitragen, während *test-wiseness*-Strategien zur konstruktirrelevanten Varianz beitragen (ebd., S. 893) und daher unerwünscht sind.

Im Fokus dieser Arbeit stehen alle Strategien, die von den Schüler/innen beim Lösen der untersuchten Testaufgaben aus dem Projekt IFB eingesetzt wurden (vgl. Kap. 3.2). In den folgenden Unterkapiteln werden daher sowohl die Lernerstrategien wie auch die Teststrategien, d.h. die Testmanagementstrategien und *test-wiseness*-Strategien, näher beleuchtet. Diese Strategietypen wurden bei der empirischen Untersuchung berücksichtigt, indem die Kategorisierung der ermittelten Strategien gemäss dem hier dargestellten konzeptuellen Vorschlag von A. D. Cohen (u.a. 2007b, 2011b, 2012, 2014) vorgenommen wurde (vgl. Kap. 3).

2.2.3.1 LERNERSTRATEGIEN

Bei der Auseinandersetzung mit den Testaufgaben können die Sprachlernenden auf verschiedene Strategien zurückgreifen, die sie auch sonst bei der Anwendung ihrer Fremdsprachenkenntnisse nutzen können (vgl. A. D. Cohen, 2014, S. 895). Es handelt sich also nicht um testbezogene Strategien im engeren Sinne. Stattdessen umfassen die Lernerstrategien alle Strategien, die die Sprachlernenden beim Hören, Lesen, Sprechen, und Schreiben in einer L2 einsetzen, inklusive der Strategien zum Umgang mit Wortschatz, Grammatik und Übersetzungen (A. D. Cohen, 2012, S. 263). Zu den Lernerstrategien, die beim Lösen eines Leseverstehenstests vorkommen können, gehören beispielsweise die Identifizierung eines unbekannten Wortes aus dem Kontext oder die Nutzung der morphosyntaktischen Regeln für das Verstehen.

2.2.3.2 TESTSTRATEGIEN

Die Teststrategien werden von A. D. Cohen (2014), wie oben erwähnt, in Testmanagementstrategien und *test-wiseness*-Strategien unterteilt.

Testmanagementstrategien (ebd., S. 895) definiert er als

(...) strategies for responding meaningfully to test items and tasks. These are the processes consciously selected to assist in producing a correct answer responsibly. They include logistic issues such as weighing the importance of responding to different items or tasks, keeping track of the time, and determining where to look for answers.

Diese Strategien helfen also den Lernenden, eine Testaufgabe sinnvoll anzugehen, um eine korrekte Antwort zu erzeugen und sind daher, wie auch die Lernerstrategien, als konstruktrelevant zu betrachten (ebd., S. 893). Diese Strategien sind laut A. D. Cohen testbezogen, weil ihr Einsatz durch die Testsituation und/oder die Testaufgabe(n) evoziert wurde. Zu den Testmanagementstrategien im Zusammenhang mit einer

Leseverstehensaufgabe gehören z.B. der wiederholte Vergleich der Textpassage mit der Testfrage, ein sorgfältiger und systematischer Umgang mit den Multiple-Choice-Optionen oder auch Strategien zur Verwaltung und Nutzung der vorgegebenen Zeit (ebd., S. 895–896).

Zu Testmanagementstrategien kann kritisch angemerkt werden, dass es sich bei ihnen prinzipiell „um metakognitive Lernstrategien wie Planen, Beobachten, Bewerten im Kontext von Sprachtests handelt, so dass der Nutzen der Kategorie der Testmanagementstrategien fraglich ist“ (Wisniewski, 2016, S. 641; vgl. auch Holzknecht, 2019, S. 27–28). In der vorliegenden Arbeit wird dennoch die von A. D. Cohen vorgeschlagene Bezeichnung dieser Strategienkategorie beibehalten, vor allem deswegen, weil sie sich in der Teststrategieforschung mittlerweile etablieren konnte (vgl. u.a. A. D. Cohen, 2014; A. D. Cohen & Upton, 2006; L. Gu & So, 2017; Suvorov, 2018).

Als *test-wiseness*-Strategien bezeichnet A. D. Cohen (2014) Strategien zur Nutzung von Randinformationen (wie z.B. das Wissen über das Funktionieren der Testformate). Diese Strategien werden von Sprachtestteilnehmenden eingesetzt, um den Test „auszutricksen“, d.h. um (möglicherweise sogar die richtigen) Antworten auf die Testfragen zu finden, ohne dabei auf ihre L2-Kenntnisse zurückgreifen zu müssen (vgl. ebd., S. 896). Aus diesem Grund sind diese Strategien auch als konstruktirrelevant zu bezeichnen (ebd., S. 895). Die *test-wiseness*-Strategien können beispielsweise im Rahmen eines Tests mit Multiple-Choice-Antwortoptionen vor allem dann mit Erfolg eingesetzt werden, wenn die Items nicht optimal konstruiert wurden. Das kommt z.B. dann vor, wenn nur eine Antwortoption die grammatikalisch korrekte Antwort auf die Frage beinhaltet; wenn zwei Optionen eliminiert werden können, weil sie praktisch das Gleiche aussagen; oder wenn ein Item bereits die Antwort zu einem weiteren Item verrät (Allan, 1992).

2.2.4 ERFORSCHUNG VON STRATEGIEN IM RAHMEN DER TESTERPROBUNGSPHASE

Die oben dargestellten Strategien, auf die die Sprachlernenden beim Bearbeiten der Testaufgaben zurückgreifen, sind für die Sprachtestentwicklung und Sprachtestforschung von grosser Bedeutung (u.a. A. D. Cohen, 2012). Deswegen werden sie vor allem während der sog. Testerprobungsphase (*pre-operational testing*) (vgl. Kenyon & MacGregor, 2012) untersucht. Dank der Erforschung von Teststrategien konnten inzwischen laut A. D. Cohen (2007b, S. 108) wichtige Hinweise zu folgenden Aspekten gewonnen werden:

- low-level vs. higher-level processing on a test;
- the impact of using authentic vs. inauthentic texts in reading tests;

- whether the strategies employed in L2 test-taking are more typical of first-language [...] use, common to L1 and L2 use, or more typical of L2 use;
- the more effective strategies for success on tests as well as the less effective ones;
- test-takers' vs. raters' understanding of and responses to integrated language tasks; and
- the items on a test that would be susceptible to the use of test-wiseness strategies.

Im Folgenden wird zunächst aufgezeigt, was unter der Testerprobungsphase zu verstehen ist und welche Rolle diese Phase im anschliessend skizzierten Prozess der Testvalidierung spielt. Anschliessend werden fünf neuere Studien zum Lese- und Hörverstehen vorgestellt, in denen die Testbewältigungsstrategien der Sprachlernenden mittels Introspektion (teilweise auch in Kombination mit anderen Methoden) untersucht wurden (A. D. Cohen & Upton, 2006; Nikolov, 2006; Rossa, 2012; L. Gu & So, 2017; Suvorov, 2018). Diese Studien wurden bei der Planung der Datenerhebung und/oder der Datenanalyse in der vorliegenden Studie beigezogen.

2.2.4.1 TESTERPROBUNG UND TESTVALIDIERUNG

Die Testerprobungsphase definieren Kenyon und MacGregor als „the thoughtful, deliberate, and systematic process of collecting and analyzing evidence to support the validity of an assessment *prior* to that assessment's operational use“ (2012, S. 295, Hervorhebung im Original). Speziell ist laut den Autoren an diesem Prozess, dass sein Hauptfokus auf dem Test selbst und weniger auf den Testanwender/innen oder Testteilnehmer/innen liegt. Je nach Schwerpunkt und Erkenntnisinteresse können dabei sowohl qualitative als auch quantitative Datenerhebungsmethoden zum Einsatz kommen. Mit qualitativen Methoden wie Lautem Denken oder *Stimulated Recall* wird in der Regel eher in der ersten, erkundenden Etappe der Testerprobung gearbeitet, die auch als Pilotstudie bzw. Pilotierung (*pilot testing*) bezeichnet wird. Während der Pilotierung werden Daten mit einer kleineren Probandenzahl erhoben, um das Funktionieren des Tests besser kennenzulernen und seine eventuellen Schwachstellen rechtzeitig identifizieren und beheben zu können. Dies ist besonders bei der Entwicklung neuerer, d.h. weniger bekannter Item- und Aufgabentypen wichtig (ebd., vgl. auch Reed, 2014, S. 877). Im Zentrum des Interesses während dieser ersten Phase stehen Probleme mit unverständlichen Anweisungen, mit dem Layout, mit schwer nachvollziehbaren Elementen der Testaufgaben oder auch mit der Testadministration oder Bewertung. Mit quantitativen Erhebungsmethoden wird hingegen später beim Feldtest (*field testing*) gearbeitet, der normalerweise mit einer grösseren Probandenzahl durchgeführt wird. Dabei wird v.a. untersucht, ob die Items psychometrisch wie gewünscht funktionieren. Weil der Einsatz der sog. *mixed methods* oft eine reichhaltigere Informationsquelle als die einzelnen Methoden

alleine bietet und mit qualitativen und quantitativen Methoden zum Teil auch unterschiedliche Informationen erfasst werden, wird bei der Testerprobung die Kombination der beiden methodischen Ansätze empfohlen (Kenyon & MacGregor, 2012; vgl. auch Reed, 2014).

Wie bereits der oben zitierten Definition der Testerprobung (Kenyon & MacGregor, 2012, S. 295) entnommen werden kann, wird dabei Evidenz für die Validität eines Tests gesammelt. Es handelt sich bei Validität um einen theoretischen Begriff, der unterschiedlich aufgefasst werden kann (Xi & Sawaki, 2017, S. 194). Xi und Sawaki heben hervor, dass die gewählte Konzeptualisierung von Validität von grosser Bedeutung ist, weil sie die Art und den Umfang der sog. Validierung beeinflusst (ebd.).

In den Anfängen der (Sprach-)Testforschung wurde Validität nach Lado wie folgt definiert: „Does a test measure what it is supposed to measure? If it does, it is valid“ (1961, S. 321, zit. nach Chapelle, 1999, S. 255). Sie wurde damals als eine Testeigenschaft und als ein „all-or-nothing attribute“ verstanden (Chapelle, 1999, S. 255). Um die so aufgefasste Validität eines Tests aufzuzeigen, griff man auf Verfahren zurück, die Nachweise für verschiedene „types of validity“ erbringen sollten (Chapelle & Voss, 2014, S. 1083). Zu diesen Validitätstypen gehörten Inhaltsvalidität, (prädiktive oder konkurrierende) Kriteriumsvalidität und Konstruktvalidität. Inhaltsvalidität basierte auf systematisch eingeholten Expertenmeinungen, die sich dazu äusserten, inwiefern die Testaufgaben für die Messung der jeweiligen Aspekte des Testkonstrukts geeignet sind. Kriteriumsvalidität wurde auf Basis der Korrelationen zwischen dem Test und anderen Tests untersucht, die das gleiche Konstrukt zu messen intendieren. Bei der Konstruktvalidität griff man ebenfalls auf quantitative Evidenz zurück: Analysen der statistischen Daten sollten aufzeigen, dass die Ergebnisse der Testteilnehmenden mit den theoriebasierten Erwartungen übereinstimmen (ebd.). Die Tatsache, dass Validität nicht als einheitliches Konzept sondern als aus verschiedenen Typen bestehend aufgefasst wurde, führte laut Xi und Sawaki dazu, dass Testanbieter bequemerweise nur einen Typ von Validität für die Unterstützung einer bestimmten Testverwendung beiziehen konnten (2017, S. 195). Die Testvalidierungsmethoden waren folglich auf Korrelationsanalysen, Inhaltsanalysen der Testaufgaben oder faktoranalytische Techniken beschränkt; die Testbewältigungsprozesse und -strategien der Testteilnehmenden sowie Testkonsequenzen wurden hingegen nicht untersucht (ebd.).

Seit der einflussreichen Definition von Messick (1989), die im Bereich der Sprachtestforschung ein Jahr später von Bachman (1990) übernommen wurde (vgl. Chapelle, 2012, S. 23), wird dieses ursprüngliche Verständnis von Validität viel breiter aufgefasst. Messick definiert Validität als „an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and

appropriateness of inferences and actions based on test scores or other modes of assessment“ (1989, S. 13, Hervorhebungen im Original). Validität wird dementsprechend nicht mehr als eine Testeigenschaft angesehen. Stattdessen stehen nun die Interpretationen und Verwendungen eines Tests und seiner Ergebnisse im Fokus, die sich als mehr oder weniger valide erweisen können (Chapelle, 2012, S. 24), und „wozu ein *assessment use argument* herangezogen wird“ (Stadler & Kremmel, 2018, S. 46 mit Verweis auf Bachman, 2005; Weir, 2005; Xi, 2008). Auch wird Validität als ein einheitliches Konzept mit Konstruktvalidität im Mittelpunkt verstanden, das nicht nur die technischen Überlegungen, sondern auch die Relevanz, den Nutzen, die Werteimplikationen und die sozialen Folgen von Tests umfasst (Chapelle, 2012, S. 24). Messick betont darüber hinaus, dass Validität als „a matter of degree, not all or none“ (1989, S. 13) zu verstehen ist.

Als Folge dieser komplexeren Definition von Validität wird auch der Prozess der Testvalidierung umfassender und breiter als bisher angesehen, und zwar als „an ongoing process of inquiry“ (Chapelle, 2012, S. 24), was die Validierung entsprechend anspruchsvoll gestaltet (Cumming, 2012). Der Bereich der Sprachtestforschung erlebte dabei eine Verschiebung „from positivistic research focusing on properties of tests and scores toward a broader and more critical examination of a wide range of validity issues embracing construct definition as well as language testing practice and policy“ (Lumley & Brown, 2005, S. 833). Im Mittelpunkt der Testvalidierung stehen also neu nicht nur die Testeigenschaften, sondern u.a. auch die Antwortprozesse (*response processes*) der Testteilnehmenden (Ercikan & Pellegrino, 2017a), über deren Kompetenz man mithilfe der Testergebnisse Aussagen machen will (siehe z.B. die Studie von Rossa, 2012), sowie die Vorgehensweisen der Rater/innen, die diese Kompetenzen beurteilen (vgl. z.B. die Untersuchung von Arras, 2007).

Ercikan und Pellegrino verstehen unter den *response processes* „the thought processes, strategies, approaches, and behaviors of examinees when they read, interpret, and formulate solutions to assessment tasks“ (2017b, S. 2). Sie bezeichnen es als zentral, im Rahmen einer Testvalidierung (auch) zu untersuchen, wie die Testteilnehmenden bei der Bewältigung der Aufgaben vorgehen, weil ihre Antwortprozesse Aufschluss darüber geben können, inwieweit die Items und Aufgaben von den Testteilnehmenden in der beabsichtigten Weise bearbeitet werden und ob folglich die Schlussfolgerungen, die bei der Erstellung und Interpretation der Testergebnisse gezogen werden, auch gerechtfertigt sind. Besonders relevant ist es dabei zu erfahren, 1) wie die Testteilnehmenden die Testaufgaben lesen und interpretieren, 2) wie sie bei der Bewältigung der Testaufgaben vorgehen und welche Strategien sie dabei anwenden, 3) auf welches Wissen und auf welche Kompetenzen sie bei der Bewältigung der Testaufgaben zurückgreifen, 4) ob sie sich wirklich mit den Testaufgaben beschäftigen und ihre Antworten nicht etwa erraten, wie auch 5) wie sie die im Test verfügbaren Ressourcen

und Informationen nutzen (ebd.). In den einflussreichen „Standards for Educational and Psychological Testing“ (American Educational Research Association et al., 2014, S. 13–20, 26) werden die *response processes* als eine der möglichen Quellen für die Validitätsevidenz eines Tests aufgeführt. Ercikan und Pellegrino (2017b, S. 3) weisen darauf hin, dass die aus den *response processes* der Testteilnehmenden gewonnene Evidenz als ergänzend zu anderen Arten von Validitätsevidenz (z.B. statistische Untersuchungen der Testaufgaben) zu verstehen ist, wobei sie weder andere Formen von Validitätsnachweisen ersetzen noch durch diese ersetzt werden kann. Wenn aber keine aus den *response processes* gewonnene Evidenz für die Validität eines Tests vorliegt, „other types of evidence have questionable value“ (ebd., S. 4). Reed weist beispielsweise darauf hin, dass gemäss mehreren Studien die von Rater/innen vorgenommenen Einschätzungen von Item- und Aufgabenschwierigkeiten nicht gut mit dem übereinstimmen, was die Testteilnehmer/innen tatsächlich beim Lösen dieser Aufgaben als schwierig oder leicht empfinden (2014, S. 879).

Auch A. D. Cohen betont, wie wichtig es ist, bei der Testvalidierung die Strategien der Testteilnehmenden empirisch zu untersuchen (2011b, S. 305). Die Vorteile, die sich daraus ergeben, sind mannigfaltig: Der genaue Blick auf die Strategien kann helfen, Lernerstrategien präziser von Teststrategien (d.h. den Testmanagementstrategien und *test-wiseness*-Strategien) zu unterscheiden. Dank solcher Untersuchungen kann festgestellt werden, ob die Leistung bei einem bestimmten Beurteilungsanlass das L2-Sprachverhalten im beurteilten Bereich widerspiegelt oder eher ein Verhalten repräsentiert, das nur zum Bestehen des Tests eingesetzt wird. Auch können empirische Studien zu Teststrategien wertvolle Information darüber liefern, was die Testaufgaben tatsächlich messen und wie vergleichbar die Ergebnisse verschiedener Testmethoden und Aufgabentypen im Hinblick auf ihren Schwierigkeitsgrad, auf die damit gemessenen Fähigkeiten und auf die bei ihrer Bewältigung eingesetzten Strategien sind; und zwar im Zusammenhang mit den Merkmalen der (Gruppen von) Testteilnehmenden (vgl. ebd., S. 331). Für Testentwickler/innen ist es laut A. D. Cohen besonders interessant zu erfahren, ob und auf welche konstruktirrelevante *test-wiseness*-Strategien die Testteilnehmenden beim Lösen ihrer Aufgaben zurückgreifen, weil diese Strategien zu einer „inflated performance“ bei einem Test führen können (2014, S. 901). Mit den Testaufgaben sollten nämlich die Sprachkenntnisse der Testteilnehmenden gemessen werden und nicht ihre Fähigkeit, den Test „auszutricksen“. Auf den Einsatz solcher Strategien sind Lese- und Hörverstehensaufgaben wegen ihrer Aufgabenformate besonders anfällig. Es sollte daher überprüft werden, ob und warum sich die Testteilnehmenden der *test-wiseness*-Strategien beim Lösen der Testaufgaben bedienen (vgl. ebd.).

Erfahrungsgemäss setzen die Testteilnehmenden *test-wiseness*-Strategien ein, wenn sie die Aufgaben nicht bewältigen *können* (z.B. aufgrund geringer Sprachkenntnisse oder aus

Zeitgründen) oder auch, wenn sie die Aufgaben – aus welchen Gründen auch immer – nicht bewältigen *wollen*. Es kann auch vorkommen, dass die Testteilnehmenden eine Aufgabe deswegen nicht richtig lösen, weil sie den Text, die Instruktion oder die Frage nicht wie von den Testentwickler/innen vorgesehen interpretieren. A. D. Cohen (2011b, S. 303) merkt an, dass es für die Testentwickler/innen evtl. von geringem Interesse sei, wenn das Problem mit der Beantwortung einer Testfrage allein bei einer Testteilnehmerin oder einem Testteilnehmer liege. Sollte sich aber zeigen, dass die falschen Schlussfolgerungen der Lernenden durch eine unklare Textpassage oder eine mehrdeutige Testfrage ausgelöst werden, sollten diese überarbeitet werden (ebd.). Fragen, die die Erforschung von Teststrategien leiten, lassen sich laut Arras wie folgt zusammenfassen: „Wie gehen Prüflinge mit Testaufgaben um? Wie lösen sie die Aufgaben, wie gehen sie konkret vor? Welche Strategien setzen sie ein? Und sind diese Strategien zielführend im Sinne des Testkonstrukts?“ (2013, S. 76). Zentral ist dabei aus der Sicht der Testforschung die Frage, ob die Testaufgaben tatsächlich das elizitieren, was gemessen werden soll (vgl. ebd., S. 77).

Um sicherzustellen, dass die Testaufgaben das gewünschte Verhalten sowie die gesuchten Sprachkompetenzen der Testteilnehmenden zum Vorschein bringen und nicht (bzw. möglichst wenig) für den Einsatz der *test-wiseness*-Strategien anfällig sind, sollen gemäss A. D. Cohen Daten gesammelt werden, die genau das widerspiegeln, was die Testteilnehmenden tun, um auf ihre Antworten zu kommen (2014, S. 901). Nur wenige Strategien der Sprachlernenden wie z.B. das Nachschlagen eines unbekannten Wortes lassen sich aber direkt beobachten. Die meisten Strategien müssen daher entweder aus dem Verhalten der Lernenden abgeleitet oder anhand ihrer Berichte eruiert werden (Griffiths, 2018, S. 64). Bei der Untersuchung der Vorgehensweisen und Strategien der Testteilnehmenden wird deswegen oft – neben Strategiefragebögen (z.B. Phakiti, 2007; Purpura, 1999) oder neuerdings auch dem Einsatz von *eye tracking* (z.B. Suvorov, 2018) – auf introspektive Verfahren zurückgegriffen (vgl. Lumley & Brown, 2005; Arras, 2013; Xi & Sawaki, 2017; Ercikan & Pellegrino, 2017a; Lazaraton, 2017; Alison Green, 1998; Lazaraton & Taylor, 2007; Taylor, 2005; Banerjee, 2004; A. D. Cohen, 2011b; Sasaki, 2014). Zu den introspektiven Verfahren gehören das Laute Denken, *Stimulated Recall* bzw. Lautes Erinnern und retrospektive Befragungen (vgl. Heine & Schramm, 2016).

2.2.4.2 (NEUERE) STUDIEN ZUR ERFORSCHUNG VON TESTSTRATEGIEN MITTELS INTROSPEKTIVER VERFAHREN

Die Strategien der Testteilnehmenden wurden bereits seit den 1980er Jahre mithilfe von introspektiven Verfahren untersucht (siehe A. D. Cohen, 2011b für einen umfassenden Überblick über die Teststrategieforschung). Im Folgenden werden fünf ausgewählte neuere

Studien in chronologischer Reihenfolge ihrer Erscheinung vorgestellt, in denen 1) (u.a.) die Strategien der Testteilnehmenden in einem Lese- und/oder Hörverstehenstest erforscht wurden, und 2) Introspektion zum Einsatz kam, teilweise auch in Kombination mit anderen Methoden.

A. D. Cohen und Upton (2006)

A. D. Cohen und Upton (2006; vgl. auch 2007) untersuchten mittels Lautem Denken Strategien, die von den Testteilnehmenden beim Lösen von Leseverstehensaufgaben eingesetzt wurden. Die Aufgaben stammten aus den Materialien, die die potenziellen Testteilnehmenden mit dem neuen TOEFL-Test vertraut machen sollten. Die Studie fokussierte Strategien, die bei der Bewältigung 1) der bereits bekannten Multiple-Choice-Formate (mit Grundverständnis- und Inferenzitems) und 2) der neuen Aufgabenformate (*selected response reading to learn items*, mit den Formaten *multiple selection* und *drag-and-drop*) zum Einsatz kamen. Die neuen Aufgabenformate wurden entwickelt, um die akademischen Lesefähigkeiten der Testteilnehmenden zu überprüfen. In der Studie sollte insbesondere untersucht werden, ob sich Unterschiede in den Strategien bei der Beantwortung der Grundverständnis-, Inferenz und *reading-to-learn*-Items feststellen lassen.

An der Datenerhebung nahmen 32 fortgeschrittene, erwachsene Englischlernende teil, die u.a. Chinesisch, Japanisch, Koreanisch und Türkisch als Erstsprache hatten. Sie wurden vor der eigentlichen Datenerhebung im Lauten Denken trainiert und wurden für die Teilnahme an der Studie mit einem kleinen Stipendium entschädigt. Den Proband/innen wurden nach dem Zufallsprinzip zwei der sechs Subtests zugeteilt, bestehend aus je einem Text mit 600-700 Wörtern und 12-13 Items. Die Teilnehmenden konnten die Aufgaben in ihrem eigenen Tempo lösen und sollten währenddessen ihre Gedanken verbalisieren (dieser Teil der Datenerhebung dauerte ca. 3 Stunden pro Teilnehmer/in). Die meisten Proband/innen hatten die Wahl, in Englisch oder in ihrer Erstsprache zu verbalisieren, da die erforderlichen Sprachkenntnisse im Forschungsteam vorhanden waren. Die Aufnahmen zu 13 kriterienbasiert ausgewählten Items wurden anschliessend übersetzt/transkribiert und die so entstandenen Lautdenkprotokolle von den trainierten Forschungsassistent/innen codiert. Der dafür eingesetzte Leitfaden wurde aufgrund der Literatur zu Lese- und Teststrategien sowie datenbasiert entwickelt. Die ermittelten Strategien wurden in Lesestrategien (n=28) und Teststrategien, d.h. Testmanagement- (n=28) und *test-wiseness*-Strategien (n=3), unterteilt. Die Codierungen wurden jeweils zuerst von mind. einer anderen Forschungsassistentin oder einem anderen Forschungsassistenten und dann auch von einem der beiden Autoren überprüft. Die festgestellten Unstimmigkeiten wurden während einer Diskussion gelöst. Codiert wurden dabei lediglich Strategien, auf die von den Teilnehmenden bewusst verbal

hingewiesen wurde und nicht solche, von denen die Codierenden vermuteten, dass sie evtl. vorkamen.

Auch wenn es sich bei der Studie um eine primär qualitative Untersuchung handelte, wurde von den Autoren eine Quantifizierung der codierten Daten vorgenommen. Dies wurde gemacht, um gewisse Tendenzen in der Häufigkeit der codierten Strategien aufzeigen zu können. Beim Betrachten der ermittelten Zahlen soll jedoch bedacht werden, dass „this quantitative analysis does not reflect hard and fast numbers, but trends since not all strategies were verbalised or verbalised every time they were used“ (A. D. Cohen & Upton 2006, S. 42). Für die quantitativen Analysen wurden die Hauptvariablen der Studie (d.h. Strategietypen und Itemtypen) als nominale Variablen erfasst und die Gesamtzahl der codierten Strategien ermittelt. Es kam dabei vor, dass zu einer Testfrage mehrmals die gleiche Strategie codiert wurde, was die Anwendung von typischen statistischen Verfahren verhinderte. Um ein System zur Unterscheidung der Häufigkeitsstufen des Auftretens einer Strategie zu ermitteln, wurde die Gesamtsumme der codierten Strategien mithilfe einer *type/token* Analyse in Verhältniswerte umgewandelt, d.h. das Verhältnis der Anzahl der Vorkommnisse jedes Strategietyps im Verhältnis zur Gesamtzahl der Items eines Typs berechnet. Aus dieser Analyse resultierten die Verhältniswerte (*ratio scores*). Diese wurden in einem weiteren Schritt nach Häufigkeit kategorisiert, was nach qualitativen Kriterien und „in a partly empirical and partly intuitive way“ (ebd., S. 42) geschah.

In der quantitativen Analyse wurde auch die Beziehung ermittelt, die verschiedene Strategien miteinander zu haben schienen, und zwar unabhängig davon, wie häufig diese Strategien vorkamen. Die ermittelten spezifischen Muster des Strategiegebrauchs pro Itemtyp wurden anschliessend detaillierter und aus einem qualitativen Blickwinkel analysiert, wobei es hier zu berücksichtigen gilt, dass „strategies for reading and for test-taking invariably cluster together in response patterns, sometimes in sequence and sometimes in groups“ (ebd., S. 43). Die vorgenommene qualitative Beschreibung der eingesetzten Strategien soll daher laut Autoren eine Beschreibung dessen liefern, woraus sich die *response processes* der Teilnehmenden tatsächlich zusammensetzten. Bei der Analyse wurde besonders darauf geachtet, ob die ermittelten Vorgehensweisen der Testteilnehmenden mit den Absichten der Testentwickler/innen übereinstimmten.

Aufgrund der Analyse der eingesetzten Strategien pro Itemtyp konnte u.a. festgestellt werden, dass die Forschungsteilnehmenden die Leseaufgaben im Allgemeinen wie von den TOEFL-Autor/innen gewünscht bewältigten. Die Testteilnehmenden mussten also ihre akademischen Leseverstehensfähigkeiten einsetzen, um die Aufgaben richtig zu bewältigen, d.h. sie mussten die Texte sowohl auf der globalen wie auch auf der lokalen Ebene verstehen und

interpretieren können. Es stellte sich zudem heraus, dass die Teilnehmenden auch die neuen TOEFL-Leseverstehensaufgaben als reine Testaufgaben ansahen: Sie waren primär an der richtigen Lösung interessiert und nicht etwa daran, etwas vom gelesenen Text zu lernen. Die Testteilnehmenden griffen beim Lösen der Aufgaben auf vielfältige Teststrategien zurück, wobei es sich in erster Linie um Testmanagementstrategien handelte. *Test-wiseness*-Strategien wurden relativ selten beobachtet, was laut den Autoren möglicherweise darauf zurückgeführt werden kann, dass es sich bei den Forschungsteilnehmenden um relativ fortgeschrittene Lernende handelte, die den Einsatz von *test-wiseness*-Strategien evtl. gar nicht nötig hatten. Auch wurde das Verhalten der Proband/innen genau beobachtet und sie wurden für ihre Teilnahme an der Studie entschädigt, was laut den Autoren ebenfalls zu einer reduzierten Anzahl an *test-wiseness*-Strategien während der Datenerhebungssituation führen konnte. Weiter zeigte sich auch, dass sich die Testteilnehmenden trotz der deutlichen Unterschiede zwischen den Grundverständnis-, Inferenz- und *reading-to-learn*-Items auf die gleichen Strategien stützten, um diese Items zu beantworten. Die Autoren schlussfolgerten daraus, dass die neuen TOEFL-Aufgaben die Fähigkeiten evozieren, die es den Testteilnehmenden erlauben, mit einer Vielzahl an Lese- und Testaufgaben aus dem akademischen Bereich umzugehen. Der Mehrwert der neuen Aufgaben gegenüber den traditionellen Aufgaben konnte aber aufgrund der Untersuchung nicht gezeigt werden, was laut A. D. Cohen und Upton evtl. auch auf die Reihenfolge der Aufgaben im Testbooklet zurückgeführt werden könnte.

Nikolov (2006)

In ihrem qualitativ und explorativ angelegten Projekt untersuchte Nikolov, von welchen Strategien 12- und 13-jährige ungarische Schüler/innen Gebrauch machen, wenn sie Testaufgaben in Englisch bewältigen. Dank der Analyse der einzelnen Strategien und Fallbeispiele sollten u.a. Einblicke in die Verwendung von Strategien durch junge Lernende gewonnen werden. Gemäss Nikolov lag zum Zeitpunkt der Untersuchung wenig Forschung zum Strategiegebrauch dieser Zielgruppe in einer Testsituation vor, weshalb die Ergebnisse potenziell auch für Testentwickler/innen und Lehrpersonen von Interesse sein sollten.

Das Projekt wurde als eine *follow-up*-Untersuchung zu einer *large-scale*-Studie (n≈30'000) durchgeführt, die im Rahmen eines Bildungsmonitorings die Englischkompetenzen der ungarischen Schüler/innen im Lesen, Hören und Schreiben auf Niveau A1 überprüfte. An der *follow-up*-Studie nahmen 52 Schüler/innen aus öffentlichen Schulen in ganz Ungarn teil. Die Lernenden wurden für die Teilnahme an der Untersuchung von ihren Lehrpersonen ausgewählt. Die Stichprobe setzte sich aus Lernenden mit einem hohen (n=22), mittleren (n=20) und niedrigen (n=10) Lernniveau (gemäss ihrer Schulnoten) zusammen. Die

Schüler/innen bearbeiteten nach dem anfänglichen Training im Lauten Denken ausgewählte Testaufgaben aus der *large-scale*-Studie zum Lesen und Schreiben und verbalisierten dabei ihre Gedanken in Ungarisch. Die Daten wurden von insgesamt 26 Studierenden erhoben und transkribiert, die einen Fortgeschrittenenkurs zum unterrichtsgesteuerten Spracherwerb für Kinder besuchten. Die Transkripte wurden unabhängig voneinander von der Autorin und von einer oder einem der Studierenden codiert. In der anschliessenden Diskussion wurden Codierungen aufeinander abgestimmt und die so entstandene Liste der identifizierten Strategien wurde mit den Strategietaxonomien von Anderson (1991) und Oxford (1990) verglichen. Danach wurden alle Daten von der Autorin nochmals gesichtet und mit dem endgültigen Codierleitfaden nachcodiert. Die in den Daten gefundenen Strategien wurden dann in drei Gruppen eingeteilt, d.h. in 1) metakognitive Strategien (n=16, z.B. Ausschluss oder Überprüfung von Antwortoptionen), 2) soziale und affektive Strategien (n=5, z.B. positive oder negative Selbstevaluation, Bitten um Hilfe) und 3) kognitive Strategien (n=13, z.B. Vorlesen des Inputtextes, Raten, sich auf Kognaten verlassen).

Nikolov berichtet, das sich die Codierung und Zählung der Strategien oft als schwierig erwies: „For example, reading out or translating text meant an item (a sentence or short paragraph) in the case of good performers and isolated words with low performers. If counted by turns, low performers applied these strategies more frequently, but if counted by items, the number is the same, although the quality of the strategy is different“ (2006, S. 22). Bei der Darstellung der Resultate entschied sich die Autorin wohl deshalb gegen eine Quantifizierung der ermittelten Strategien. Stattdessen präsentierte sie die Resultate in Form 1) einer Kurzanalyse (*short analysis*) der Vorgehensweisen von vier Lernenden (ausgewählt wurden dafür zwei Lernende mit den besten und zwei mit den schlechtesten Testergebnissen) sowie 2) einer qualitativen Beschreibung von Tendenzen, die in den Daten gefunden wurden.

Die Vorgehensweisen der vier ausgewählten Lernenden zeigten, dass sie an die Testaufgaben unterschiedlich herangingen und auch verschiedenartige Kombinationen von Strategien einsetzten. Ein Schüler betrachtete den Text als Ganzes und erriet die Bedeutung in einer selbstbewussten Art und Weise aus dem Kontext, während drei andere Lernende mehr zögerten, weniger sicher in Bezug auf ihre Leistung waren und beim Bearbeiten der Testaufgaben eher den *bottom-up*-Ansatz verfolgten. Bei den zwei *low performers* wurde beobachtet, dass sie negativ eingestimmt waren, kein Selbstvertrauen zu haben schienen und auf der Ebene der Einzelworte an die Texte herangingen.

Aufgrund der Analyse aller Daten stellte Nikolov in Bezug auf den Strategienegebrauch zwei Haupttendenzen fest: 1) Einige Schüler/innen setzten die gleichen Strategien beim Bearbeiten aller Testaufgaben ein, während andere nur gelegentlich auf dieselben Strategien

zurückgriffen (daher zeigt die Auswertung der Häufigkeiten der Strategien zwar einige aussagekräftige Trends auf, sie gibt aber keine Einblicke in die tatsächlichen Prozesse und individuellen Unterschiede); und 2) die Strategien traten nicht isoliert, sondern meistens in Kombination auf. Die Daten zeigten auch viele individuelle Unterschiede zwischen den einzelnen Schüler/innen: Diese setzten nicht nur unterschiedliche Strategien ein, die in manchen Fällen zur korrekten und in manchen zur falschen Antwort führten, sondern verbalisierten auch unterschiedlich viel. Es schien dabei auch, dass „good performers most probably benefit more from the use of strategies, and this is in line with previous research (Anderson, 1991; Purpura, 1999)“ (Nikolov, 2006, S. 42). Die Erkenntnisse zeigten laut der Autorin auch, dass die Aufgaben die vom Test gesuchten Fähigkeiten der Lernenden elizitierten, wobei die Testautor/innen in der Zukunft mehr darauf achten sollten, dass die Aufgaben nicht nur aufgrund des Hintergrundwissens oder eines Matching von Wörtern und Strukturen gelöst werden können.

Rossa (2012)

In seinem Dissertationsprojekt untersuchte Rossa „die Validität der Testaufgaben zum Hörverstehen in der Fremdsprache Englisch, die in der DESI-Studie eingesetzt wurden“ (Rossa, 2012, S. 2; für weiterführende Informationen zur DESI-Studie und der Hörverstehensaufgaben vgl. Nold & Rossa, 2007; Nold et al., 2008). Insbesondere wollte seine explorative Studie Antworten darauf liefern, 1) welche mentalen Prozesse der Testteilnehmenden mit der Bearbeitung der Testaufgaben einhergehen, 2) inwiefern diese mentalen Prozesse den relevanten Facetten des Testkonstrukts entsprechen und 3) wie die Aktivierung konstruktrelevanter und konstruktirrelevanter Strategien und Prozesse mit der erfolgreichen bzw. fehlerhaften Aufgabenlösung zusammenhängt. Im Unterschied zu den anderen vier Studien, die in diesem Kapitel näher vorgestellt werden, interessierte sich Rossa sowohl für die eher bewusst ausgeführten Strategien als auch für die grösstenteils automatisierten kognitiven Prozesse der Testteilnehmenden beim Lösen der Testaufgaben (vgl. Rossa, 2012, S. 87).

In der Studie wurde ein *mixed-methods*-Forschungsdesign eingesetzt, das Rossa nach Creswell et al. (2007) als „concurrent triangulation“ bezeichnet (Rossa, 2012, S. 84). Um die Forschungsfragen zu beantworten, analysierte und interpretierte er 1) die verbalen Daten der Testteilnehmenden, die aus Lautdenkprotokollen während des Aufgabenlösens sowie aus *Stimulated-Recall*-Interviews und mündlichen Zusammenfassungen der Inputtexte im Nachhinein bestanden, und 2) die psychometrischen Daten zu den untersuchten Testaufgaben.

An der Studie nahmen 18 Schüler/innen (15-16 Jahre alt) aus dem Ruhrgebiet teil, die aus der Gesamtstichprobe von fünf Klassen (n=121)²² nach der Methode des „absichtsvollen und zufälligen Samplings von Extremfällen“ (ebd., S. 87) ausgewählt wurden. Zunächst wurden dafür die sprachlichen Kompetenzen und die „auf das Unterrichtsfach Englisch bezogene Einstellungen und Wissensbestände“ (ebd., S. 91) der Schüler/innen aus der Gesamtstichprobe mittels C-Tests aus der DESI-Hauptstudie (vgl. Harsch & Schröder, 2007) erfasst. Aufgrund der Resultate wurden anschliessend aus allen Schulklassen der Gesamtstichprobe neun Schüler/innen mit besonders hoch (Teilstichprobe „HI“) und neun Schüler/innen mit besonders schwach (Teilstichprobe „LO“) ausgebildeter allgemeiner Englischkompetenz zufällig ausgewählt. Dadurch wurde u.a. „die Erfassung eines möglichst breiten Spektrums unterschiedlicher kognitiver Prozesse“ (Rossa, 2012, S. 92) angestrebt, die wahrscheinlich mit der unterschiedlich ausgeprägten Sprachkompetenz der Schüler/innen einhergingen. Fast alle ausgewählten Schüler/innen hatten Deutsch als Erstsprache, bis auf eine Schülerin der Zufallsstichprobe „HI“ mit Polnisch und drei Schüler/innen der Zufallsstichprobe „LO“ mit Serbokroatisch, Türkisch bzw. Dari als L1.

Am Vortag der Datenerhebung wurden die Schüler/innen in Kleingruppen von 2-3 Personen während eines einstündigen und standardisierten Trainings in das Verfahren des Lauten Denkens eingeführt. Am Tag der eigentlichen Datenerhebung konnten die Lernenden das Laute Denken nochmals kurz üben, bevor sie die Hörverstehensaufgaben lösten und dabei ihre Gedanken verbalisierten. Bei den Aufgaben handelte es sich um 16 von insgesamt 52 Items aus der DESI-Hauptstudie, „die sich auf beide Textsorten (Dialog und Erzählung) beziehen und mit Ausnahme der höchsten Stufe [...] alle in DESI ausgewiesenen Kompetenzniveaus abdecken“ (ebd., S. 60). Einer weiteren Stelle im Buch ist zu entnehmen, dass die Datenerhebung in zwei Phasen gegliedert war und dass „der Forscher bestrebt [war], die Informanten in den Gesprächsphasen vor den zwei Aufgabenblöcken (acht kurze Dialogausschnitte, eine knapp zweiminütige Erzählung) hinsichtlich ihrer Verbalisierungstechniken positiv zu bestärken und wiederholt zu betonen, dass in dieser Untersuchung die Angemessenheit der Aufgaben und nicht etwa die Leistung der Testteilnehmer überprüft wird“ (ebd., S. 105). Die Bearbeitung der Testaufgaben dauerte pro Aufgabensatz ca. 20 Minuten.

Kurz nach der Bearbeitung der Testaufgaben hörten sich die Teilnehmenden die Hörtexte nochmals an und wurden gebeten, diese auf Deutsch zusammenzufassen. Dadurch sollte das Textverständnis der Schüler/innen zum Zeitpunkt des Aufgabenlösens besser als nur auf der

²² Es handelte sich dabei um Klassen der Schulformen Hauptschule, Realschule, Gesamtschule und Gymnasium, wobei im Gymnasium zusätzlich eine Klasse des bilingualen Zweigs dazugehörte (Rossa, 2012, S. 92–93).

Grundlage der teilweise fragmentarischen Verbalprotokolle eingeschätzt werden können. Im Anschluss daran wurden die Teilnehmenden in *Stimulated-Recall*-Interviews „zum Verlauf ihrer aufgabenbezogenen Kognitionen befragt“, was ihnen ermöglichte, „im Sinne einer kommunikativen Validierung in die ersten Schritte der Analyse ihrer Daten mit einbezogen zu werden“ (ebd., S. 101–102). In den *Stimulated-Recall*-Interviews wurden nur diejenigen Aufgaben besprochen, die die Schüler/innen als besonders anspruchsvoll einschätzten. Als Stimulus dienten die abgespielten Ausschnitte der Lautdenkprotokolle zu den ausgewählten Aufgaben und die jeweiligen Hörtexte. Die Schüler/innen wurden in den Interviews gefragt, was die ausgewählten Aufgaben für sie schwierig machte, was ihnen darin unklar war und ob sie wissen, warum ausgerechnet diese Aufgaben ihnen Schwierigkeiten bereiteten. Danach wurden ihnen die Transkripte der Audioaufnahmen vorgelegt, mit der Bitte, jene Textstellen zu markieren, mit denen sie Mühe hatten.

Die Verbalprotokolle wurden anschliessend transkribiert und segmentiert. Die Segmentierung orientierte sich an der Untersuchungseinheit der „Proposition“²³ nach Kintsch und van Dijk (1978) und als maximale Grösse eines Datensegments wurden zwei Propositionen definiert. Die drei Typen der verbalen Daten (Lautdenkprotokolle, *Stimulated-Recall*-Interviews und mündliche Zusammenfassungen) wurden danach getrennt und den zwei Teilstichproben zugeordnet, wodurch sechs Datensätze entstanden. Anschliessend wurden die Daten mit einem Kategoriensystem codiert, welches aufgrund von relevanten theoretischen und empirischen Arbeiten sowie induktiv entstand und 156 Kategorien beinhaltete, die 12 Oberkategorien zugeordnet wurden. Die *Stimulated-Recall*-Interviews und mündlichen Zusammenfassungen wurden genutzt, um die Codierungen der Lautdenkprotokolle „zu überprüfen, zu bestätigen oder zu revidieren und gegebenenfalls zu vervollständigen“ (Rossa, 2012, S. 111). Als Oberkategorien für die Codierung der mentalen Prozesse der Schüler/innen wurden folgende Kategorien definiert:

1. Kognitive Prozesse: Textinformationen aus dem Gedächtnis abrufen (n=8, z.B. Textfragmente aus dem phonologischen Arbeitsgedächtnis abrufen);
2. Kognitive Prozesse: Inferenzen generieren (n=4, z.B. elementare oder konfabulierende Interferenzen);
3. Fehler bei der Konstruktion des Textverständnisses (n=5, z.B. Worterkennungsfehler);

²³ „Propositionen sind als Einheiten zu verstehen, die die Ergebnisse mentaler Prozesse bei der Verarbeitung von Bedeutungen widerspiegeln [...] und sowohl semantische Beziehungen auf der Satzebene als auch funktionale Beziehungen zwischen Sätzen und Textsegmenten repräsentieren [...]. Eine Proposition besteht in der Notation computerbasierter psycholinguistischer Modelle aus einem Prädikat (Relationskonzept: z. B. Verben, Adjektive, Konnektoren) und mindestens einem Argument (Inhaltskonzept: z. B. Nomen, eingebettete Präposition), das eine semantische Rolle erfüllt (z. B. Agens, Patiens, Attribut)“ (Rossa, 2012, S. 26–27).

4. Metakognitive Strategien: Überprüfung der eigenen Verstehensprozesse (n=3, z.B. Nachvollziehen der eigenen Verstehensprozesse);
5. Metakognitive Strategien: Evaluieren der eigenen Verstehensprozesse (n=2, darunter a) Ursachen erfolgreicher und b) unsicherer und gescheiterter Aufgabenlösungen);
6. Das Phänomen Aufgabenschwierigkeit aus der Perspektive der Testteilnehmenden (n=4, z.B. Merkmale der Items);
7. Affektive Strategien (n=2, z.B. emotionale Reaktionen); und
8. Teststrategien²⁴ (n=3, z.B. Raten).

Das Kategoriensystem wurde zuerst an einem Drittel der Daten erprobt und überarbeitet. Auch wurde ein Teil der Lautdenkdaten zusätzlich von einer studentischen Mitarbeiterin codiert, die vorher ein einstündiges Training durchlief. Die ermittelte Übereinstimmung der codierten Segmente betrug 78,62%, was der Autor als einen vorsichtigen „Hinweis auf die Anwendbarkeit der Kernkategorien des Kodierschemas“ interpretierte (ebd., S. 115).

In den verbalen Daten wurden insgesamt 4121 Codierungen vorgenommen. Die Datenanalyse zeigte eine breite Palette von kognitiven, metakognitiven und affektiven Verarbeitungsprozessen der Schüler/innen beim Lösen der Testaufgaben auf. Rossa stellt diese entlang seines Kategoriensystems ausführlich dar, illustriert sie anhand prototypischer Datenauszüge (ebd., S. 117–192) und bezeichnet sie als „umfassende – wenn auch nicht lückenlose – Einblicke in das sich entwickelnde Textverständnis und in die mentale Auseinandersetzung mit den Multiple-Choice-Items“ (ebd., S. 231). Die Resultate der qualitativen Analyse wurden darüber hinaus quantifiziert und „mit Blick auf Zusammenhänge mit den Testleistungen der Informanten und den psychometrischen Parametern der untersuchten Testaufgaben untersucht und inferenzstatistisch überprüft“ (ebd., S. 91). Die Ergebnisse dieser Analyse deuten u.a. darauf hin, dass es generell gelang, die relevanten Aspekte des Testkonstrukts mit den eingesetzten Aufgaben zu erfassen und dass die konstruktrelevanten Sprachverarbeitungsprozesse auch „konsistent stabile Zusammenhänge mit der Testleistung“ zeigen (ebd., S. 236). Die Aktivierung konstruktirrelevanter Prozesse und Strategien führte hingegen üblicherweise zur Auswahl einer falschen Antwortoption. Der Vergleich der beiden Teilstichproben bestätigte laut Rossa die Ergebnisse der Strategieforschung, „wonach Sprachverwender auf niedrigeren Stufen der Entwicklung ihres Sprachwissens in der L2 stärker unter Verstehensproblemen auf den hierarchieniedrigeren Ebenen der Sprachverarbeitung leiden und auf diese entweder mit verstärkter Aufmerksamkeit für die bottom-up Prozesse oder mit der Generierung elaborierender

²⁴ Rossa versteht darunter konstruktirrelevante Strategien, die A. D. Cohen (u.a. 2014) als *test-wiseness*-Strategien bezeichnen würde.

Hypothesen zur Rekonstruktion der bruchstückhaften Textbasis reagieren“ (ebd., S. 237). Die Ergebnisse lassen den Autor auch vermuten, dass es mit den eingesetzten Aufgaben nicht ausreichend gelang, die Hörverstehensfähigkeiten derjenigen Schüler/innen zu erfassen, deren Englischkompetenzen am schwächsten entwickelt waren. Ausserdem wurden bei der Analyse der Verbalprotokolle einzelne Items identifiziert, die Validitätsprobleme aufweisen und deswegen überarbeitet oder aus dem Aufgabenpool ausgeschlossen werden sollten (ebd., S. 236).

L. Gu und So (2017)

Bei dieser Studie handelte es sich um eine explorative Untersuchung der Strategien, auf die Englischlernende auf verschiedenen Leistungsniveaus (*proficiency levels*) beim Bearbeiten der Lese- und Hörverstehensaufgaben aus dem neuen TOEFL-Primary-Tests für junge Lernende ab 8 Jahren zurückgreifen. Die Studie wurde anlässlich der Pilotierung der neuen, prototypischen Testaufgaben durchgeführt und sollte sicherstellen, dass das Design der Itemtypen für die Zielgruppe geeignet ist. Im Hauptfokus stand die Frage der Konstruktvalidität: Aufgrund der Strategien der Testteilnehmenden sollte untersucht werden, ob mit den Items tatsächlich das beabsichtigte Konstrukt gemessen wird, d.h. die Lese- und Hörverstehensfähigkeiten der Kinder in Englisch als Fremdsprache. Über die Teststrategien von so jungen Lernenden lagen bisher gemäss L. Gu und So keine empirischen Erkenntnisse vor. Die Autorinnen erhofften sich daher, neben den Hinweisen zum Funktionieren der prototypischen TOEFL-Testaufgaben auch allgemein mehr über die Vorgehensweisen der Kinder beim Lösen von Testaufgaben zu erfahren, um auf dieser Grundlage nützliche Anregungen sowohl für die Gestaltung geeigneter Beurteilungen wie auch weiterer möglichen Strategieuntersuchungen für diese Zielgruppe formulieren zu können.

An der Studie nahmen 16 Kinder mit Chinesisch als L1 teil, die in China ein ausserschulisches Englischlernprogramm besuchten und noch nie in einem englischsprachigen Land gelebt hatten. Für die Stichprobe wurden Schüler/innen unterschiedlichen Alters (6 bis 11 Jahre alt) und mit unterschiedlich ausgeprägten Englischkenntnissen (gemäss Einschätzungen ihrer Lehrpersonen) ausgewählt. Die individuellen Datensitzungen dauerten eine Stunde pro Schüler/in. Eine 11-jährige Schülerin konnte in dieser Zeit sowohl den Lese- als auch Hörverstehensteil des Prototyp-Tests lösen, alle anderen Lernenden bewältigten währenddessen entweder den Lese- (n=8) oder den Hörverstehensteil (n=7). Die Forscherinnen gingen in Anbetracht des Alters der Proband/innen davon aus, dass das Laute Denken die Kinder mental anstrengen würde und sie sich deshalb bei der Datenerhebung unwohl fühlen würden. Als Datenerhebungstechnik wurde deswegen die retrospektive Befragung gewählt. Die Interviews wurden auf Chinesisch von einer der Forscherinnen

durchgeführt, deren Erstsprache ebenfalls Chinesisch war. Lernende, die die Leseverstehensaufgaben bewältigt hatten, wurden während dieser Tätigkeit nicht unterbrochen und erst danach befragt; beim Hörverstehen fand hingegen die Befragung nach jeder Aufgabe statt. Die Lernenden wurden in den Interviews gefragt, warum sie die gewählte Antwortoption für richtig hielten, warum sie die alternative Antwortoptionen ablehnten und ob sie im Nachhinein ihre Antwort ändern würden. Alle Interviews wurden anschliessend ins Englische übersetzt und transkribiert.

Die transkribierten Interviews wurden doppelt codiert und die eventuellen Unterschiede während der anschliessenden Diskussion geklärt. Der Kategorienleitfaden wurde ursprünglich auf der Grundlage der Fachliteratur erstellt und dann im Prozess der Codierung überprüft und modifiziert. Die codierten Strategien wurden in einem weiteren Schritt gemäss dem Vorschlag von A. D. Cohen (u.a. 2011a, 2012; A. D. Cohen & Upton 2006) in Lernerstrategien (n=12) und Teststrategien unterteilt, wobei sich die letzteren aus den Testmanagementstrategien (n=11) und *test-wiseness*-Strategien (n=8) zusammensetzten. Um die eventuellen Zusammenhänge zwischen den ermittelten konstruktrelevanten und konstruktirrelevanten Strategien und dem Sprachniveau der Lernenden feststellen zu können, wurden die Schüler/innen aufgrund der Testresultate, die sie bei der Untersuchung erzielten, in *low*-, *medium*- und *high-proficiency*-Gruppen eingeteilt: Die Schüler/innen der *low-proficiency*-Gruppe beantworteten weniger als 50% der Items korrekt, die *medium-proficiency*-Schüler/innen hatten zwischen 50% und 75%, die *high-proficiency*-Schüler/innen über 75% der Antworten richtig. Aufgrund der kleinen Stichprobe wurden keine Signifikanztests durchgeführt; stattdessen wurden die Häufigkeiten der ermittelten Strategien über die drei Leistungsniveaus hinweg und pro Fertigkeit verglichen.

Aufgrund der Berichte der Schüler/innen konnten insgesamt 308 Codierungen über alle Lese- und Hörverstehensitems hinweg von den Autorinnen vorgenommen werden. Am häufigsten wurden in den Daten Lernerstrategien (123 Codierungen, d.h. 40% aller Codierungen) sowie *test-wiseness*-Strategien (114 Codierungen, d.h. 37% aller Codierungen), und etwas weniger oft Testmanagementstrategien (71 Codierungen, d.h. 23% aller Codierungen) ermittelt. L. Gu und So listen in ihrem Artikel die jeweiligen Strategien auf und illustrieren die häufigsten mit einem Beispiel. Die Analyse der Häufigkeit der verwendeten Strategien pro Fertigkeit und Gruppe zeigte, dass die drei Lernenden mit den besten Resultaten im Lese- bzw. Hörverstehen keine *test-wiseness*-Strategien beim Hören und nur wenige (insgesamt 4) *test-wiseness*-Strategien beim Leseverstehen einsetzten. Die Schüler/innen der *medium*- und *low-proficiency*-Gruppen berichteten hingegen über einen häufigeren Einsatz von *test-wiseness*-Strategien. Über unterschiedliche Fertigkeiten hinweg betrachtet zeigte sich ausserdem, dass die leistungsstarke Gruppe die Lernerstrategien häufiger beim Lese- als beim Hörverstehen

nutzte und dass die Häufigkeit der Lernerstrategien beim Leseverstehen mit steigender Kompetenz zunahm. Beim Hörverstehen griffen hingegen die Schüler/innen der mittleren Leistungsgruppe am häufigsten auf die Lernerstrategien zurück.

L. Gu und So kommen aufgrund ihrer Resultate zu dem Schluss, dass die Mehrheit der Strategien, die sie in ihren Daten finden konnten, als konstruktrelevant angesehen werden kann, und betrachten dies als eine ermutigende Nachricht für die Testentwickler/innen. Sie merken aber auch an, dass der häufigere Gebrauch von *test-wiseness*-Strategien durch die Schüler/innen aus den *low*- und *medium-proficiency*-Gruppen darauf hindeuten könnte, dass diese Items für diese Gruppe zu anspruchsvoll waren und daher nicht gut geeignet sind, um zwischen Lernenden auf niedrigeren Leistungsniveaus zu diskriminieren. Ihre Ergebnisse betrachten sie daher als übereinstimmend mit der Entscheidung der Testentwickler/innen, in der Endfassung des TOEFL-Primary-Tests die Hör- und Leseverstehenstests auf zwei Levels anzubieten.

Suvorov (2018)

Suvorov untersuchte Teststrategien, die Englischlernende beim Lösen von ausgewählten Aufgaben (insgesamt 58 Items) zum Leseverstehen, Hörverstehen und zur Grammatik aus dem Michigan English Test (MET) einsetzen. Die Studie hatte zum Ziel, herauszufinden, auf welche Strategien die Testteilnehmenden bei diesem Test zurückgriffen, ob sich beim Lösen von insgesamt fünf verschiedenen Itemtypen Unterschiede in den angewandten Strategien zeigen lassen und inwieweit durch den Einsatz von *test-wiseness*-Strategien konstruktirrelevante Varianz verursacht wird.

Die MET-Aufgaben, die normalerweise auf Papier gelöst werden und aus 135 Multiple-Choice-Items pro Testbooklet bestehen, wurden für Untersuchungszwecke auf das Kursmanagementsystem Moodle übertragen und am Computer gelöst. An der Studie nahmen 15 Student/innen teil, die im Durchschnitt seit ca. 11 Jahren Englisch lernten und Englischkurse an der Universität Hawai'i at Manoa besuchten. Ihre Teststrategien wurden unter Verwendung des Konvergenzmodells des Datentriangulationsdesigns nach Creswell und Plano Clark (2006) mittels Analyse von 1) *eye-tracking*-Daten, 2) retrospektiven Befragungen mit Aufnahmen der Augenbewegungen als Stimulus (Suvorov spricht hier von „cued retrospective reporting“, auch bekannt als „eye-movement supported verbal retrospection“ oder „post-experience eye-tracked protocol“ (Suvorov, 2018, S. 5)) und 3) erzielten Testresultaten untersucht.

Die Proband/innen lösten die 58 MET-Items in individuellen, ca. zweistündigen Datensitzungen und bekamen für die Teilnahme an der Studie einen kleinen Gutschein. Sie

sahen jeweils ein Item auf einmal und konnten selbst bestimmen, wann sie zum nächsten Item übergehen bzw. die Audioaufnahmen anhören wollten, sie konnten aber im Test nicht zurückgehen. Nach dem Lösen aller Items eines Itemtyps wurden den Testteilnehmenden die Aufzeichnungen ihrer Augenbewegungen gezeigt, mit der Aufforderung, zu verbalisieren, 1) welche Strategien sie beim jeweiligen Item einsetzten, 2) warum sie eine bestimmte Antwortoption wählten, und 3) ob sie sicher seien, dass die gewählte Antwort auch richtig ist. Die Interviews wurden in Englisch durchgeführt, aufgenommen (Audio- und Bildschirmaufnahmen) und transkribiert.

Für die Untersuchung der Teststrategien wurde zuerst eine „qualitative scanpath analysis of eyetracking data via visual inspection“ durchgeführt (ebd., S. 7). Für jedes Testitem wurden also die zeitliche und räumliche Abfolge der Augenbewegungen eines Testteilnehmers oder einer Testteilnehmerin analysiert und dazu eine kurze Beschreibung angefertigt. Anschliessend wurden in den transkribierten Interviews die Testmanagement- und *test-wiseness*-Strategien (A. D. Cohen, 2014) codiert, die von den Testteilnehmenden erwähnt wurden. Eine Strategie wurde als *test-wiseness*-Strategie betrachtet, wenn 1) der/die Testteilnehmende nicht einmal ansatzweise erklären konnte, warum er/sie eine bestimmte Antwortoption auswählte und warum diese Antwortoption richtig sei; 2) der/die Testteilnehmende nicht sicher war, ob die gewählte Antwortoption richtig war; oder 3) wenn die Strategie nicht eindeutig mit den intendierten kognitiven Prozessen übereinstimmte, die durch das bestimmte Item aktiviert werden sollten. Die Ergebnisse der Verbalprotokollanalyse wurden in einem weiteren Schritt mit den Ergebnissen der Analyse der Augenbewegungen der jeweiligen Proband/innen zusammengeführt. Bei Diskrepanzen zwischen den beiden Datenquellen wurden bei der Bestimmung der Strategie die *scanpath*-Daten den verbalen Daten vorgezogen. Die Tatsache, dass die ermittelten Strategien teilweise auch in Clustern vorkamen, wurde dabei nicht berücksichtigt. Um den Einfluss der *test-wiseness*-Strategien auf die konstruktirrelevante Varianz zu untersuchen, wurden die Ergebnisse der Augenbewegungs- und Transkriptanalysen sowie die erzielten (d.h. beobachteten) Testpunkte quantitativ analysiert. Es wurde dafür zuerst die Anzahl der korrekt gelösten Items und dann die Anzahl der mithilfe von *test-wiseness*-Strategien gelösten Items ermittelt. Anschliessend wurden die *adjusted scores* berechnet, d.h. von der Anzahl der korrekt gelösten Items wurde die Anzahl der mithilfe von *test-wiseness*-Strategien korrekt gelösten Items abgezogen. Aufgrund des statistischen Zusammenhangs zwischen diesen *adjusted scores* und den eigentlichen Testresultaten wurden Schlussfolgerungen über den Einfluss von *test-wiseness*-Strategien auf die beobachteten Testresultate und über den Beitrag dieser Strategien zur konstruktirrelevanten Varianz gezogen.

Die Resultate der Analyse von Augenbewegungen und verbalen Daten zeigten, dass die Testteilnehmenden bei der Beantwortung der Testaufgaben auf vielfältige Testmanagementstrategien (n=20, z.B. Lesen der Antwortoptionen beim Hören des Inputtextes, Suche nach einem Schlüsselwort im Inputtext und ausschliessliche Lektüre des Textteils mit diesem Schlüsselwort) und *test-wiseness*-Strategien (n=7, z.B. zufälliges Raten, Wählen einer anders aussehenden Antwortoption) zurückgriffen, wobei es zwischen den Teilnehmenden individuelle Unterschiede gab. Die ermittelten Strategien wurden bei der Darstellung der Resultate detailliert beschrieben und pro Itemtyp aufgeführt. Es stellte sich heraus, dass die Mehrheit der Teststrategien im Zusammenhang mit allen Itemtypen verwendet wurden, aber einige der Strategien nur bei bestimmten Itemtypen (und Fertigkeiten) beobachtet werden konnten. Dies ist laut Suvorov teilweise auf bestimmte Eigenschaften des Testdesigns zurückzuführen, z.B. auf die Komplexität der Inputtexte, die Länge der Antwortoptionen oder die fehlenden Zeitvorgaben im Test. Des Weiteren deckte die nähere Betrachtung der *test-wiseness*-Strategien auf, dass alle Testteilnehmenden solche Strategien verwendeten (zwischen 4 und 27 Mal pro Testteilnehmer/in) und mit deren Hilfe zwischen 1 und 10 Items richtig (aber auch zwischen 1 und 17 Items falsch) lösen konnten. Am häufigsten konnten die *test-wiseness*-Strategien bei einem Probanden bzw. einer Probandin mit deutlich unterdurchschnittlichen Englischkompetenzen im Vergleich zum Rest der Stichprobe beobachtet werden. Die Ergebnisse einer statistischen Analyse zeigten ausserdem, dass der Einsatz von *test-wiseness*-Strategien einen statistisch signifikanten Effekt auf die beobachteten Testergebnisse hatte und damit zur konstruktirrelevanten Varianz beitrug.

Aufgrund der Resultate seiner Untersuchung kommt Suvorov zu dem Schluss, dass die Testentwickler/innen den potenziellen Einfluss ihres Testdesigns auf die Verwendung von bestimmten Teststrategien nicht ausser Acht lassen sollten: Einige der vom Testdesign evozierten Strategien können sich nämlich auf das gemessene Konstrukt auswirken. Es soll daher sichergestellt werden, dass dies auch in einer vom Test gewünschten Art und Weise geschieht. Ausserdem könnten die Resultate auf eine mögliche Interaktion zwischen dem Kompetenzniveau und den eingesetzten Strategien hindeuten, d.h. es könnte sein, dass Lernende mit geringeren Sprachkompetenzen eher auf *test-wiseness*-Strategien zurückgreifen. Die Untersuchung zeigte auch, dass sich der Einsatz von *eye-tracking* und retrospektiver Befragung als komplementäre Datenerhebungsmethoden für Testvalidierungszwecke bewährte.

Zusammenfassung der Studien

Die fünf Beispielstudien untersuchten u.a. mittels unterschiedlicher introspektiven Verfahren (Lautes Denken, *Stimulated Recall* und retrospektive Befragungen) die Teststrategien, die von Englischlernenden verschiedenen Alters, auf verschiedenen Niveaus und beim Lösen von Testaufgaben zu verschiedenen Teilkompetenzen angewandt wurden. Bei den Studien, in denen Lautes Denken als Datenerhebungsmethode eingesetzt wurde, ist der Datenerhebung jeweils ein Training im Lauten Denken vorangegangen. Auch wurden die erhobenen Daten in allen Studien transkribiert (und auch ins Englische übersetzt, wenn es nötig war) und codiert - dort, wo es möglich war, auch doppelt. Allen Beispielstudien ist der erhebliche Aufwand zu entnehmen, der bei der Erhebung, Aufbereitung, Codierung und Analyse von introspektiven Daten in Kauf genommen werden muss.

Interessant ist, dass die jeweiligen Autor/innen unterschiedliche Vorgehensweisen bei der Datenanalyse wählten. Dies zeigt sich schon bei der Entscheidung über die Codierung der Strategien: Drei Studien (A. D. Cohen & Upton, 2006; L. Gu & So, 2017; Suvorov, 2018) orientierten sich bei der Codierung von Strategien an die Klassifizierung von A. D. Cohen (vgl. u.a. A. D. Cohen & Upton, 2006; A. D. Cohen, 2012). In den Studien von Nikolov (2006) und Rossa (2012) wurde hingegen bei der Konstruktion der Kategorien zur Erfassung der Teststrategien auf Vorschläge aus der Lernstrategieforschung zurückgegriffen (u.a. O'Malley et al., 1989 (Rossa); Oxford, 1990 (Nikolov)). Rossa ergänzte diese u.a. um Kategorien, die er aufgrund der Erkenntnisse aus der Fachliteratur zum Hörverstehen, Textverstehen und zu den Teststrategien modellierte (vgl. 2012, S. 112–113). Die Tatsache, dass die Analyse und Darstellung der Daten in allen fünf Studien jeweils anders ausfielen, ist aber u.a. auf die unterschiedlichen Forschungsfragen zurückzuführen, die im Zentrum der Untersuchungen standen.

Erwähnenswert ist dabei auch, dass sich bis auf Nikolov alle Autor/innen dafür entschieden, die codierten Strategien auch quantitativen Analysen zu unterziehen. Diese quantitativen Analysen sind jedoch – wie die Autor/innen teilweise selber anmerken – mit Vorsicht zu genießen, und zwar wegen der jeweils kleinen Stichproben und dem Vorkommen der Strategien in Strategieclustern- bzw. -ketten, nicht zuletzt aber auch wegen der Tatsache, dass die Teilnehmenden beim Lauten Denken bekannterweise nicht alles verbalisieren, was sie denken (vgl. Heine & Schramm, 2016; Knorr & Schramm, 2012).

Im Folgenden werden die fünf oben dargestellten Studien in einem vergleichenden, tabellarischen Überblick veranschaulicht.

Tabelle 1: Vergleichender Überblick über die neueren Studien zur Überprüfung der Teststrategien mittels introspektiver Verfahren.

	A. D. Cohen & Upton (2006)	Nikolov (2006)	Rossa (2012)	L. Gu & So (2017)	Suvorov (2018)
n	32	52	18	16	15
Zielgruppe	<ul style="list-style-type: none"> Fortgeschrittene, erwachsene Englischlernende mit verschiedenen L1 	<ul style="list-style-type: none"> 12- und 13-jährige Schüler/innen aus Ungarn 	<ul style="list-style-type: none"> 15- und 16-jährige Schüler/innen aus Deutschland 	<ul style="list-style-type: none"> 6- bis 11-jährige Kinder aus China 	<ul style="list-style-type: none"> Fortgeschrittene, erwachsene Englischlernende mit verschiedenen L1
Untersuchte Testaufgaben	<ul style="list-style-type: none"> TOEFL-Test, Trainingsmaterial Testteil: Leseverstehen 	<ul style="list-style-type: none"> Ausgewählte Testaufgaben aus dem ungarischen Bildungsmonitoring, Fremdsprache Englisch Testteile: Leseverstehen und Schreiben 	<ul style="list-style-type: none"> Ausgewählte Testaufgaben aus der DESI-Studie (16 Items), Fremdsprache Englisch Testteil: Hörverstehen 	<ul style="list-style-type: none"> Aufgaben aus dem neuen TOEFL-Primary-Test für junge Lernende ab 8 Jahren, Fremdsprache Englisch Testteile: Lese- und Hörverstehen 	<ul style="list-style-type: none"> Aufgaben aus dem Michigan English Test (58 Items) Testteile: Leseverstehen, Hörverstehen, Grammatik
Methoden der Datenerhebung	<ul style="list-style-type: none"> Lautes Denken in der L1 der Proband/innen bzw. auf Englisch 	<ul style="list-style-type: none"> Lautes Denken in der L1 der TN 	<ul style="list-style-type: none"> Lautes Denken und <i>Stimulated Recall</i> (inkl. mündlicher Zusammenfassung der rezipierten Texte) zu ausgewählten Testaufgaben in der L1 der TN Psychometrische Daten zu den untersuchten Testaufgaben 	<ul style="list-style-type: none"> Retrospektive Befragung in der L1 der TN 	<ul style="list-style-type: none"> Eye-tracking Retrospektive Befragungen auf Englisch mit Aufnahmen der Augenbewegungen als Stimulus
Vorgehen bei der Datenerhebung	<ul style="list-style-type: none"> Training im Lauten Denken Dauer der Datenerhebung: ca. 3 Stunden pro Teilnehmer/in (TN) Datenerhebung durch 4 trainierte Assistent/innen Entschädigung der TN²⁵ 	<ul style="list-style-type: none"> Training im Lauten Denken Dauer der Datenerhebung: 2 Lektionen à 45 Min. pro Schüler/in Datenerhebung durch 26 trainierte Assistent/innen 	<ul style="list-style-type: none"> Training im Lauten Denken Datenerhebung durch den Autor Dauer der Aufgabenbearbeitung pro TN: ca. 40 Minuten, im Anschluss daran fanden die Interviews statt 	<ul style="list-style-type: none"> Dauer der Datenerhebung: 1 Stunde pro Schüler/in Datenerhebung durch eine der Autorinnen 	<ul style="list-style-type: none"> Dauer der Datenerhebung: ca. 2 Stunden pro TN Datenerhebung durch den Autor Entschädigung der TN

²⁵ Wird hier erwähnt, weil es sich gemäss A. D. Cohen und Upton (2006) evtl. auf den geringeren Einsatz von *test-wiseness*-Strategien auswirken könnte.

	A. D. Cohen & Upton (2006)	Nikolov (2006)	Rossa (2012)	L. Gu & So (2017)	Suvorov (2018)
Datenauswertung und Datenauswertungsmethoden	<ul style="list-style-type: none"> • Übersetzung ins Englische und Transkription der Daten zu 13 Items • Codierung der Daten durch trainierte Assistent/innen, Überprüfung von einer anderen Assistentin oder einem anderen Assistenten und einem der beiden Autoren • Qualitative und quantitative Auswertung der codierten Daten 	<ul style="list-style-type: none"> • Transkription der Daten • Codierung der Daten durch je eine Assistentin bzw. einen Assistenten und die Autorin • Qualitative Analyse der codierten Daten 	<ul style="list-style-type: none"> • Transkription (GAT 2-Basistranskript) und Segmentierung der Daten (Propositionen) • Codierung der Daten durch den Autor, ein Teil der Daten wurde zusätzlich durch eine Assistentin codiert • Qualitative und quantitative Auswertung der codierten Daten 	<ul style="list-style-type: none"> • Übersetzung ins Englische und Transkription der Daten • Doppelcodierung der Transkripte durch die Autorinnen • Qualitative und quantitative Auswertung der codierten Daten 	<ul style="list-style-type: none"> • Analyse und zeitliche und räumliche Beschreibung der Augenbewegungen der TN • Transkription und Codierung der verbalen Daten durch den Autor • Qualitative und quantitative Auswertung der codierten Daten
Kategorienleitfaden zur Erfassung von Strategien	<ul style="list-style-type: none"> • Entwicklung auf der Grundlage der Literatur zu Lese- und Teststrategien und datenbasiert • Strategien unterteilt in Lernerstrategien (n=28), Testmanagementstrategien (n=28) und <i>test-wiseness</i>-Strategien (n=3) 	<ul style="list-style-type: none"> • Entwicklung datenbasiert, Abgleich mit der Literatur zu (Lerner)Strategien • Strategien unterteilt in metakognitive, (n=16), soziale und affektive (n=5) und kognitive Strategien (n=13) 	<ul style="list-style-type: none"> • Entwicklung auf der Grundlage der relevanten Literatur und datenbasiert • Strategien unterteilt in metakognitive Strategien: Überprüfung der eigenen Verstehensprozesse (n=3), metakognitive Strategien: Evaluieren der eigenen Verstehensprozesse (n=2), affektive Strategien (n=2) und Teststrategien (n=3) 	<ul style="list-style-type: none"> • Entwicklung auf der Grundlage der Literatur zu Teststrategien und datenbasiert • Strategien unterteilt in Lernerstrategien (n=12), Testmanagementstrategien (n=11) und <i>test-wiseness</i>-Strategien (n=8) 	<ul style="list-style-type: none"> • Entwicklung auf der Grundlage der Literatur zu Teststrategien und datenbasiert • Teststrategien unterteilt in Testmanagementstrategien (n= 20) und <i>test-wiseness</i>-Strategien (n=7)

	A. D. Cohen & Upton (2006)	Nikolov (2006)	Rossa (2012)	L. Gu & So (2017)	Suvorov (2018)
Ausgewählte Resultate	<ul style="list-style-type: none"> Die TN bewältigten die Aufgaben allgemein wie von TOEFL-Autor/innen beabsichtigt. Die TN nahmen die Aufgaben als Test- und nicht als Lernaufgaben wahr. Die TN setzten viele Strategien ein, vor allem Testmanagement-strategien. Der Mehrwert der neuen TOEFL-Aufgaben gegenüber den traditionellen Aufgaben konnte nicht gezeigt werden. 	<ul style="list-style-type: none"> Die Aufgaben elizitierten die vom Test gesuchten Fähigkeiten der TN. Die TN gingen an die Texte unterschiedlich heran: Einige setzten die gleichen Strategien bei allen Aufgaben ein, andere setzten die gleichen Strategien nur gelegentlich ein. Die Strategien traten nicht isoliert, sondern in Kombination auf. 	<ul style="list-style-type: none"> Die eingesetzten Aufgaben erfassten generell die relevanten Aspekte des Testkonstrukts. Die Aktivierung von konstruktirrelevanten Prozessen und Strategien führte üblicherweise zur Auswahl einer falschen Antwortoption. Dank der Analyse der Verbalprotokolle konnten einzelne Items identifiziert werden, die Validitätsprobleme aufwiesen. 	<ul style="list-style-type: none"> Die Mehrheit der angewandten Strategien war konstruktrelevant. Kinder mit den besten Testresultaten setzten beim Hörverstehen keine und beim Leseverstehen nur wenige <i>test-wiseness</i>-Strategien ein; Kinder mit mittelmäßigen und schwachen Resultaten machten das deutlich häufiger. Leistungsstarke Kinder wandten öfters Lernerstrategien an. Die Testitems schienen nicht gut zwischen Lernenden auf niedrigeren Leistungsniveaus zu diskriminieren. 	<ul style="list-style-type: none"> Die TN wandten bei der Bewältigung der Aufgaben vielfältige Teststrategien an. Es gab individuelle Unterschiede im Strategieneinsatz der TN. Einige Strategien wurden nur bei bestimmten Itemtypen, andere wiederum itemtypübergreifend eingesetzt. Der Einsatz von <i>test-wiseness</i>-Strategien hatte einen statistisch signifikanten Effekt auf die beobachteten Testergebnisse und trug damit zur konstruktirrelevanten Varianz ein.

Trotz der Einschränkungen der Studien, auf die alle Autor/innen ausführlich hinweisen, können den ermittelten Strategien der Testteilnehmenden wertvolle Erkenntnisse entnommen werden. Diese Erkenntnisse sind in erster Linie für die Entwickler/innen von Testaufgaben von Bedeutung, sie sind aber auch über den jeweiligen Kontext hinaus interessant, nicht zuletzt deswegen, weil sie die Wichtigkeit der Erprobung der Testaufgaben und der genaueren Betrachtung der Vorgehensweisen der Testteilnehmer/innen zutage fördern. Ausserdem könnten von diesen Erkenntnissen auch andere Stakeholder wie z.B. Lehrpersonen oder Curriculumentwickler/innen profitieren.

2.2.5 ZUSAMMENFASSUNG

Den Gegenstand dieses Kapitels bildeten die theoretischen Grundlagen der vorliegenden Studie, in der die Teststrategien jugendlicher Französischlernenden beim Lösen von neu erstellten Lese- und Hörverstehensaufgaben untersucht wurden.

In Unterkapitel 2.1 wurden zentrale Begriffe und Konzepte zum Lese- und Hörverstehen sowie zur kommunikativen Beurteilung dieser Kompetenzen in einer Fremdsprache dargestellt, und zwar vor allem aus der Perspektive des *large-scale-assessments* im Kontext des schulischen Fremdsprachenunterrichts. Es wurde aufgezeigt, dass es sich beim Lese- und Hörverstehen um komplexe Prozesse handelt, die viele Gemeinsamkeiten haben (beide laufen intentional und zielgerichtet ab und können als Zusammenspiel zwischen dem Inputtext, den angemessenen kognitiven Prozessen und den bereits bekannten Informationen beschrieben werden (vgl. Grabe, 2009, S. 74)), aber auch wichtige Unterschiede aufweisen (u.a. wegen der ephemerischen Natur der gesprochenen Sprache im Unterschied zur geschriebenen Sprache). Wie sich Lese- und Hörverstehen (vgl. Grotjahn, 2012) modellieren lässt, wurde mithilfe des im Sprachtestbereich anerkannten Modells von Khalifa und Weir (2009) dargestellt. Das Modell stellt sowohl die kognitiven und metakognitiven Prozesse als auch das Sprach- und Hintergrundwissen der Leser/innen dar und umfasst drei Hauptteile: 1) eine metakognitive Komponente, 2) eine zentrale Verarbeitungskomponente und 3) eine Wissenskomponente, die ihrerseits jeweils weitere Unterprozesse mit einschliessen.

In diesem Unterkapitel wurde auch darauf hingewiesen, dass die Ergebnisse des Lese- und Hörverstehens nicht direkt beobachtbar sind, sondern mithilfe von verständnisüberprüfenden Aufgaben eliziert werden müssen: Dies führt dazu, dass die Beurteilung dieser beiden Kompetenzen und die seriöse Konstruktion von geeigneten Testaufgaben eine nicht zu unterschätzende Herausforderung darstellen. Die Lese- und Hörverstehensaufgaben müssen nicht nur das Lesen bzw. Hören evozieren, sondern es den Testteilnehmenden auch zu

demonstrieren ermöglichen, die vom Test gesuchten Fähigkeit erfolgreich eingesetzt respektive das Gelesene bzw. Gehörte tatsächlich verstanden zu haben.

Ein weiterer Teil dieses Unterkapitels (2.1.3) widmete sich zwei neueren Entwicklungen im Bereich der Beurteilung von rezeptiven Sprachkompetenzen, die für diese Studie von Bedeutung sind, d.h. den digitalen Textsorten und der szenariobasierten Beurteilung. Digitale Textsorten wie Chats oder Podcasts haben im realweltlichen Lesen und Hören eine zunehmend wichtige Bedeutung, was u.a. Auswirkungen auf den (Fremdsprachen-)Unterricht und somit auch auf die Beurteilung der Sprachkompetenzen hat. Mithilfe der szenariobasierten Beurteilung, die ursprünglich zur Erfassung von Leseverstehenskompetenzen der Schüler/innen aus den USA in Englisch entwickelt wurde, sollten u.a. das Leseverstehenskonstrukt breiter gefasst, die Testaufgaben zum Leseverstehen besser auf die theoretischen Grundlagen abgestützt und die Leseverstehenskompetenzen der Testteilnehmenden gefördert werden.

Das Unterkapitel 2.2 befasste sich mit den Strategien von Sprachlernenden. Zuerst wurde hier die Rolle der strategischen Kompetenz dargestellt, die in den Modellen der kommunikativen Kompetenz (Bachman, 1990; Bachman & Palmer, 1996, 2010; Canale, 1983; Canale & Swain, 1980) eine wichtige Rolle spielt. Danach wurden die verschiedenen aktuellsten Definitions- und Klassifizierungsversuche von L2-Strategien zusammengefasst (u.a. Arbeiten von A. D. Cohen, 2011b; Y. Gu, 2012; Oxford, 2017; Griffiths, 2018), da die Forschung aus dem allgemeineren Gebiet der Sprachlernstrategien auf die Erforschung der sprachtestbezogenen Strategien einen grossen Einfluss hatte. Ein weiterer Teil dieses Unterkapitels war den Strategien gewidmet, die bei der Bewältigung von Sprachtests zum Einsatz kommen können. Hier wurde der von A. D. Cohen (u.a. 2007b, 2011b, 2012, 2014) vorgeschlagene konzeptuelle Rahmen zur Klassifizierung von sprachtestbezogenen Strategien vorgestellt, der für den empirischen Teil der vorliegenden Arbeit übernommen wurde. A. D. Cohen (2014) unterteilt die Strategien, auf die die Testteilnehmenden bei der Bewältigung von Testaufgaben zurückgreifen können, in Lernerstrategien und Teststrategien, wobei sich die letzteren in Testmanagementstrategien und *test-wiseness*-Strategien unterteilen lassen. Die Lernerstrategien und Testmanagementstrategien sind relevant für das Testkonstrukt, die *test-wiseness*-Strategien sind hingegen konstruktirrelevant.

Teststrategien werden oft im Rahmen der Erprobung und Validierung von neuen Tests und Testaufgaben erforscht. Deswegen wurde in diesem Teil geschildert, was unter Testerprobung und -validierung zu verstehen ist, nämlich der durchdachte, bewusste und systematische Prozess des Sammelns und Analysierens von Beweisen zur Untermauerung der Validität einer Beurteilung vor ihrem operativen Einsatz (Kenyon & MacGregor, 2012,

S. 295); und wie dieser Prozess in der Praxis ablaufen kann. Zum Schluss wurden fünf neuere Studien zusammengefasst (A. D. Cohen & Upton, 2006; Nikolov, 2006; Rossa, 2012; L. Gu & So, 2017; Suvorov, 2018), in denen u.a. mithilfe von introspektiven Verfahren die Strategien der Testteilnehmenden beforscht wurden. Diese Studien waren hilfreich, um die Präzisierung und Differenzierung des Erkenntnisinteresses der vorliegenden Untersuchung vorzunehmen.

3 VORGEHEN BEI DER UNTERSUCHUNG DER STRATEGIEN DER TESTTEILNEHMENDEN

In diesem Kapitel wird die qualitative Untersuchung, die im Rahmen dieser Dissertation und im Zusammenhang mit dem Forschungsprojekt IFB durchgeführt wurde (Karges, Lenz et al., 2021; KFM, 2020a), ausführlich und im Sinne grösstmöglicher Transparenz vorgestellt. Im Projekt IFB, das bereits in Kap. 1 kurz dargestellt wurde, wurden computer- und szenariobasierte Testaufgaben für die Überprüfung der Lese- und Hörverstehenskompetenzen von Deutschschweizer Schüler/innen in ihren schulischen Fremdsprachen Englisch und Französisch am Ende der Sekundarstufe I erstellt. Schweizer Schüler/innen sollen gemäss den nationalen Bildungsstandards am Ende der Sekundarstufe I im Lese- und Hörverstehen in beiden schulischen Fremdsprachen mindestens das Niveau A2 gemäss GeR erreicht haben (EDK, 2011), weshalb die Aufgaben auf Sprachkompetenzen in diesem Bereich abzielen (A2/B1). Diese neu erstellten Aufgaben wurden im Projekt IFB überwiegend mit quantitativen Forschungsmethoden untersucht. In der vorliegenden, ergänzenden und zugleich für sich stehenden Untersuchung wurde mittels introspektiver Verfahren (Lautes Denken und *Stimulated Recall*) qualitative Evidenz zum Funktionieren und zur Validität der neuen Testaufgaben gesammelt. Im Fokus standen dabei Strategien (vgl. u.a. A. D. Cohen, 2014), auf die die Lernenden beim Bearbeiten der Französischaufgaben zurückgriffen.

Damit der Kontext der vorliegenden Studie besser nachvollzogen werden kann, werden in Kap. 3.1 zuerst die Testaufgaben und die Datenerhebung im Projekt IFB detaillierter beschrieben. In Kap. 3.2 werden die Forschungsfragen und daran anschliessend (3.3) introspektive Forschungsmethoden dargestellt, die oft bei der Erforschung von Teststrategien angewandt werden und auch in dieser Studie zum Einsatz kamen. Es folgen die Darstellungen zur Erhebung (3.4), Aufbereitung (3.5) und Auswahl der introspektiven Daten (3.6), welche anschliessend genauer auf die Strategien hin untersucht wurden. Dem Ablauf der Datenauswertung ist Kap. 3.7 gewidmet und eine Zusammenfassung (3.8) rundet diesen Teil ab.

3.1 FORSCHUNGSKONTEXT: DATENERHEBUNG UND TESTAUFGABEN IM PROJEKT IFB

An den Datenerhebungen im Projekt IFB, die zwischen Mitte Oktober 2017 und Anfang Februar 2018 stattfanden, nahmen insgesamt 39 Deutschschweizer Klassen (n=631) teil. Die

Klassen stammten sowohl aus der sog. *Passepartout*-Region²⁶, in der ab der 3. Klasse Französisch und ab der 5. Klasse Englisch unterrichtet wird, als auch aus anderen Gebieten der Deutschschweiz, wo zuerst Englisch und dann Französisch unterrichtet wird. Da die Schweizer Schüler/innen am Ende der obligatorischen Schulzeit in beiden Fremdsprachen die gleichen Grundkompetenzen erreichen sollen (EDK, 2011), kamen in beiden Regionen in beiden Fremdsprachen Parallelversionen der gleichen Aufgaben zum Einsatz. Jede Aufgabe wurde innerhalb jeder Klasse in einer französischen und einer englischen Sprachversion eingesetzt, wobei der Test einzelnen Schüler/innen nach dem Zufallsprinzip²⁷ entweder nur auf Französisch oder nur auf Englisch zugeteilt wurde.

Die im Projekt IFB erstellten Testaufgaben wurden in ein Szenario, d.h. in einen gemeinsamen Handlungsrahmen, eingebettet. Es entstanden insgesamt sechs Szenarien, „die aus vier inhaltlich zusammenhängenden, aber unabhängig voneinander lösbaren Aufgaben (A-D) im Niveaubereich A2/B1 bestehen, welche alle selbstständig am Computer gelöst werden können“ (Karges, Lenz et al., 2021). Innerhalb eines Szenarios bearbeiteten die Lernenden immer Aufgaben sowohl zum Leseverstehen als auch zum Hörverstehen, wobei am Anfang des Szenarios immer eine Leseverstehensaufgabe stand (Aufgabe A) und in einer letzten Aufgabe Hörverstehen und Leseverstehen kombiniert vorkamen (Aufgabe D). Je nach Szenario lösten die Schüler/innen zwei Leseverstehensaufgaben (Aufgabe A und C) und eine Hörverstehensaufgabe (Aufgabe B) oder zwei Hörverstehensaufgaben (Aufgaben B und C) und eine Leseverstehensaufgabe (Aufgabe A). Im Hörverstehensteil konnten die Schüler/innen die Hörtexte je nach Item und Aufgabe entweder einmal oder zweimal hören. Alle IFB-Aufgaben können im Forschungsdatenarchiv des KFM eingesehen werden²⁸ und die in der vorliegenden Studie eingesetzten Szenario-Aufgaben werden in Kap. 3.4.2 vorgestellt (siehe auch Anhang 2 und 3).

Als inhaltliche Vorarbeit für die Aufgabenentwicklung wurden die Lehrwerke und Lehrpläne der beiden Deutschschweizer Regionen gesichtet, in denen später Daten erhoben wurden. Die Grundlagen für die Erstellung der Szenarien bildeten in erster Linie die Minimalziele der

²⁶ Die Kantone entlang der deutsch-französischen Sprachgrenze: Basel-Stadt und Basel-Land, Solothurn, Bern, Fribourg und das Wallis.

²⁷ In Klassen, wo Lernende eine der beiden Fremdsprachen nicht (mehr) belegten (meist Französisch), erhielten diese die Aufgaben in der anderen Fremdsprache (meist Englisch). Lernende mit Englisch oder Französisch als Erstsprachen erhielten die Aufgaben ebenfalls in der jeweils anderen Sprache. Diese Prinzipien konnten jedoch wegen Lücken in den vorher erhaltenen Angaben zu den Schüler/innen nicht ganz durchgängig eingehalten werden.

²⁸ Die Aufgaben können unter <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000003403> heruntergeladen werden.

Deutschschweizer Lehrpläne, die für die beiden Fremdsprachen formuliert wurden (D-EDK, 2016; Passepartout: Fremdsprachen an der Volksschule, 2015).

Für die IFB-Aufgaben wurden vier (z.T. innovative) Formate ausgewählt:

- zu einer Frage aus drei möglichen Antwortoptionen, Texten oder Lösungen nur eine auswählen (Multiple-Choice-Frage); in einer Variante der Multiple-Choice-Frage zusätzlich angeben, in welcher Passage eines Lesetextes oder in welchem von mehreren Hörtexten die gesuchte Information gefunden wurde;
- aufgrund von neuen Informationen eine Tabelle mit Vorschlägen aus einer Liste ergänzen;
- aufgrund von Informationen in den Inputtexten eine Grafik (z.B. mit einem Zeitplan) ergänzen;
- bestimmte Textteile markieren (z.B. Programmteile, die gemäss dem Vorschlag eines Mitschülers geändert werden sollten).

In jedem Szenario kamen mindestens drei dieser vier Aufgabenformate vor. Es wurden insgesamt sechs Szenarien entwickelt: „Kinoabend“, „Berufswahl“, „Walkies“, „Schulfest“, „Stadttrip“ und „Waitaha“.

Jedes Szenario wurde den Schüler/innen zu Beginn auf einem Bildschirm jeweils kurz schriftlich vorgestellt, damit sie sich besser in die Situation hineinversetzen konnten. Dies sollte sie dazu ermutigen, ihre kognitiven Ressourcen besser zu aktivieren und dadurch ihre Kompetenzen optimal im Test zum Ausdruck zu bringen (vgl. Sabatini et al., 2014a).

Die Testinstruktionen, die Szenarioeinbettung und die Items wurden den Schüler/innen in ihrer Schulsprache Deutsch vorgelegt. Diese Entscheidung wurde auf der Grundlage der Erkenntnisse des IFB-Teams aus früheren, ähnlichen Forschungs- und Entwicklungsprojekten getroffen (insb. das Projekt „Task Lab“, KFM, 2020b; vgl. auch Karges, Barras et al., 2021). Die Ergebnisse dieser Projekte zeigen nämlich, dass – zumindest für Lernende auf einem tiefen Sprachkompetenzniveau – mithilfe der Items in der Schulsprache die zu ermittelnde Fähigkeit, zielsprachige Texte zu verstehen, zuverlässiger gemessen werden kann (vgl. Barras, 2018; Barras et al., 2016; Karges et al., i. V.). Auch die Erkenntnisse aus der einschlägigen Fachliteratur sprachen für diese Entscheidung (vgl. Alderson, 2000, S. 81–82, 86–87; Koda, 2012, S. 230; Shohamy, 1984).

Pro Klasse löste normalerweise ca. die Hälfte der Lernenden die französische und die andere Hälfte die englische Version der Testaufgaben. Die einzelnen Schüler/innen bearbeiteten insgesamt sechs Szenarien mit je vier Lese- und/oder Hörverstehensaufgaben (mindestens

drei Items pro Aufgabe). Darüber hinaus lösten die Schüler/innen weitere, „herkömmliche“ Testaufgaben zum Lese- und Hörverstehen²⁹ sowie Tests zu Teilkompetenzen, und füllten mehrere Fragebögen aus. Das Instrumentarium³⁰ (siehe Karges, Lenz et al., 2021) umfasste die folgenden Teile:

- „herkömmliche“ Lese- und Hörverstehenstests in der jeweils getesteten Fremdsprache (u.a. um Hinweise zur *concurrent validity* der neu kreierten szenariobasierten Testaufgaben zu bekommen);
- Tests bzw. Fragebögen zu bekannten, sowohl sprachspezifischen wie auch sprachunabhängigen Korrelaten der Lese- und Hörverstehenskompetenz (vgl. Alderson et al., 2015):
 - rezeptiver Wortschatz (Hören, Lesen, *sight word recognition*);
 - Textsegmentierung (Hören und Lesen);
 - Grammatikwissen (*grammaticality judgement*);
 - fluide Intelligenz (Ravens Matrizen);
 - Lese- und Hörstrategien (Fragebögen);
- Fragebogen zu Hintergrundvariablen (sozioökonomischer Status, ausserunterrichtlicher Kontakt mit der Zielsprache usw.).

Sämtliche Aufgaben wurden mit der Testentwicklungssoftware CBA Item Builder (DIPF & Nagarro, 2018b) erstellt und mit der zugehörigen Testauslieferungsplattform EE4CBA (DIPF & Nagarro, 2018a) in einem lokalen Netzwerk ausgeliefert. Das Auslesen der Ergebnisse, die als XML-Dateien vorliegen, erfolgte in R (R Core Team, 2014) mit einem eigens dafür erstellten Skript.

Die Schüler/innen bearbeiteten das IFB-Instrumentarium während insgesamt 6 Lektionen à 45 Minuten. Während des Bearbeitens der Testaufgaben konnten die Schüler/innen ihre Antworten nicht mehr verändern, sobald sie eine Bildschirmseite verlassen hatten. Je nach Itemformat waren auf den einzelnen Bildschirmseiten eines oder mehrere Items mit den dazugehörigen Hör- oder Lesetexten untergebracht.

Es wurde an Computern gearbeitet, die von den Projekt-Mitarbeitenden mitgebracht und vorbereitet wurden (siehe Abbildung 2). Es handelte sich um Klassensätze von HP Pavilion x2 2-in-1-Laptops, mit einem 10-Zoll-Touch-Bildschirm und abnehmbaren Tastaturen. Mäuse

²⁹ Es handelte sich dabei um öffentliche Aufgaben aus den KET- und PET-Tests der Cambridge General English Exams. Diese im Original englischen Aufgaben wurden für die französische Testversion adaptiert. Anders als in den Originalaufgaben wurden die Fragen zum Text z.T. in der Schulsprache Deutsch gestellt.

³⁰ Auf diese Instrumente wird hier nicht weiter eingegangen, weil sie nicht den Hauptgegenstand der vorliegenden Studie bilden.

wurden nicht zur Verfügung gestellt. Für die Hörverstehensaufgaben verwendeten die Lernenden von der Testleitung mitgebrachte oder eigene Kopfhörer. Als Testauslieferungsserver für die Laptops diente ein INTEC NUC6I, auf dem die schon erwähnte Testauslieferungssoftware EE4CBA lief. Die Laptops der Schüler/innen griffen per WLAN auf den Testserver zu, welcher mit einem eigenen WLAN-Router verbunden war. Die Lernenden sahen sämtliche Aufgaben in Vollbild. Items und Anweisungen wurden üblicherweise in der Schriftart „Arial“, Grösse 15, der übrige Text mindestens in „Arial“, Grösse 12 dargestellt. Der Aufbau und das Design der Aufgaben waren ähnlich wie in früheren Projekten des KFM (u.a. Karges, Barras et al., 2021) und wurden zusätzlich im Vorfeld qualitativ und quantitativ erprobt. Es ist davon auszugehen, dass die Lernenden visuell keine grösseren Schwierigkeiten bei der Wahrnehmung der Aufgaben oder der Bedienung der Testoberfläche hatten.



Abbildung 2: IFB-Studie, Schulzimmer kurz vor der Datenerhebung.

Um den Schüler/innen Abwechslung zu bieten und ihre Motivation für die Testteilnahme möglichst hoch zu halten, wurde die Reihenfolge der verschiedenen Instrumente gemischt und auf zwei Tage verteilt (pro Tag drei Lektionen mit Pausen dazwischen, die pro Klasse individuell gehandhabt wurden). Die Tests wurden pro Klasse von einer Testleiterin oder einem Testleiter geführt, die für diese Aufgabe im Voraus trainiert wurden. Sie achteten auf die Zeit und gaben den Schüler/innen Hinweise dazu, wann sie idealerweise zu einer nächsten Aufgabe übergehen sollen; die Schüler/innen konnten aber beim Bearbeiten der Aufgaben nach eigenem Ermessen vorgehen. Die Testleitenden halfen den Schüler/innen bei technischen Problemen, nicht aber bei inhaltlichen Fragen zum Test.

3.2 FORSCHUNGSFRAGEN UND FORSCHUNGSDESIGN

Bei der Konzipierung von neuen Itemtypen und Testaufgaben für standardisierte Prüfungen ist es zentral, diese Aufgaben und Itemtypen im Rahmen einer Testerprobung ganz genau zu untersuchen (Kenyon & MacGregor, 2012; Reed, 2014). Die während der Testerprobung gewonnenen Erkenntnisse tragen nämlich zur wichtigen Evidenz für die Validität eines Tests bei (Kenyon & MacGregor, 2012).

Testvalidität wird seit der einflussreichen Definition von Messick (1989) nicht mehr als Eigenschaft eines Tests angesehen: Es sind vielmehr die Interpretationen und Verwendungen dieses Tests, die als mehr oder weniger valide bezeichnet werden können (Chapelle, 2012, S. 24). Durch diese Neudefinierung von Validität wurde u.a. der Prozess der Testvalidierung umfangreicher und anspruchsvoller (vgl. Chapelle, 2012; Cumming, 2012). Neben den psychometrischen Erkenntnissen zum Funktionieren des Tests und seiner Elemente wird für eine Testvalidierung nun u.a. ergänzende Evidenz zu den Antwortprozessen (*response processes*) der Testteilnehmenden gesammelt (vgl. American Educational Research Association, 2011; Ercikan & Pellegrino, 2017a). Die *response processes* der Sprachlernenden beim Lösen von Testaufgaben werden häufig mittels introspektiver Verfahren wie dem Lauten Denken oder *Stimulated Recall* erhoben (vgl. u.a. Lumley & Brown, 2005; Arras, 2013; Ercikan & Pellegrino, 2017a; Lazaraton, 2017; Alison Green, 1998; Lazaraton & Taylor, 2007; Taylor, 2005; A. D. Cohen, 2011b).

Auf Basis der Erkenntnisse, die in Kap. 1 und 2 ausführlich dargelegt wurden, lassen sich für die qualitative Untersuchung der in Kap. 3.1 vorgestellten IFB-Testaufgaben für Französisch als Fremdsprache folgende Forschungsfragen formulieren:

1. Welche Strategien können bei Schüler/innen beim Lösen der IFB-Testaufgaben beobachtet werden?
2. Lassen sich Zusammenhänge zwischen konstruktirrelevanten Strategien und erfolgreichen bzw. nicht erfolgreichen Aufgabenlösungen erkennen?
3. Zeigen sich Unterschiede beim Einsatz der Strategien, wenn szenariobasierte und nicht szenariobasierte Testaufgaben gelöst werden, und, falls ja, worauf könnte dies zurückgeführt werden?

Im Fokus dieser Studie steht die Untersuchung der Antwortprozesse der Testteilnehmenden beim Lösen der Französischaufgaben aus dem Projekt IFB. Allgemeines Ziel ist es, Erkenntnisse zum Funktionieren und zur Validität der IFB-Testaufgaben sowie zur Einbettung dieser Aufgaben in ein übergreifendes Szenario zu gewinnen, die mit quantitativen Methoden

unzugänglich wären. Für die Untersuchung der *response processes* der Testteilnehmenden wurde in dieser Arbeit in Anlehnung an A. D. Cohen (u.a. 2006, 2011b, 2012, 2014), ebenso wie in den Studien von A. D. Cohen und Upton (2006), L. Gu und So (2017) und Suvorov (2018) der von Purpura (2014) als *the “Learner Strategies” approach* bezeichnete Ansatz gewählt. In diesem *bottom-up*-Ansatz, der auch mit anderen Herangehensweisen kombinierbar ist (Purpura, 2014; vgl. z.B. die Studie von Suvorov, 2018), stehen die internen mentalen Prozesse der Testteilnehmenden beim Lösen der Testaufgaben im Zentrum. Informationen über diese Prozesse werden mithilfe introspektiver Verfahren gewonnen, d.h. die Testteilnehmenden werden gebeten, ihre Gedanken beim Bearbeiten der Testfragen zu verbalisieren (Lautes Denken) und/oder im Nachhinein zu äussern (*Stimulated Recall* bzw. retrospektive Befragungen). Aus den Verbalisierungen der Testteilnehmenden entstehen sog. Verbalprotokolle, die anschliessend im Hinblick auf die Strategien, d.h. auf die von den Testteilnehmenden mehr oder weniger bewusst gewählten Prozesse zur Bewältigung von sprachbezogenen Aspekten und Anforderungen der Sprachtestaufgaben (A. D. Cohen, 2011b, S. 305), untersucht werden. Bei der Klassifizierung der Strategien wird in diesem Ansatz explizit zwischen konstruktrelevanten Lernerstrategien und Testmanagementstrategien sowie konstruktirrelevanten *test-wiseness*-Strategien unterschieden.

In der vorliegenden Studie wurden im Sinne der Triangulation (vgl. Knorr & Schramm, 2016) zum einen Lautdenkprotokolle der Schüler/innen beim Lösen der Testaufgaben erhoben und zum anderen direkt im Anschluss daran *Stimulated-Recall*-Interviews mit den Testteilnehmenden durchgeführt. Dies diente dazu, 1) die typischerweise fragmentarischen und bruchstückartigen Verbalisierungen aus den Lautdenkprotokollen valider interpretieren zu können und 2) durch gezielte Fragen in den *Stimulated-Recall*-Interviews ergänzende und vertiefende Informationen gewinnen zu können. Um den Einfluss der Szenarioeinbettung auf die Lösungswege der Lernenden besser einschätzen zu können, bearbeiteten 17 teilnehmende Schüler/innen Aufgaben, die in ein übergreifendes Szenario eingebettet waren, und 13 weitere Lernende eine ähnliche Anzahl an Aufgaben, die aber aus verschiedenen Szenarien stammten und daher keine kontextuelle Einbettung mehr aufwiesen, sonst aber vergleichbar waren.

Im folgenden Kapitel werden die introspektiven Verfahren näher dargestellt. Im Fokus stehen dabei das Laute Denken und die *Stimulated-Recall*-Interviews, die bei der Datenerhebung im vorliegenden Projekt zum Einsatz kamen. In anschliessenden Kapiteln werden Datenerhebung, Datenaufbereitung, Auswahl der Datengrundlage für die Analyse und Vorgehensweise bei der Datenauswertung beschrieben.

3.3 INTROSPEKTIVE VERFAHREN ALS DATENERHEBUNGSMETHODEN

Bei den introspektiven Verfahren (wie sie in der Fremdsprachenforschung aufgefasst werden) handelt es sich um Methoden, bei denen durch lautes Aussprechen von Gedanken, Beobachtungen und Emotionen „gezielt Daten bezüglich einer bestimmten (mentalen oder interaktionalen) Tätigkeit erhoben werden“ (Heine & Schramm, 2016, S. 173). Die so erhobenen Daten werden als Verbalprotokolle bezeichnet. Der Introspektion liegen gemäss Gass und Mackey zwei Annahmen zugrunde: Erstens, „it is possible to tap into and document a learner's internal processes in much the same way as one can observe external real-world events“, und zweitens, „humans have access at some level to their internal thought processes and can verbalize those processes“ (2017, S. 1). Dank Introspektion können folglich wertvolle Daten über die internen Verarbeitungsprozesse erhoben werden, die durch andere Methoden wie Beobachtung oder quantitative Analysen von Testergebnissen nicht erzeugt werden könnten (Sasaki, 2014, S. 1340). Die Typen von verbalen Daten (*verbal report*), die mit introspektiven Verfahren gewonnen werden können, unterteilt A. D. Cohen (2011b, S. 79–80) in 1) Selbstoffenlegung (*self-revelation*), d.h. die introspektive und unreflektierte (*nonmentalist*) Offenlegung der Denkprozesse während der Bearbeitung einer Information, die auch als Lautes Denken bekannt ist und als genauestes Abbild der kognitiven Prozesse der Proband/innen angesehen wird; 2) Selbstbeobachtung (*self-observation*), also die introspektive (innerhalb von 20 Sekunden nach dem mentalen Ereignis) oder retrospektive Erforschung eines bestimmten sprachlichen Verhaltens; und 3) Selbstbericht (*self-report*), der rückblickend und reflektiert (*mentalist*) ist und bei dem die Proband/innen generelle Aussagen zu ihren (sprachlichen) Handlungen machen.

Introspektive Verfahren gewannen im Bereich der Sprachlehrforschung seit der 1980er Jahre dank der Publikationen von Ericsson und Simon (1980, 1993) sowie Færch und Kasper (1987) zunehmend an Bedeutung (Lumley & Brown, 2005, S. 844). Im Bereich der Sprachtestforschung wird seit der wirkungsreichen Validitätsdefinition von Messick (1989) vermehrt auf Introspektion zurückgegriffen, weil die nach seinem Vorschlag aufgefassten Validierungsbestrebungen das Gewinnen von Evidenz über die Denkprozesse der Teilnehmenden erforderlich machen (Sasaki, 2014, S. 1342). Erkenntnisse, die aufgrund von introspektiven Datenerhebungsmethoden gewonnen werden, können aber auch über Testvalidierungszwecke hinaus genutzt werden, u.a. um die zukünftigen Lernprozesse der Testteilnehmenden optimaler zu gestalten (vgl. Arras, 2013, S. 83–87; A. D. Cohen, 2011b, S. 361; Sasaki, 2014, S. 1354).

Zu den introspektiven Verfahren, die im Folgenden kurz beleuchtet werden, gehören laut Heine und Schramm (2016) 1) Lautes Denken (vgl. u.a. Bowles, 2010; Heine, 2005; Heine & Schramm, 2007; Aguado, 2018; Alison Green, 1998), 2) *Stimulated Recall* bzw. Lautes Erinnern (vgl. u.a. Gass & Mackey, 2017; Knorr & Schramm, 2012; Knorr, 2013; Finkbeiner & Schluer, 2018) und 3) retrospektive Befragungen (vgl. u.a. Heine, 2014). Für eine umfassendere Diskussion der Vor- und Nachteile dieser Verfahren sowie Hinweise zur angemessenen Vorgehensweise bei der Erhebung von introspektiven Daten muss an dieser Stelle auf die soeben aufgeführte Literatur verwiesen werden.

Beim Lauten Denken wird „die aus dem Arbeits- oder Kurzzeitgedächtnis erfolgende simultane, ungefilterte Verbalisierung einer Person von Gedanken während einer (mental, interaktionalen oder aktionalen) Handlung“ angestrebt (Knorr & Schramm, 2012, S. 185). Die Forschungsteilnehmenden werden bei der Handlung und beim gleichzeitigen, unreflektierten Verbalisieren ihrer Gedanken (*inner speech*) möglichst nicht beeinflusst: Man erhofft sich, so wichtige Einblicke in ihre mentalen Abläufe zu erhalten (Heine & Schramm, 2016, S. 174). Die Verbalprotokolle der Proband/innen (die typischerweise fragmentarisch und inkohärent sind, weil sie nicht explizit an die Forschenden gerichtet sind) werden in Form von Audio- und/oder Videoaufnahmen festgehalten, um auf dieser Grundlage in einem anschließenden interpretatorischen Prozess auf die kognitiven Prozesse der Proband/innen während der Ausführung einer Tätigkeit rückschließen zu können (Heine, 2014, S. 124). Es muss dabei beachtet werden, dass beim Lauten Denken „nur diejenigen Gedankengänge analytisch rekonstruiert werden können, die im Arbeitsgedächtnis verarbeitet wurden; automatisierte Prozesse sind dagegen nicht zugänglich“ (Heine & Schramm, 2016, S. 175). Auch wird beim Lauten Denken wohl nicht alles verbalisiert, was gedacht wird, z.B. wegen dem mehr oder weniger parallelen Ablauf von mehreren Denkprozessen oder weil nicht alle mentalen Zustände in einer verbalen Form vorliegen und erst im Nachhinein versprachlicht werden müssen, was wiederum die quantitative Aufzählung der in Lautdenkdaten festgestellten Phänomene relativiert (ebd.; vgl. auch Knorr & Schramm, 2012). Beim Einsatz dieser Methode muss auch ihre Reaktivität bedacht werden, d.h. die Tatsache, dass das Laute Denken die Denkprozesse der Forschungsteilnehmenden beeinflussen kann (Bowles, 2010; Knorr & Schramm, 2012; Leow & Morgan-Short, 2004). Das unreflektierte Verbalisieren eigener Gedanken ist gewöhnungsbedürftig. Eine sorgfältig formulierte Instruktion und ein Training der Forschungsteilnehmenden im Lauten Denken vor der eigentlichen Tätigkeit sind daher von zentraler Bedeutung, um die Validität der erhobenen Daten zu erhöhen (Bowles, 2010; Heine & Schramm, 2016). Dennoch muss beim Einsatz des Lauten Denkens beachtet werden, dass nicht alle Forschungsteilnehmende für die Arbeit mit dieser Methode geeignet sind, weil „verschiedene Menschen ihre Gedanken offenbar unterschiedlich stark mit verbalen Formen

verknüpfen“ (Heine & Schramm, 2016, S. 176) und den Forschenden unterschiedlich gern Einsichten in ihre Gedankengänge gewähren (vgl. Gass & Mackey, 2011, S. 62). Trotz dieser Einschränkungen, „it is now widely agreed that various versions of thinking aloud are the most direct and therefore best tools available in examining the on-going processes and intentions“ (Y. Gu, 2014, S. 74).

Stimulated Recall wird von Gass und Mackey definiert als „a technique in which participants are asked to recall thoughts they had had while performing a prior task or while they had participated in a prior event“ (2017, S. 14). Die Autorinnen nehmen dabei an, dass die Reaktivierung möglichst vieler Gedanken aus der Erhebungssituation durch einen Stimulus in Form einer visuellen und/oder akustischen Erinnerung an diese Situation besser gelingen kann (vgl. ebd.). In *Stimulated-Recall*-Interviews können mithilfe von Interviewfragen oder Artefakten, wie z.B. einer Videoaufnahme der Situation oder einem Blatt mit der bearbeiteten Aufgabe, zwei Typen von Gedanken elizitiert werden: 1) *recall thoughts*, also introspektive „there and then“ Gedanken, die die Forschungsteilnehmenden während der eigentlichen Handlungsausführung hatten, und 2) *hindsight report*, d.h. reflexive und erst während des Interviews entstandenen „here and now“ Gedanken; wobei nur die Ersteren für die Ermittlung der tatsächlichen Denkprozesse bei der Handlungsausführung nützlich sind (Gass & Mackey, 2017, S. 44–45, vgl. auch S. 50). Im deutschsprachigen Raum führten Knorr und Schramm (2012) für die Erfassung der nachträglichen und ungefilterten Gedanken während der Datenerhebungssituation den Begriff „Lautes Erinnern“ ein, mit dem sie sich an den Terminus *Stimulated Recall* anlehnen, wobei sie damit sowohl medienunterstützte wie auch nicht-medienunterstützte Vorgehensweise bei der Datenerhebung meinen (ebd., S. 185–186; vgl. auch Heine & Schramm, 2016). Interviews, in denen die Forschungsteilnehmenden Fragen zu einer vorher durchgeführten Handlung beantworten und die „auf die entsprechenden mentalen Aktivitäten“ der Befragten abzielen, werden hingegen von den Autorinnen als „retrospektive Befragungen“ bezeichnet (Heine & Schramm, 2016, S. 174; vgl. auch Knorr, 2013)³¹.

Da die Forschungsteilnehmenden bei *Stimulated Recall* bzw. Lautem Erinnern darum gebeten werden, ihre Gedanken während der Tätigkeit *im Nachhinein* zu rekonstruieren und zu verbalisieren, verbalisieren sie – anders als beim Lauten Denken – nicht aus dem Arbeits- bzw. Kurzzeitgedächtnis, sondern greifen auf die im Langzeitgedächtnis abgespeicherten Erinnerungen zu (Heine & Schramm, 2016, S. 176). In der Folge muss damit gerechnet werden, dass diese Erinnerungen nicht nur die Tätigkeit selbst betreffen, sondern auch mit

³¹ In Kap. 3.4.4.2 wird diskutiert, warum in der vorliegenden Arbeit der Begriff *Stimulated Recall* verwendet wird.

dem Selbstbild der Forschungsteilnehmenden oder mit ihren allgemeinen Einstellungen verknüpft sein können (ebd., S. 174). Auch wenn es bei dieser Datenerhebungsmethode darum geht, die Gedanken der Forschungsteilnehmenden bei der vergangenen Handlung zu elizitieren (*there and then responses*), kann es in der Praxis vorkommen, dass sich ihre Verbalisierungen auch auf den Moment der Datenerhebung beziehen (*here and now responses*) (Knorr & Schramm, 2012, S. 197), was bei der Datenanalyse berücksichtigt werden sollte (Heine & Schramm, 2016, S. 177). Wichtig ist beim Einsatz dieser Methode auch, die Datenerhebung möglichst zeitnah zu der zu untersuchenden Handlung durchzuführen (idealerweise im Anschluss an die Handlung und allerspätestens innerhalb von 48 Stunden), um möglichst viele und auch möglichst unveränderte Erinnerungen an die Handlung hervorrufen zu können (Gass & Mackey, 2017; Heine & Schramm, 2016).

Zusammenfassend lässt sich festhalten, dass die hier beschriebenen introspektiven Verfahren zu den wichtigsten Datenerhebungsmethoden bei der Untersuchung von Teststrategien gehören (vgl. u.a. A. D. Cohen, 2011b, 2012). Sie ermöglichen u.a. wertvolle Einblicke in die Funktionsweise von Aufgaben, wodurch unterstützende Evidenz für die Validität eines Tests gewonnen werden kann (vgl. Arras, 2013, S. 77). Introspektion kommt daher oft im Rahmen der ersten, explorativen Phase der Erprobung von Sprachtests zum Einsatz und wird auch mit anderen Datenerhebungsmethoden trianguliert (vgl. Kenyon & MacGregor, 2012). Der Einsatz der introspektiven Verfahren erweist sich vor allem bei der Untersuchung von Lese- und Hörverstehensaufgaben als besonders vorteilhaft. Da sowohl der Prozess wie auch das Produkt des Lese- und Hörverstehens nahezu vollständig internalisiert ablaufen und sich der Beobachtung entziehen, muss für die Beurteilung von Lese- und Hörverstehensfähigkeiten auf Testaufgaben zurückgegriffen werden, die diese Fähigkeiten sichtbar machen (Taylor, 2005, S. 3). Die introspektiven Verfahren bieten hierfür „one possible way of ‘making visible’ at least something of what may be going on inside a test-taker’s head when they complete a reading or listening task” (ebd.), was einzuschätzen hilft, inwiefern die Testaufgaben tatsächlich die anvisierten Lese- und Hörverstehensfähigkeiten auslösen.

Zu den Nachteilen der oben skizzierten introspektiven Verfahren gehört – neben den möglichen methodologischen Fallen, die es durch eine umsichtige Planung der Datenerhebung und der Datenauswertung zu vermeiden gilt (vgl. z.B. Gass & Mackey, 2011, S. 53–63) – ihre Praktikabilität, da die Erhebung introspektiver Daten einen beträchtlichen Aufwand darstellt. Auch müssen für ihre Aufbereitung und Auswertung erhebliche Ressourcen vorgesehen werden (vgl. z.B. Arras, 2013; Heine, 2014).

Trotz ihrer methodischen und praktischen Einschränkungen sind die introspektiven Verfahren im Bereich der Sprachtestforschung, wie den obigen Ausführungen zu entnehmen ist, äusserst nutzbringend, weil sie wichtige, ergänzende Evidenz zum Funktionieren der Testaufgaben und dem Verhalten der Testteilnehmenden hervorbringen können. Insbesondere für die Konzipierung neuer Sprachtests oder Testformate seien daher introspektive Verfahren „dringend anzuraten“ (Arras, 2013, S. 78). Die bei der Erforschung von Sprachtests mittels introspektiver Verfahren gewonnenen Erkenntnisse können darüber hinaus auch nützlich sein, um die anschliessenden Lernprozesse der Testteilnehmenden zu unterstützen und um die Kluft zwischen „assessment and education“ zu überbrücken (Sasaki, 2014, S. 1354).

3.4 DATENERHEBUNG

In diesem Teil wird der Verlauf der Datenerhebung in der qualitativen Studie dargestellt. Zunächst wird über die Erkenntnisse aus der Pilotierung der qualitativen Untersuchung berichtet (3.4.1). Daraufhin wird genauer auf die Testaufgaben eingegangen, die in der qualitativen Studie eingesetzt wurden (3.4.2). Das nächste Unterkapitel (3.4.3) ist den Schüler/innen gewidmet, die an der qualitativen Studie teilnahmen. Im Anschluss daran wird die Erhebung der Lautdenkprotokolle und der *Stimulated-Recall*-Interviews detailliert geschildert (3.4.4).

3.4.1 PILOTSTUDIE

Die Vorgehensweise und Instrumente, die bei der eigentlichen Datenerhebung zwischen November 2017 und Februar 2018 zum Einsatz kamen, wurden im Juni 2017 mit sechs Lernenden pilotiert. Diese Lernenden stammten aus Klassen, die an der Pilotierung der IFB-Testinstrumente teilnahmen. Die Pilotierung der qualitativen Studie dauerte, wie auch die Haupterhebung, drei aufeinanderfolgende Lektionen à 45 Minuten pro Schüler/in.

Bei der Pilotierung wurde das für die Hauptuntersuchung vorgesehene Szenario „Stadttrip“ in seiner englischen (2 Teilnehmende) und französischen (3 Teilnehmende) Version eingesetzt. Mit einem der Lernenden wurde auch die französische Version des Szenarios „Waitaha“ pilotiert, dies aber vor allem mit dem Ziel, die technische Seite der Datenerhebung zu erproben und dem IFB-Team vor der Haupterhebung eine unmittelbare Rückmeldung zur Usability der Antwortformate zu geben.

Die Pilotierung zeigte, dass sich das Laute Denken und die *Stimulated-Recall*-Interviews als Datenerhebungsverfahren bewährten, der vorgesehene Ablauf gut funktionierte und die zur Verfügung stehende Zeit für die anschliessende Besprechung der Testaufgaben mit diesen

sechs Schüler/innen ausreichend war. Aufgrund der Erfahrung in der Pilotierung konnte die Instruktion zum Lauten Denken präzisiert und ergänzt werden (mehr dazu in Kap. 3.4.4.1).

Es stellte sich auch heraus, dass die englischsprachige Version des Szenarios „Stadttrip“ für die Lernenden sehr einfach war. Die IFB-Aufgaben (Zielniveau: A2/B1) orientierten sich an den nationalen Bildungsstandards (EDK, 2011), die am Ende der Sekundarstufe I das Erreichen des Niveaus A2 gemäss GeR in beiden schulischen Fremdsprachen vorsehen. Es wird vermutet, dass diese Lernenden in Englisch bereits ein höheres Niveau erreicht haben könnten und ihre rezeptiven Englischkenntnisse zumindest bei Aufgaben auf A2-Niveau einen gewissen Grad an Automatisierung aufwiesen. Dies wäre wohl durch die weltweit besondere Stellung des Englischen und durch die Präsenz dieser Sprache im Alltag der (Schweizer) Jugendlichen erklärbar. Die Tatsache, dass die Englischaufgaben für die Lernenden einfach zu bewältigen waren, führte dazu, dass sie dabei nicht viel verbalisierten, was zu wenig interessanten Daten für die Analyse führte. Dieses Phänomen ist beim Einsatz des Lauten Denkens als Datenerhebungsverfahren nicht unbekannt: „[A]utomatisierte Prozesse, die normalerweise unbewusst ausgeführt werden, [sind] kein geeigneter Untersuchungsgegenstand, weil sie für eine Verbalisierung erst bewusst gemacht werden müssten; dies würde die eigentlichen Abläufe beeinflussen“ (Heine, 2014, S. 127). Für die Erhebung der Lautdenkdaten werden daher Aufgaben empfohlen, die für die Teilnehmenden einen gewissen Schwierigkeitsgrad aufweisen, denn „[n]ur wenn begründet angenommen werden kann, dass die Aufgabe nicht durch automatisierte Vorgänge gelöst werden kann, wird man in den Lautdenkprotokollen Einsichten in den Problemlösungsprozess erhalten“ (Heine & Schramm, 2007, S. 176). Auch wenn die englische Version des Szenarios nur mit zwei Lernenden erprobt wurde, wurde nach der Pilotierung entschieden, in der Haupterhebung ausschliesslich die französische Version der IFB-Aufgaben einzusetzen, deren Schwierigkeitsniveau besser mit den Zielen der Studie zusammenpassen. Auch bei sehr schwachen Französischlernenden wurde beobachtet, dass sie wenig verbalisierten, weil sie von den Inputtexten kaum etwas verstanden. Wie weiter unten beschrieben wird (Kap. 3.4.3.1), wurden deshalb für die Hauptstudie durchschnittliche Französischlernende pro teilnehmende Klasse rekrutiert.

Während der Pilotierung wurden auch wichtige technische Erkenntnisse für die Haupterhebung gewonnen. Schon bei der Pilotierungssession mit der ersten Probandin zeigte sich schnell, dass die im Projekt IFB eingesetzten Computer zu leistungsschwach waren, um gleichzeitig Aufgaben zu lösen und längere Video- und Tonaufnahmen der Datenerhebungssituation anzufertigen. Ab der zweiten Pilotierungssession wurde daher ein anderer, leistungstärkerer Laptop eingesetzt, auf dem die Aufgaben in der gleichen Grösse und Auflösung wie im IFB-Setting dargestellt wurden und der auch sonst gleich wie die

IFB-Laptops zu bedienen war. Es wurde darüber hinaus festgestellt, dass für die qualitativ guten Tonaufnahmen ein externes Mikrofon benutzt werden musste und das Anfertigen von zusätzlichen Aufnahmen mit einem Diktiergerät lohnenswert ist.

3.4.2 EINGESETZTE IFB-TESTAUFGABEN

Im Rahmen der vorliegenden Studie setzten sich 17 Lernende mit der französischen Version des Szenarios „Stadttrip“ auseinander. Weitere 13 Schüler/innen lösten insgesamt fünf Französischaufgaben, die verschiedenen Szenarien entnommen wurden und die im Folgenden als Off-Szenario-Aufgaben bezeichnet werden.

Im Szenario „Stadttrip“ geht es darum, einen Aufenthalt in Avignon zu organisieren. Die Szenarioeinbettung lautet dementsprechend: „Du bist während 3 Wochen für einen Sprachkurs in Frankreich. Nächstes Wochenende möchtest du die Stadt Avignon besuchen.“ (vgl. Abbildung 3).

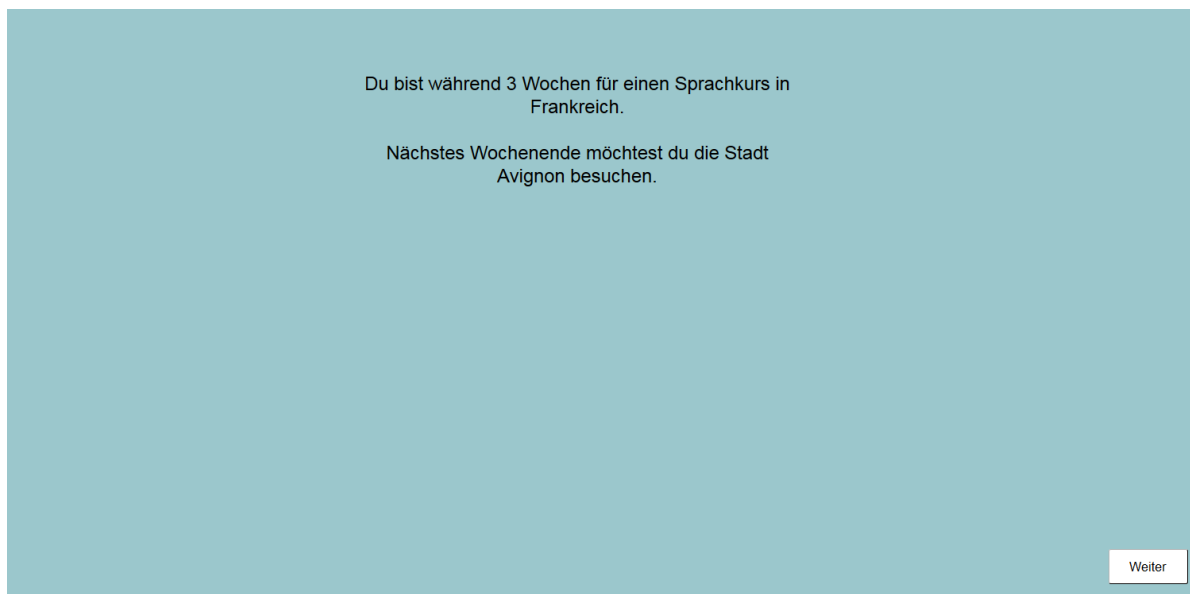


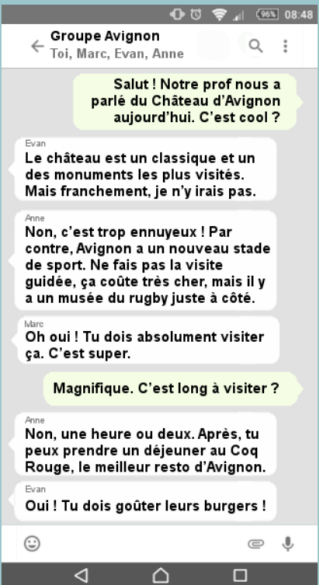
Abbildung 3: Einbettung des Szenarios „Stadttrip“.

Im Rahmen dieses Szenarios mussten die Schüler/innen:

1. Tipps zur Besichtigung von Avignon verstehen, die ihnen von Bekannten in einem Gruppenchat auf dem Smartphone gegeben werden (Task A);
2. Informationen zur Region Avignon aus einem Audioblog einer kanadischen Touristin verstehen (Task B);
3. Aufgrund von Internet-Reviews ein passendes Hotel in Avignon finden, das den vorgegebenen Kriterien entspricht (Task C);

4. Aufgrund der Informationen auf einer Internetseite *und* aus einem Radiobeitrag einen Zeitplan für den Besuch eines Comic-Festivals in Avignon erstellen (Task D) (vgl. Abbildung 4 bis Abbildung 7).

Vor jeder neuen Aufgabe wurde eine Seite eingebaut, die die Schüler/innen über die Situierung dieser Aufgabe im Rahmen des Szenarios informiert. So lautet beispielsweise die Situierung zu Task A im Szenario „Stadttrip“: „Du hast Bekannte in Avignon: Marc, Evan und Anne. In einem Gruppenchat geben sie dir Tipps, was man in Avignon unternehmen kann. Deine Aufgabe: Du liest den Chat. Auf den nächsten Seiten sammelst du die Tipps deiner Bekannten.“ Nach dieser Einführung sahen die Lernenden drei Bildschirmseiten hintereinander, auf denen Teile des Inputtextes, in diesem Fall der erwähnte Gruppenchat, und entsprechende Items abgebildet waren. Bei anderen Formaten, z.B. beim Ausfüllen einer Tabelle, konnte auf einer Bildschirmseite auch eine Reihe mehr oder weniger voneinander abhängiger Items dargestellt sein, die sich alle auf den gleichen Teil des Inputtextes bezogen. In der Aufgabenentwicklung wurde versucht, die Bearbeitungszeit pro Bildschirmseite etwa gleich zu halten. Im Folgenden wird daher im Allgemeinen von Bildschirmseiten die Rede sein und weniger von Items.



Lies den Anfang des Chats.

Ziehe die Informationen in die passende Kategorie.
Die Reihenfolge ist egal.

✓ Das empfehlen die drei:	✗ Das empfehlen die drei nicht:
?	?
?	?

Das Schloss anschauen

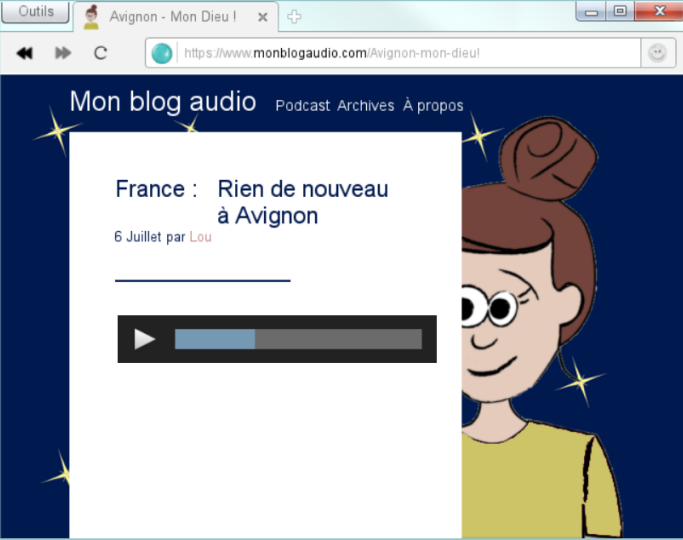
Die Stadiontour machen

Die Rugby-Ausstellung besuchen

Im 'Coq Rouge' essen

Weiter

Abbildung 4: Szenario „Stadttrip“, Task A, erste Bildschirmseite.



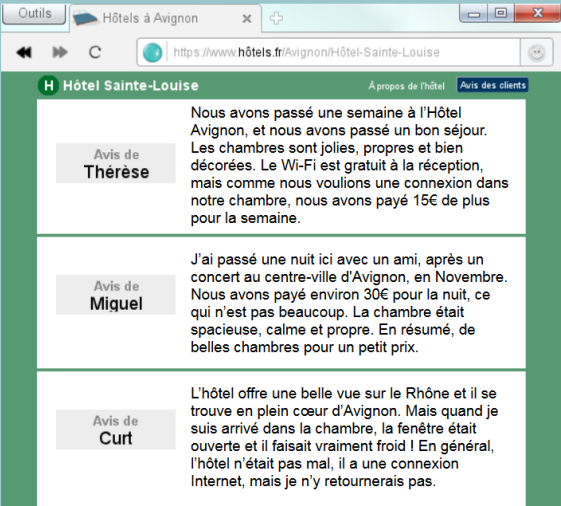
Höre den ersten Teil von Lous Podcast. Beantworte dann die Frage.

Lou war für drei Tage nicht im Internet. Warum?

- ☐ Sie war in einem kleinen Dorf ohne Internetempfang.
- ☐ Ihre neue französische SIM-Karte war noch nicht aktiviert.
- ☐ In ihrem Hotel in Avignon funktionierte das Internet nicht.

Weiter

Abbildung 5: Szenario „Stadttrip“, Task B, erste Bildschirmseite.



Lies die Bewertungen über das **Hôtel Sainte-Louise**.

Du hast dir für jedes Hotel drei Fragen gestellt.

- Beantworte die Frage mithilfe der Bewertungen. Klicke an!
- In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du in der gleichen Bewertung mehrere Antworten.

- Gibt es im Zimmer gratis WLAN?

☐ ja ✓
 ☐ nein ✗

Avis de ?
- Ist das Hotel im Stadtzentrum?

☐ ja ✓
 ☐ nein ✗

Avis de ?
- Gibt es im Hotel Zimmer unter 35€ pro Nacht?

☐ ja ✓
 ☐ nein ✗

Avis de ?

Weiter

Abbildung 6: Szenario „Stadttrip“, Task C, erste Bildschirmseite³².

³² Hier sollten die Schüler/innen die links stehenden Kästchen „Avis de Thérèse“, „Avis de Miguel“ und „Avis de Curt“ den passenden und rechts stehenden Kästchen „Avis de ?“ zuordnen.

Höre den ersten Teil des Podcasts und lies den Blogeintrag.

Welche Veranstaltungen finden am Sonntag statt? Um welche Uhrzeit? Ziehe die Kärtchen rechts zur passenden Uhrzeit.

Achtung:
Auf dieser Seite musst du nur 2 Veranstaltungen finden.

Spannende Veranstaltungen am Sonntag

9:00	?
10:00	?
11:00	?
11:30	?
13:00	?

- Ausstellung: Neuheiten in der Robotik
- Diskussion: Roboter - das Spielzeug der Zukunft
- Filmfestival: Science-Fiction-Klassiker
- Kinofestival: Superhelden
- Kurzfilmfestival mit alten Trickfilmen

Weiter

Abbildung 7: Szenario „Stadttrip“, Task D, erste Bildschirmseite.

Die Spezifikationen und Formate der Testaufgaben aus dem Szenario „Stadttrip“ können Tabelle 2 zusammenfassend entnommen werden.

Tabelle 2: Taskspezifikationen Szenario „Stadttrip“.

Task	Fertigkeit	Handlungsziel
Task A (6 Items auf 3 Bildschirmseiten)	Leseverstehen	Einen schriftlichen Dialog mit Empfehlungen verstehen
Task B (3 Items auf 3 Bildschirmseiten)	Hörverstehen	Einen längeren narrativen Text in groben Zügen verstehen
Task C (9 Items auf 3 Bildschirmseiten)	Leseverstehen	Ein Suchresultat aufgrund von vorgegebenen Kriterien auswählen
Task D (15 Items auf 3 Bildschirmseiten)	Lese- und Hörverstehen	Aus einem Hör- und einem Lesetext zum gleichen Thema einen Zeitplan ergänzen

Die sog. Off-Szenario-Aufgaben wurden aus verschiedenen Szenarien zusammengestellt, weshalb für diese Aufgaben keine Szenarioeinbettung oder Task-Situierung vorgesehen wurde. Bei der Auswahl der Off-Szenario-Tasks wurde darauf geachtet, dass die Handlungsziele, die Formate und die Gewichtung der getesteten Fertigkeiten aus den Off-Szenario-Tasks möglichst denen des Szenarios „Stadttrip“ entsprechen. Zu Vergleichszwecken wurde ausserdem Task B aus Szenario „Stadttrip“ auch in die Off-Szenario-Tasks integriert.

Die technische Vorbereitung der Off-Szenario-Aufgaben für die qualitative Studie konnte aus Zeitgründen erst während der laufenden Datenerhebung mit dem Szenario „Stadttrip“ umgesetzt werden. Die bis zu diesem Zeitpunkt gewonnene Erfahrung mit den Aufgaben aus dem Szenario „Stadttrip“ hatte gezeigt, dass die genauere Besprechung aller vier

Szenario-Tasks³³ vor allem bei sehr mitteilungsfreudigen Schüler/innen relativ viel Zeit in Anspruch nehmen kann, so dass in solchen Fällen nicht alle Items in der gewünschten Ausführlichkeit besprochen werden konnten. Aus diesem Grund wurde entschieden, die Länge der Tasks bei den Off-Szenario Aufgaben zu kürzen, d.h. bei vier der fünf Tasks jeweils nur die zwei ersten Bildschirmseiten zu nehmen (vgl. Tabelle 3). Lediglich der fünfte Task, welcher ohnehin nur auf einer Bildschirmseite abgebildet ist, wurde im Ganzen besprochen. Dieser wurde trotz der zeitlichen Einschränkung hinzugefügt, um auch qualitative Informationen über das Format dieser Aufgabe zu sammeln, welches im Szenario „Stadttrip“ nicht vorkommt.

Tabelle 3: Taskspezifikationen Off-Szenario.

Task	Fertigkeit	Handlungsziel
Szenario „Walkies“, Task C (2 Items auf 2 Bildschirmseiten)	Leseverstehen	Einen schriftlichen Dialog mit Anweisungen verstehen
Szenario „Stadttrip“, Task B (2 Items auf 2 Bildschirmseiten)	Hörverstehen	Einen längeren narrativen Text in groben Zügen verstehen
Szenario „Berufswahl“, Task C (6 Items auf 2 Bildschirmseiten)	Leseverstehen	Ein Suchresultat aufgrund von vorgegebenen Kriterien auswählen
Szenario „Waitaha“, Task C (10 Items auf 2 Bildschirmseiten)	Hörverstehen	Eine kurze, persönliche Nachricht im Detail verstehen
Szenario „Schulfest“, Task D (14 Items auf 1 Bildschirmseite)	Lese- und Hörverstehen	Widersprüchliche Informationen in einem Hörverstehens- und Leseverstehenstext identifizieren

Transkripte der Lese- und Hörverstehenstexte sowie Screenshots mit den Aufgaben, die bei der qualitativen Studie eingesetzt und vertieft analysiert wurden, sind im Anhang 2 und 3 abgebildet. Alle IFB-Szenarioaufgaben sind auch im KFM-Forschungsdatenarchiv³⁴ abrufbar.

3.4.3 UNTERSUCHUNGSTEILNEHMENDE

An der hier beschriebenen qualitativen Untersuchung nahmen insgesamt 30 durchschnittlich 15-jährige Französischlernende aus der Deutschschweiz teil. Bereits bei der Planung der Datenerhebung wurde damit gerechnet, dass die erhobenen Daten sehr umfangreich ausfallen würden und vor der Analyse reduziert werden müssen (vgl. Heine & Schramm, 2007, S. 196). Die für eine solche Untersuchung recht hohe Anzahl an Teilnehmenden wurde angestrebt, weil aufgrund der Erfahrung bei der Pilotierung und der Erkenntnisse aus der Forschungsliteratur zu introspektiven Verfahren davon auszugehen war, dass ein Teil der

³³ Mehr Informationen zum Ablauf der qualitativen Studie sind in Kap. 3.4.4 zu finden.

³⁴ <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000003403>

erhobenen Daten für bestimmte Analyseteile unbrauchbar sein würde, sei es aus technischen Gründen oder weil sich nicht alle Untersuchungsteilnehmende für die Erhebung der Lautdenkdaten gut eignen würden.

Die qualitative Studie wurde an einem der beiden Halbtage durchgeführt, der in der jeweiligen Klasse für die Teilnahme am Projekt IFB vorgesehen war; d.h. die ausgewählten Schüler/innen bearbeiteten die Testaufgaben an diesem Tag individuell in einem separaten Raum und in Anwesenheit der Forscherin (ein/e Schüler/in pro Halbtage), während ihre Klasse am quantitativen Teil der IFB-Untersuchung teilnahm. Am anderen Untersuchungstag bearbeiteten diese Lernenden während drei weiteren Lektionen zusammen mit ihrer Klasse die „herkömmlichen“ Aufgaben zum Hör- und Leseverstehen, die Aufgaben zu Teilkompetenzen sowie die Fragebögen (vgl. Kap. 3.1). Auf diese Weise konnten alle Hintergrundinformationen zu den Lernenden miterhoben werden, z.B. in Bezug auf ihre Wortschatz- und Grammatikkenntnisse. Aus Zeitgründen wurden bei diesen 30 Schüler/innen die restlichen Szenarien weggelassen.

Im Folgenden wird die Auswahl der Schüler/innen für die qualitative Studie begründet (Kap. 3.4.3.1). In Kap. 3.4.3.2 werden die Testresultate der 30 Lernenden, die an der qualitativen Studie teilnahmen, mit den Testresultaten aller anderer Schüler/innen verglichen, die sich mit den IFB-Französischtests auseinandersetzten.

3.4.3.1 AUSWAHL DER TEILNEHMENDEN FÜR DIE QUALITATIVE UNTERSUCHUNG

Die 30 Lernenden aus der qualitativen Stichprobe stammten aus 19 der insgesamt 39 achten und neunten Klassen, die an der IFB-Studie teilnahmen. Pro Klasse wurden maximal zwei Schüler/innen ausgewählt. Nicht alle Lehrpersonen der IFB-Klassen stimmten der Teilnahme an der qualitativen Studie zu, u.a. weil die Durchführung dieser Zusatzstudie etwas mehr Organisationsaufwand bedeutete. Auch musste für diese Untersuchung ein separater Raum zur Verfügung gestellt werden, was in einigen Schulen aus Platzgründen nicht möglich war.

Eine schriftliche Einverständniserklärung für die Teilnahme an der IFB Studie wurde, je nach Vorgaben des Kantons bzw. der Schule, von den Lehrpersonen, Schuldirektionen oder den Erziehungsberechtigten der Schüler/innen eingeholt. Da bei der hier beschriebenen, qualitativen Untersuchung auch Videoaufnahmen von den minderjährigen Lernenden gemacht wurden, unterschrieben ihre Erziehungsberechtigten im Voraus eine zusätzliche Einverständniserklärung für die Verwendung der erhobenen Daten für die Forschungszwecke (siehe Anhang 1). Zusätzlich wurden die Lernenden von der Forscherin direkt vor der

Untersuchung noch einmal mündlich über die Studie informiert und um ihr mündliches Einverständnis gebeten.

An der quantitativen IFB-Untersuchung nahmen insgesamt 631 Schüler/innen aus 39 Klassen teil, die mittels eines breit gestreuten *call for participants* gewonnen wurden. Die Teilnahme der Klassen und der Lernenden an der Studie war freiwillig: Es handelte sich bei diesen Klassen also um eine willkürliche Auswahl (*convenience sample*) (vgl. Diekmann, 2008, S. 378). Unter den Klassen, die an der IFB-Studie teilnahmen, waren sowohl Klassen aus den tiefsten Leistungsanforderungen als auch Klassen aus den höchsten Leistungsanforderungen und die Datenerhebung fand sowohl in eher städtischen Regionen als auch auf dem Land statt. Einzelne Schulen aus der Nähe der deutsch-französischen Sprachgrenze waren ebenfalls beteiligt. Somit kann vorsichtig vermutet werden, dass die IFB-Stichprobe eine gewisse Bandbreite der Klassen mit unterschiedlichen Leistungsanforderungen am Ende der obligatorischen Schulzeit abdeckt. Bei der Auswahl der 30 Schüler/innen für die vorliegende Studie wurde darauf geachtet, dass Lernende aus Klassen mit unterschiedlichen Leistungsanforderungen vertreten waren.

Dem Forschungsteam standen vor der Datenerhebung in den teilnehmenden Klassen neben der Anzahl der Schüler/innen die folgenden Informationen zur Klasse zur Verfügung:

- allgemeine Leistungsanforderungen der Klasse:
 - Grundanforderungen (in dieser Studie „C“ genannt)
 - allgemeine Anforderungen (in dieser Studie „B“ genannt)
 - erweiterte Anforderungen (in dieser Studie „A“ genannt)
 - Klasse mit unterschiedlichen Anforderungen (sog. kooperatives Modell) oder ohne Selektion (sog. integriertes Modell) (vgl. EDK, 2019; educa, 2019)³⁵;
- Reihenfolge der gelernten Fremdsprachen:
 - Englisch ab der 3. und Französisch ab der 5. Klasse, oder
 - Französisch ab der 3. Klasse und Englisch ab der 5. Klasse;
- Anzahl der Lernenden in der Klasse;
- Lernende ohne Französisch oder Englisch.

In einigen Fällen informierten die Lehrpersonen auch von sich aus über Lernende mit DaZ-Hintergrund oder Lese-Rechtschreibschwächen.

³⁵ Die Klassen mit verschiedenen Ansprüchen werden in der Schweiz je nach Region anders bezeichnet, worauf hier übersichtshalber verzichtet wird. Für einen vollständigen Überblick siehe EDK (2019).

Die klassenbezogenen Informationen waren in Bezug auf die einzelnen Schüler/innen meist nicht sehr aussagekräftig. Aus diesem Grund wurde entschieden, bei der Auswahl der Proband/innen für die qualitative Studie die Lehrpersonen miteinzubeziehen, da diese ihre Lernenden in der Regel gut kennen (sollten). Die Lehrpersonen wurden im Voraus gebeten, für die Teilnahme an der Untersuchung zwei bis drei Schüler/innen aus ihrer Klasse vorzuschlagen, die an der qualitativen Studie teilnehmen möchten und die in Französisch weder allzu schwach noch allzu stark waren. Die Lehrpersonen wurden gebeten, jeweils mehr Schüler/innen für die qualitative Studie zu rekrutieren als nötig, für den Fall, dass am Tag der Untersuchung eine/r der ausgewählten Lernenden nicht anwesend wäre. Da die insgesamt 30 ausgewählten Schüler/innen aus Klassen mit unterschiedlichen Leistungsanforderungen stammten, konnte angenommen werden, dass die Französischkompetenzen dieser Lernenden auch unterschiedlich stark ausgeprägt sein würden.

Die Entscheidung, bei der qualitativen Untersuchung möglichst „durchschnittliche“ Lernende pro Klasse zu berücksichtigen, wurde aufgrund der Beobachtungen bei der Pilotierung der vorliegenden Studie getroffen (vgl. Kap. 3.4.1). Während der Pilotierung hatte sich gezeigt, dass die schwächsten Schüler/innen von den rezipierten Texten kaum etwas verstanden. Dies hatte zur Folge, dass sie beim Lauten Denken kaum etwas verbalisieren und beim *Stimulated Recall* ihre Entscheidungen kaum begründen konnten. Für die besten Lernenden wiederum waren die IFB-Aufgaben, die für die Überprüfung des Erreichens der Grundkompetenzen in den schulischen Fremdsprachen (also des Niveaus A2.2) konzipiert worden waren, eher einfach. Dadurch waren die Ergebnisse des Lauten Denken bei diesen Schüler/innen sehr spärlich, weil für sie die rezipierten Texte weitgehend klar waren und sie dazu nicht viel mitzuteilen hatten.

Die untenstehende Tabelle gibt einen Überblick über alle Schüler/innen, die an der qualitativen Studie teilgenommen haben. Die in der Tabelle aufgelisteten Pseudonyme wurden von den Schüler/innen selbst gewählt, falls gewünscht mithilfe einer Liste mit Vorschlägen. Bei der vertiefenden Analyse wurden die Transkripte der 20 grau hinterlegten Schüler/innen berücksichtigt (mehr dazu in Kap. 3.6).

Tabelle 4: Übersicht über die Teilnehmenden an der qualitativen Studie.

Nr.	Selbstgewähltes Pseudonym	w/m	Testaufgaben beim Lauten Denken	Leistungsanforderung	Reihenfolge der schulischen Fremdsprachen ³⁶	Datum der Datenerhebung	Erstsprache ³⁷
1	Willy	m	Szenario „Stadttrip“	A	EN vor FR	13.11.2017	Deutsch ³⁸
2	Matthew	m	Szenario „Stadttrip“	C	FR vor EN	14.11.2017	Portugiesisch
3	Arya	w	Szenario „Stadttrip“	A	FR vor EN	15.11.2017	Deutsch
4	Amy	w	Szenario „Stadttrip“	A	EN vor FR	16.11.2017	Deutsch
5	Aria	w	Szenario „Stadttrip“	A	EN vor FR	20.11.2017	Deutsch
6	Elena	w	Szenario „Stadttrip“	C	FR vor EN	21.11.2017	Farsi ³⁹
7	Alison	w	Szenario „Stadttrip“	A	FR vor EN	22.11.2017	Deutsch
8	Billy	m	Szenario „Stadttrip“	B	FR vor EN	07.12.2017	Deutsch
9	Sheldon	m	Szenario „Stadttrip“	A	EN vor FR	08.12.2017	Deutsch
10	Cybermen	m	Szenario „Stadttrip“	B	FR vor EN	12.12.2017	Deutsch
11	Spencer	w	Szenario „Stadttrip“	B	FR vor EN	15.12.2017	Deutsch
12	Howard	m	Szenario „Stadttrip“	B	FR vor EN	11.01.2019	Deutsch
13	Böbu	m	Szenario „Stadttrip“	C	FR vor EN	15.01.2019	Deutsch
14	Tyler	w	Szenario „Stadttrip“	B	FR vor EN	18.01.2018	Deutsch
15	Sofia	w	Szenario „Stadttrip“	C	EN vor FR	22.01.2018	Spanisch ⁴⁰
16	Toby	m	Szenario „Stadttrip“	A	FR vor EN	24.01.2018	Deutsch
17	Cersei	w	Szenario „Stadttrip“	C	FR vor EN	09.02.2018	Deutsch

³⁶ EN vor FR: Englisch als erste schulische Fremdsprache ab der 3. Klasse und Französisch als zweite schulische Fremdsprache ab der 5. Klasse. FR vor EN: Französisch als erste schulische Fremdsprache ab der 3. Klasse und Englisch als zweite schulische Fremdsprache ab der 5. Klasse.

³⁷ Selbstauskunft; angekreuzte bzw. eingetippte Antwort auf die Frage „In welche(r) Sprache(n) hast du als kleines Kind zuerst sprechen gelernt?“.

³⁸ Die Schüler/innen, bei denen in der Tabelle „Deutsch“ steht, haben im Fragebogen die Antwortmöglichkeit „Deutsch bzw. deutscher Dialekt“ ausgewählt oder die Option „andere“ ausgewählt und z.B. „Walliserdeutsch“ eingetippt.

³⁹ Diese Schülerin ging seit 6 Jahren in der Schweiz zur Schule. Sie wurde früher auf Farsi unterrichtet.

⁴⁰ Diese Schülerin besuchte erst seit 3 Jahren eine Schule in der Schweiz und wurde vorher auf Spanisch unterrichtet, sie lernte auch erst seit 3 Jahren Französisch.

Nr.	Selbstgewähltes Pseudonym	w/m	Testaufgaben beim Lauten Denken	Leistungsanforderung	Reihenfolge der schulischen Fremdsprachen ³⁶	Datum der Datenerhebung	Erstsprache ³⁷
18	Mogli	m	Off-Szenario	B/C	EN vor FR	24.11.2017	Deutsch
19	Loup	w	Off-Szenario	A	EN vor FR	24.11.2017	Serbisch
20	Ezra	m	Off-Szenario	B/C	EN vor FR	05.12.2017	Deutsch
21	Jon	m	Off-Szenario	C	FR vor EN	06.12.2017	Deutsch
22	Nancy	w	Off-Szenario	B	FR vor EN	13.12.2017	Italienisch
23	Hector	m	Off-Szenario	B/C	EN vor FR	14.12.2017	Deutsch
24	Bonnie	w	Off-Szenario	B/C	EN vor FR	14.12.2017	Deutsch
25	Dustin	w	Off-Szenario	A	FR vor EN	19.12.2017	Deutsch
26	Picard	w	Off-Szenario	B	EN vor FR	22.12.2017	Deutsch
27	Omega	m	Off-Szenario	B	FR vor EN	11.01.2019	Deutsch
28	Chiara	w	Off-Szenario	B	FR vor EN	18.01.2018	Deutsch
29	Steve	m	Off-Szenario	C	FR vor EN	19.01.2018	Deutsch
30	Eleven	m	Off-Szenario	C	FR vor EN	26.01.2018	Deutsch

Die in Tabelle 4 aufgelisteten Schüler/innen haben, wie bereits erwähnt, neben der über zweistündigen Datenerhebungssitzung mit der Forscherin auch die „herkömmlichen“ Aufgaben zum Hör- und Leseverstehen, die Aufgaben zu verschiedenen Teilkompetenzen und Fragebögen der Hauptstudie bearbeitet. Dadurch ist es möglich, die Kompetenzen der Schüler/innen der qualitativen Untersuchung zu den Kompetenzen der IFB-Stichprobe für Französisch in Beziehung zu setzen, was im nächsten Unterkapitel erfolgt.

3.4.3.2 KOMPETENZEN DER SCHÜLER/INNEN DER QUALITATIVEN STICHPROBE IM VERGLEICH ZUR IFB-STICHPROBE FÜR FRANZÖSISCH

Im Folgenden werden die Testresultate der 30 Lernenden, die an der qualitativen Studie teilgenommen haben, mit denjenigen verglichen, die von allen Teilnehmenden der Französischtests erzielt wurden. Durch diesen Vergleich soll sichergestellt werden, dass sich diese 30 Lernenden nicht erheblich von den anderen Schüler/innen der IFB-Stichprobe unterscheiden. Dafür wurden alle erhobenen Ergebnisse aus den Aufgaben zu den Teilkompetenzen sowie die „herkömmlichen“ Aufgaben zum Hör- und Leseverstehen mittels *Item Response Theory* (IRT) psychometrisch analysiert. Die dabei berechneten *Weighted Likelihood Estimates* (WLEs) nach Warm (1989) geben Auskunft über die relative Fähigkeit, die die Lernenden in dem jeweiligen Test gezeigt haben (z.B. das gemessene Wortschatzwissen). Werte um 0 situieren die Lernenden in der Mitte der getesteten Stichprobe, negative Werte im Bereich der schwächeren und positive im Bereich der stärkeren Schüler/innen. WLEs haben gegenüber anderen Massen der Personenfähigkeit den Vorteil, dass sie auch für Extrem-Ergebnisse (null Punkte bzw. maximale Punktzahl) einen Wert schätzen. Die im Folgenden verwendeten WLEs wurden durch die Schätzung von 2-PL-Modellen (*2 parameter logistic model*) der IRT gewonnen (Birnbaum, 1968). In einem 2-PL-Modell werden zwei Itemparameter geschätzt, die Itemschwierigkeit und die Itemdiskriminierung, d.h. dessen Trennschärfe. Das 2-PL-Modell unterscheidet sich vom einfacheren und bekannteren Rasch-Modell dadurch, dass es genauere Personenfähigkeiten ausgeben kann, wenn die Trennschärfe zwischen den Items variiert und folglich „die Schwierigkeitsreihenfolge der Items für Personen mit verschiedenen Fähigkeiten nicht mehr identisch“ ist (Robitzsch, 2015, S. 13).

In den folgenden Histogrammen wird für jeden Einzeltest die Verteilung der WLEs (x-Achse) über alle Lernenden hinweg dargestellt, die an den Französischtests teilgenommen haben. An der Höhe der Balken ist die Anzahl der Lernenden ersichtlich, die jeweils einem Fähigkeitsbereich zugeordnet wurden.

Darüber werden die einzelnen WLEs der 30 Lernenden aus der qualitativen Studie dargestellt: An der Position des jeweiligen Pseudonyms auf der x-Achse wird sichtbar, welches WLE diesem Lernenden zugeordnet wurde. Die Mitte jedes Pseudonyms befindet sich über dem statistisch geschätzten WLE-Wert der betreffenden Schüler/innen. Die Situierung der Pseudonyme auf der y-Achse ist hingegen nicht relevant. Wichtig ist in der Darstellung lediglich die Stellung der Pseudonyme in Bezug auf die x-Achse, die Auskunft über den WLE-Wert (Kompetenzmass) der Schüler/innen gibt.

Abbildung 8 und Abbildung 9 stellen die Resultate bei den „herkömmlichen“ Lese- und Hörverstehenstests (vgl. Kap. 3.1) dar.

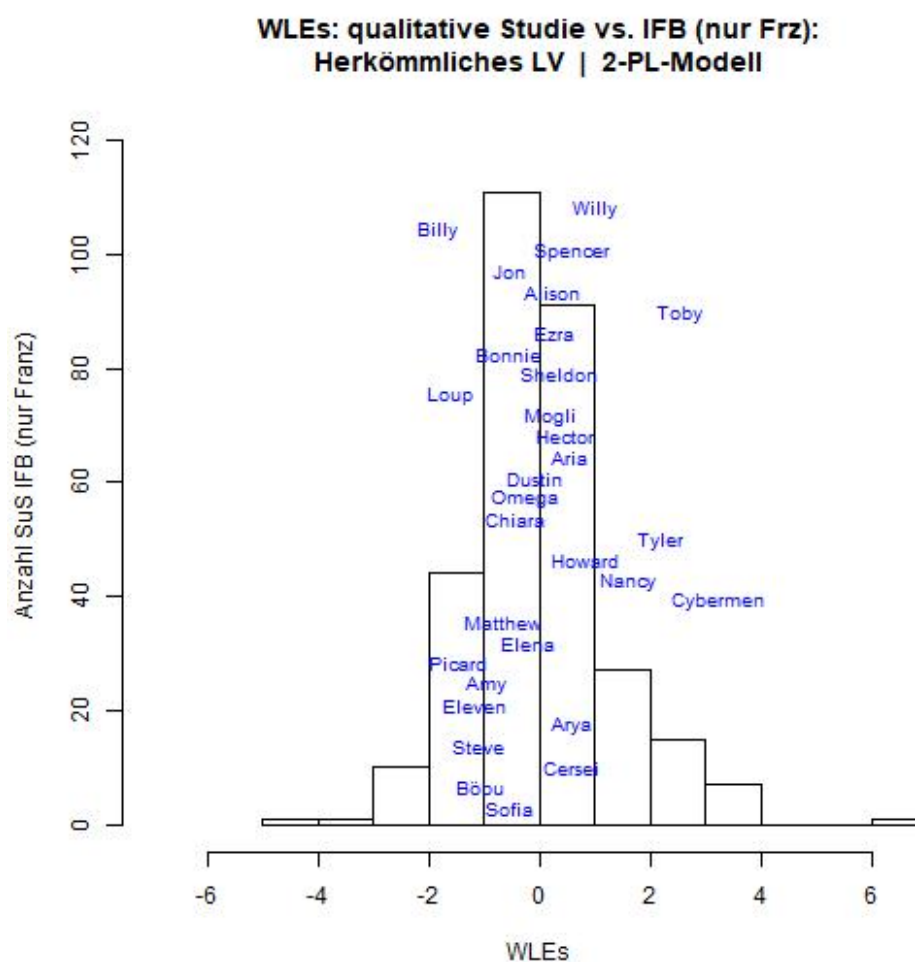


Abbildung 8: „Herkömmliches“ Leseverstehen.

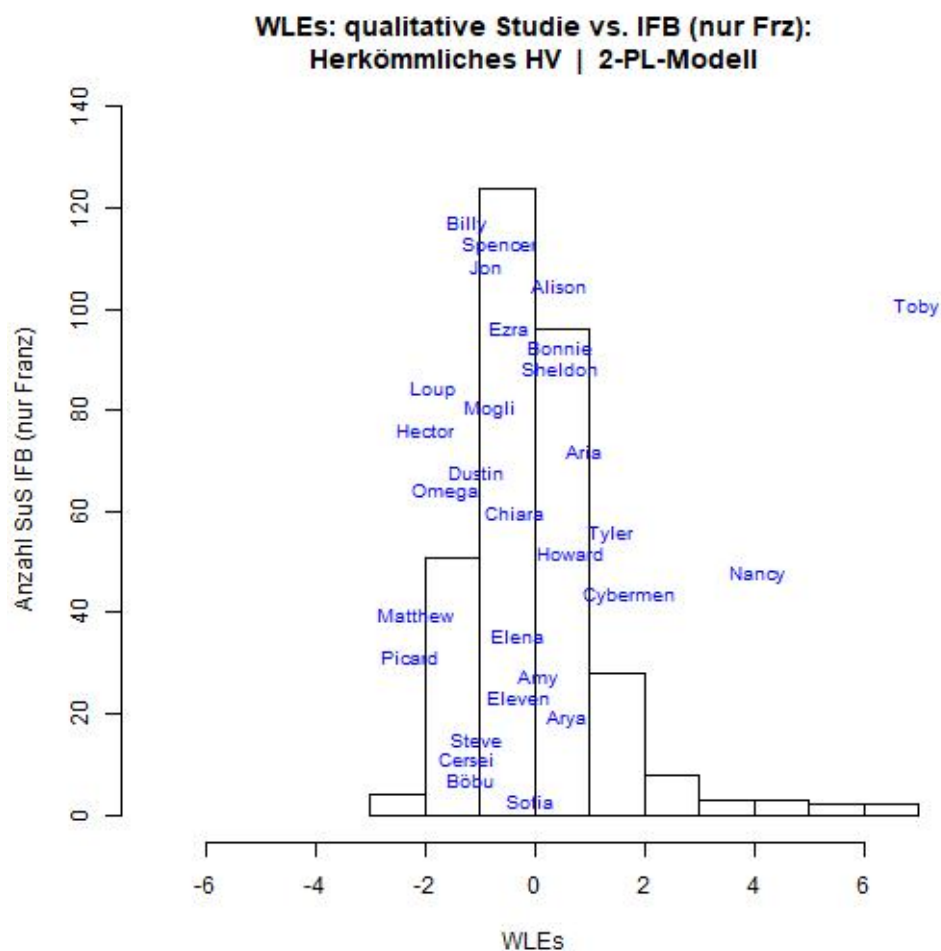


Abbildung 9: „Herkömmliches“ Hörverstehen.

Den Histogrammen ist zu entnehmen, dass die Schüler/innen der qualitativen Studie recht gut über die Gesamtgruppe der IFB-Schüler/innen, die den Französischtest lösten, verteilt sind.⁴¹ Dieses Ergebnis kann als Indiz dafür interpretiert werden, dass die Leistungsbreite der Schüler/innen der qualitativen Stichprobe mit der Leistungsbreite der IFB-Gesamtpopulation ziemlich gut vergleichbar ist. Es gibt also unter den Schüler/innen der qualitativen Stichprobe - wie erhofft - solche, die in Französisch sehr stark, mittelmässig oder eher schwach waren. Somit kann behauptet werden, dass in der qualitativen Studie das Funktionieren der IFB-Testaufgaben mit Schüler/innen untersucht wurde, die nicht nur zu den besten oder schwächsten innerhalb der IFB-Stichprobe für Französisch gehörten.

⁴¹ Auffällig gut hat in den „herkömmlichen“ Lese- und Hörverstehenstests (vor allem im Hörverstehen) der Schüler Toby abgeschnitten, der nicht weit von der deutsch-französischen Sprachgrenze lebt und gemäss eigenen Aussagen auch ausserhalb der Schule viel Kontakt zum Französischen hat. Interessanterweise haben sowohl die Lehrperson als auch er selbst angegeben, er sei – im Vergleich zu den anderen Lernenden in seiner Klasse – nicht besonders gut im Französischen. Dies lässt sich in den Komponententests nur teilweise nachweisen: Zwar sind die Lernenden in Tobys Klasse tatsächlich im Schnitt in allen Tests recht gut, Toby gehört aber meist zu den besten.

Auch die unten abgebildeten Resultate der Tests zu Teilkompetenzen zeigen eine recht breite Streuung der Schüler/innen aus der qualitativen Studie innerhalb der IFB-Teilstichprobe für Französisch. Zu diesen Tests gehören ein Test zur Grammatikalität von Äusserungen (*grammaticality judgement test*), ein Wortschatztest, ein Test zum Sichtwortschatz (*sight-word recognition test*), ein Test zur Segmentierung von gesprochener Sprache und ein nichtverbaler Intelligenztest mit Raven-Matrizen (vgl. Karges, Lenz et al., 2021).

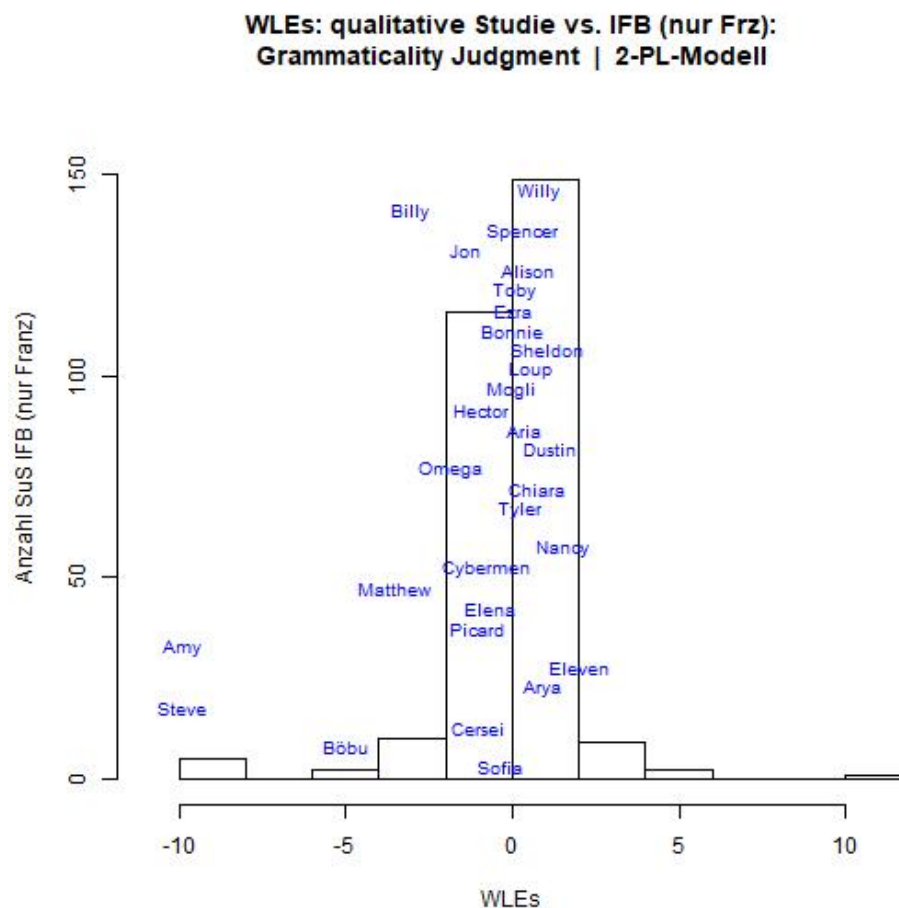


Abbildung 10: *Grammaticality judgement*.

**WLEs: qualitative Studie vs. IFB (nur Frz):
Wortschatz | 2-PL-Modell**

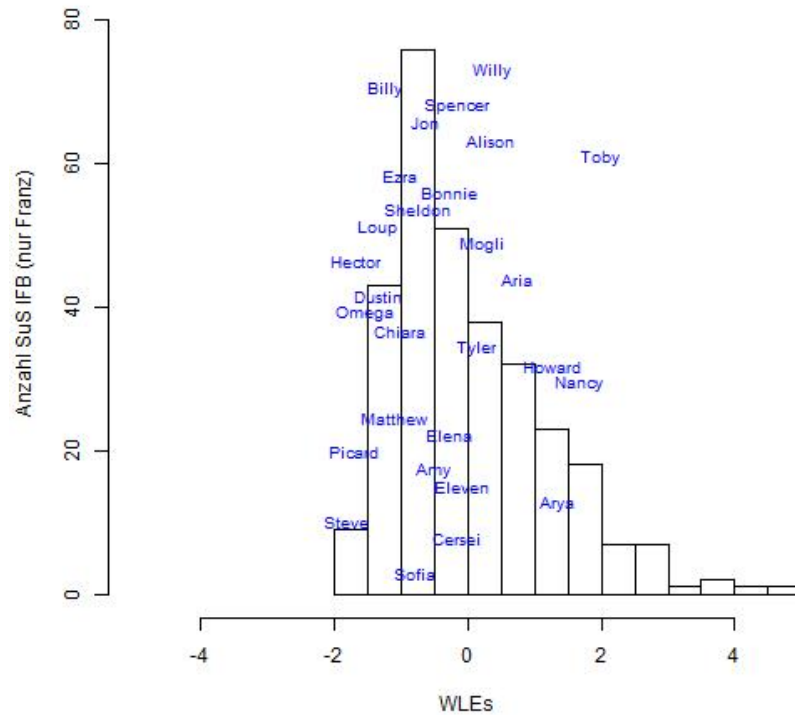


Abbildung 11: Wortschatztest.

**WLEs: qualitative Studie vs. IFB (nur Frz):
Sight-word recognition | 2-PL-Modell**

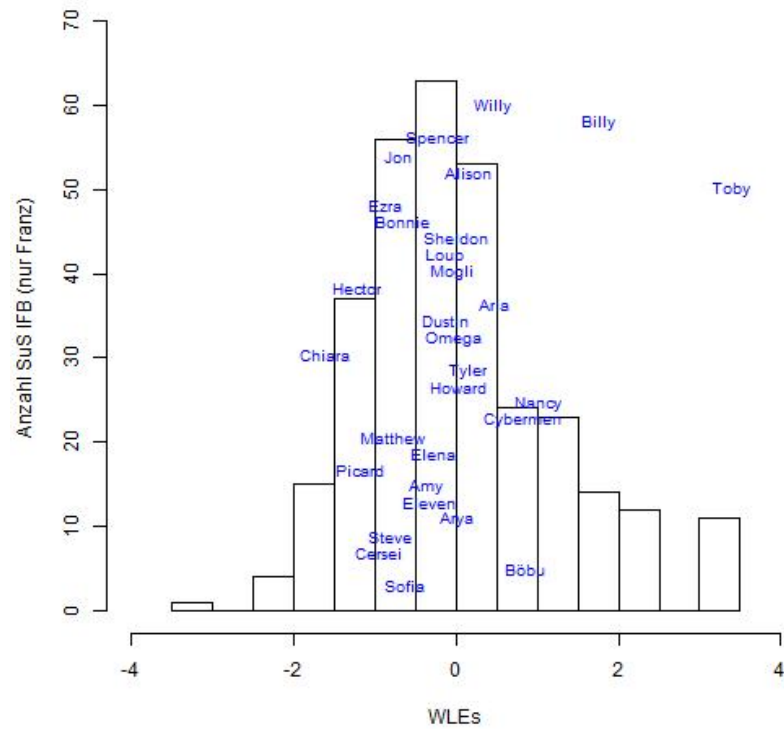


Abbildung 12: Sight-word recognition.

**WLEs: qualitative Studie vs. IFB (nur Frz):
Segmentierung Hören | 2-PL-Modell**

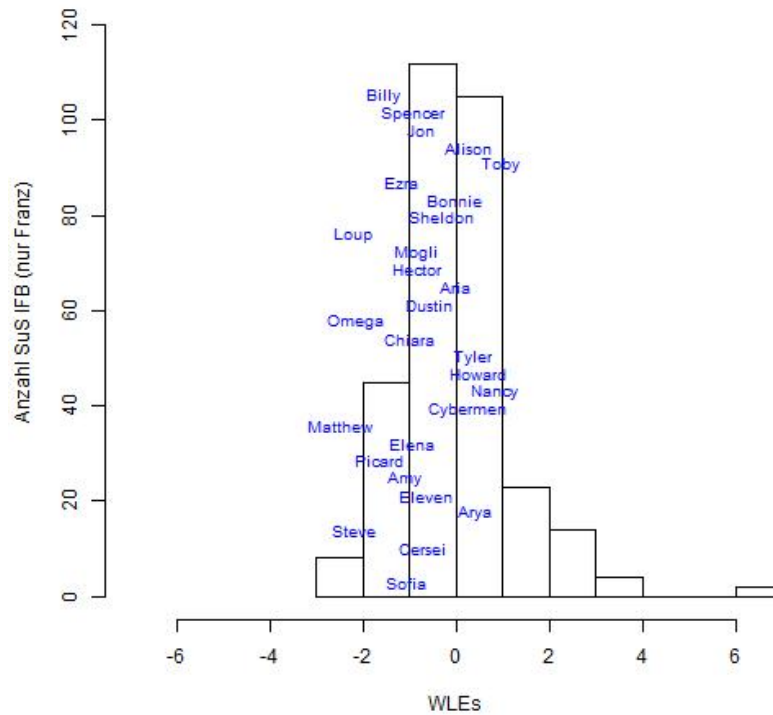


Abbildung 13: Hörverstehen Segmentierung.

**WLEs: qualitative Studie vs. IFB (nur Frz):
Raven Matrizen | 2-PL-Modell**

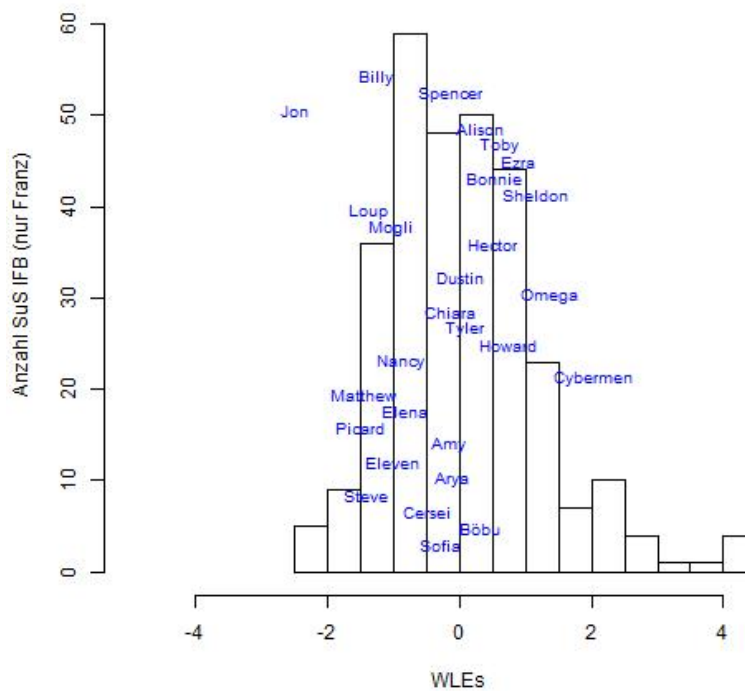


Abbildung 14: Nichtverbaler Intelligenztest mit Ravens Matrizes.

Aufgrund der WLEs aus den verschiedenen Tests wurden für den Vergleich der beiden Teilnehmergruppen auch Rangfolgen gebildet, absteigend von den höchsten zu den niedrigsten WLEs innerhalb eines Tests. Tabelle 5 zeigt den Rang der 30 Schüler/innen aus der qualitativen Studie innerhalb der Teilstichprobe jener IFB-Schüler/innen, die den Französischtest gelöst haben (n=331). Die Tabelle wurde in Microsoft Excel übertragen, nach dem durchschnittlichen Rang geordnet und zur besseren Visualisierung eingefärbt: Grüne Werte stehen für sehr niedrige Ränge (d.h. hohe WLE-Werte im Vergleich zur gesamten Französisch-Stichprobe), rote Werte für sehr hohe Ränge (d.h. sehr tiefe WLE-Werte). Gelbe Zellen schliesslich repräsentieren die Mitte der Skala.

Tabelle 5: Rang der Schüler/innen aus der qualitativen Studie (n=30) in den „herkömmlichen“ Lese- und Hörverstehensaufgaben und in Teilkomponententests im Vergleich zur IFB-Teilstichprobe für Französisch (n=331).

Pseudonym der Schüler/innen ⁴²	Rang „herkömmliches“ HV	Rang „herkömmliches“ LV	Rang <i>grammaticality judgement</i>	Rang Segmentierung HV	Rang <i>sight-word recognition</i>	Rang Wortschatztest	Rang Raven Matrizen
Toby	2	14	150	44	6	21	84
Nancy	7	31	19	50	59	31	237
Willy	NA ⁴³	49	57	NA	93	100	NA
Cyberman	22	6	231	125	70	NA	20
Howard	75	64	NA	94	133	43	75
Arya	80	75	47	104	135	40	149
Aria	51	84	118	149	87	73	NA
Tyler	34	20	135	112	122	117	136
Alison	93	117	95	119	121	101	105
Sheldon	91	103	38	184	138	221	39
Bonnie	89	213	158	151	223	153	92
Spencer	230	74	122	250	161	142	154
Eleven	176	267	11	223	175	134	251
Dustin	278	150	33	211	150	291	144
Mogli	248	119	155	244	143	113	255
Ezra	200	115	154	271	242	257	57
Elena	186	171	223	256	167	151	231
Böbu	288	255	288	NA	63	NA	106
Chiara	190	199	61	259	300	258	152
Sofia	150	211	193	264	216	225	169
Cersei	297	78	249	232	250	141	191
Amy	132	247	291	265	184	182	160
Omega	313	180	276	306	142	303	30
Hector	318	95	247	242	275	306	93
Jon	256	210	261	233	224	201	301
Loup	316	287	86	309	152	294	279
Billy	296	296	282	292	29	286	272
Matthew	320	224	284	311	233	275	285
Steve	277	261	290	307	236	308	281
Picard	321	281	250	294	272	307	288

Die Testergebnisse der Schüler/innen Toby, Nancy, Willy, Cyberman, Howard, Arya, Aria, Tyler und Alison lag bei den meisten Aufgaben im stärksten Bereich (vorwiegend grün). Ein besonders interessantes Ergebnis zeigt sich bei der Schülerin Nancy: Diese mehrsprachige

⁴² Bei der vertiefenden Analyse der Verbalprotokolle wurden die Transkripte der 20 fett markierten Schüler/innen berücksichtigt (mehr dazu in Kap. 3.6).

⁴³ „NA“ bedeutet „not available“. Diese/r Lernende hat diesen Testteil nicht gelöst bzw. die Resultate sind nicht verwertbar.

Schülerin mit Italienisch als L1 gehörte zwar im Raven-Matrizen-Test zu den eher schwächeren Schüler/innen der Teilstichprobe, in allen Tests aber, in denen die Teilkompetenzen in der Fremdsprache Französisch getestet wurden, schnitt sie sehr gut ab. Die Lernenden Picard, Steve, Matthew, Billy, Loup und John gehörten hingegen über alle Tests hinweg zu den schwächsten Schüler/innen (vorwiegend rot), wobei zwei dieser Schüler/innen in einzelnen Tests doch recht gut abschnitten (Billy in *sight-word recognition* und Loup in *grammaticality judgement*).

Betrachtet man die Verteilung der Farben in der Tabelle, so wird deutlich, dass die Lernenden der qualitativen Stichprobe insgesamt alle Kompetenzbereiche abdecken, die in der Teilstichprobe Französisch vertreten waren. Es konnte also gezeigt werden, dass die 30 Schüler/innen, die an der qualitativen Studie teilnahmen, bezüglich ihrer allgemeinen Französischkompetenzen und Intelligenz ähnlich stark variieren wie die gesamte Gruppe der Testteilnehmenden in Französisch. Damit kann vorsichtig vermutet werden, dass die Erkenntnisse aus der qualitativen Befragung die Tendenzen in der grösseren Gruppe der Französischlernenden im IFB Projekt widerspiegeln könnten.

3.4.4 ERHEBUNG DER INTROSPEKTIVEN DATEN

Die Haupterhebung der qualitativen Daten fand zwischen November 2017 und Februar 2018 statt. Während eine Klasse am quantitativen Teil der IFB-Untersuchung teilnahm, arbeiteten der oder die ausgewählte Schüler/in aus dieser Klassen in einem separaten Raum individuell mit der Forscherin (vgl. Abbildung 15).



Abbildung 15: Eine prototypische Datenerhebungssituation, kurz vor der Datenerhebung. Rechts: Sitzplatz des/der Lernenden, links: Sitzplatz der Forscherin.

Der oder die Lernende löste zuerst nach einem anfänglichen Training im Lauten Denken entweder die Aufgaben aus dem Szenario „Stadttrip“ oder die Off-Szenario-Testaufgaben und sprach dabei die eigenen Gedanken laut aus. Danach wurde er oder sie in einem *Stimulated-Recall*-Interview zu seinen oder ihren Lösungsprozessen und dem Verständnis der Inputtexte befragt. Ziel der Triangulation (vgl. Knorr & Schramm, 2016) der beiden

Datenerhebungsverfahren war es, durch die Äusserungen der Schüler/innen in den *Stimulated-Recall*-Interviews einerseits die fragmentarischen Verbalisierungen in den Lautdenkprotokollen valider interpretieren zu können und andererseits durch gezieltes Nachfragen während der Interviews weitere Erkenntnisse zu den Testaufgaben gewinnen zu können, die in den Lautdenkprotokollen nicht zum Vorschein kommen würden.

Die Datensitzungen wurden mithilfe der Software OBS (*OBS Studio*, 2017) und der Bildschirmkamera aufgenommen. Die Aufnahmen zeigen die Gesichter der Schüler/innen sowie alles, was bei der Bearbeitung der Aufgaben auf dem Bildschirm zu sehen war, inklusive des Mauszeigers (vgl. Abbildung 16).

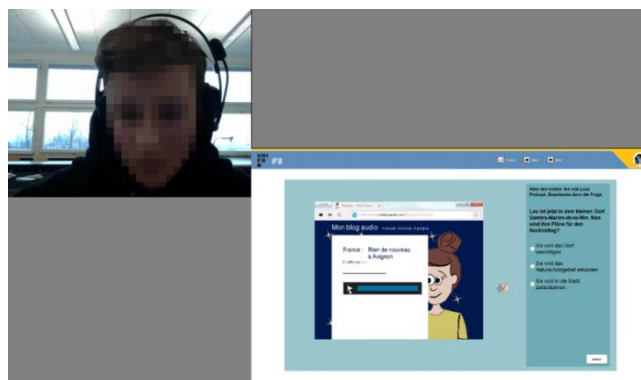


Abbildung 16: Screenshot einer anonymisierten Videoaufnahme aus der Hauptuntersuchung.

Diese individuellen Datenerhebungssessionen, deren Verlauf in folgenden Unterkapiteln näher beschrieben und reflektiert wird, dauerten genauso lange wie die quantitative Untersuchung im Klassenverband am gleichen Halbtage, d.h. insgesamt drei Lektionen à 45 Minuten. Während dieser Zeit konnten die Lernenden an bestimmten Punkten Pausen einlegen⁴⁴, auch wurden ihnen kleine Snacks und Wasser zur Verfügung gestellt. Direkt nach der Datenerhebung wurde ein Protokoll mit den Besonderheiten der Datenerhebungssituation und des Interviewverlaufs angefertigt. Diese Informationen wurden bei Bedarf in das Transkript in Form eines Memos integriert, um das bessere Verständnis der zu analysierenden Daten während der Codierung zu unterstützen.

3.4.4.1 LAUTES DENKEN

Am Anfang jeder Datenerhebung wurden die Schüler/innen über den Ablauf und das Ziel der nächsten drei Lektionen informiert. Es wurde ihnen mitgeteilt, dass 1) sie nach einem Training in ihrem eigenen Tempo einige Französischaufgaben am Computer lösen werden, dabei laut

⁴⁴ Die Lernenden konnten vor oder nach dem Lauten Denken bzw. an einer beliebigen Stelle beim *Stimulated-Recall*-Interview eine Pause einlegen. Es wurde ihnen überlassen, ob und wann sie davon Gebrauch machen wollten.

denken sollen und im Nachhinein Fragen dazu beantworten werden; 2) im Zentrum dieser Untersuchung die neu entwickelten Testaufgaben und die Lösungswege der Schüler/innen beim Bearbeiten dieser Aufgaben sowie ihre Einschätzung dieser Aufgaben, nicht aber ihre Französischkompetenzen stehen; 3) sie für diese Untersuchung zufällig ausgewählt wurden und daran in der Rolle einer Expertin oder eines Experten teilnehmen werden; 4) die Teilnahme an der Untersuchung freiwillig ist; 5) die Testresultate der Lehrperson nicht mitgeteilt werden und auch keinen Einfluss auf die Schulnoten haben; und dass 6) während der Untersuchung Ton- und Videoaufnahmen angefertigt werden, die dann in einer anonymisierten Form für Forschungszwecke verwendet werden. Die Erziehungsberechtigten der betroffenen Schüler/innen waren bereits im Voraus über die Forschungsziele und die Vorgehensweise bei der Datenerhebung schriftlich informiert worden und hatten eine Einwilligung für die Studienteilnahme ihrer Schützlinge unterzeichnet (vgl. Anhang 1). Die Lernenden wurden vor der Datenerhebung auch mündlich gefragt, ob sie mit der Teilnahme einverstanden seien, was bei allen der Fall war. Die Schüler/innen wählten dann ein Pseudonym, unter dem sie in dieser Studie figurieren.

Anschliessend wurde ein ca. 20-minütiges Training im Lauten Denken durchgeführt. Das Training war weitgehend standardisiert und wurde eingeplant, um die Validität der erhobenen Daten zu erhöhen (vgl. Bowles, 2010; Heine & Schramm, 2016). Zuerst wurde den Lernenden die Instruktion zum Lauten Denken vorgelesen. Grosse Teile dieser Instruktion wurden aus Rossa (2012, S. 103) sowie aus Heine und Schramm (2007, S. 178) wortwörtlich übernommen, um Hinweise zur Erhebung der Lautdenkdaten aus der Fachliteratur ergänzt (v.a. Bowles, 2010; Heine, 2014) und aufgrund der Erfahrungen aus der Pilotierung sowie im Hinblick auf die Besonderheiten der Datenerhebungssituation wo nötig angepasst. Die Instruktion lautete:

Sag' mir bitte alles, das dir durch den Kopf geht, während du die folgenden Aufgaben bearbeitest, wenn du nicht nur einfach die Wörter des Textes liest oder hörst⁴⁵. Das Wichtigste ist, laut zu reden, weil ich alles mitbekommen möchte, egal wie unwichtig

⁴⁵ Der Anfang lautete ursprünglich wie in Rossa (2012, S. 103): „Sag' mir bitte alles, das dir durch den Kopf geht, während du die folgenden Aufgaben bearbeitest. Das wichtigste ist, laut und **dauernd** zu reden...“ (Hervorhebung M.B.). Als Ergebnis der Pilotierung wurde die hervorgehobene Stelle entfernt und die Formulierung „wenn du nicht nur einfach die Wörter des Textes liest oder hörst“ hinzugefügt. In der Pilotierung zeigte sich nämlich, dass manche Schüler/innen die ursprüngliche Formulierung so interpretierten, dass sie beim Bearbeiten der Leseaufgaben die französischen Inputtexte *vorlesen* sollen. Das Vorlesen war hier jedoch nicht intendiert, denn, wie Alderson anmerkt, „reading aloud is not the 'normal' way in which people read, and the process of reading aloud may be very different from reading silently[,] [...] such externalising risks distorting and changing the nature of the process“ (2000, S. 4). Auch dachten manche Lernende, dass sie beim Hören der Inputtexte *gleichzeitig* sprechen sollen. Infolgedessen bekamen sie die Hälfte des Inputtextes nicht mit und konnten dann die Aufgabe auch nicht lösen.

es dir erscheinen mag. Mich interessiert nicht so sehr, ob du die richtige Lösung findest, sondern all deine Versuche, alles, was dir in den Sinn kommt, egal, ob du deine Ideen gut oder weniger gut findest. Ich zähle nicht, wie oft du etwas falsch verstehst oder wie oft du etwas nicht kannst. Versuche nicht, jemandem etwas zu erklären. Du solltest nicht planen, was du sagen sollst, sondern deine Gedanken frei fließen lassen und einfach alles aussprechen, so wie es dir durch den Kopf geht. Stell dir vor, du bist alleine im Raum und sprichst einfach mit dir selbst. Es ist egal, ob du auf Deutsch, Schweizerdeutsch oder Französisch sprichst, du kannst auch gerne jederzeit zwischen den Sprachen wechseln. Wenn du einen Moment lang nichts sagst, werde ich dich daran erinnern, dass du laut denken sollst.

Sprich also laut, wenn du z.B.⁴⁶:

- über den Inhalt der Texte oder über die Fragen nachdenkst; oder*
- wenn du dir überlegst, welche Antwort die richtige ist;*
- wenn du dir den Inhalt der Texte noch mal auf Französisch oder Deutsch zusammenfasst;*
- wenn du Schwierigkeiten hast, etwas zu verstehen, oder dich etwas fragst;*
- wenn du Aha-Erlebnisse hast;*
- wenn du etwas wichtig findest;*
- wenn du Verbindungen mit dem erkennst, was du vorher gelesen hast;*
- wenn dir etwas in den Sinn kommt, was mit der Aufgabe zu tun hat, z.B. aus der Schule, aus den Ferien usw.;*
- wenn du dich über etwas freust oder ärgerst.*

Damit ich weiss, wo du im Text bist, zeig bitte immer mit der Maus auf die Textstelle oder Frage, über die du gerade nachdenkst⁴⁷.

Als erste Trainingsaufgabe, bei der die Lernenden das Laute Denken ausprobieren konnten, wurden in Anlehnung an Rossa (2012, S. 104) sieben Spielkarten aus dem Gesellschaftsspiel *Anno Domini: Erfindungen*⁴⁸ eingesetzt. Bei dieser Aufgabe mussten die zufällig gewählten Karten mit Erfindungen, deren Jahreszahl verdeckt war, in eine chronologische Reihenfolge gebracht werden. Zuerst wurde die Vorgehensweise des Auflegens von Karten und des gleichzeitigen Lauten Denkens von der Forscherin demonstriert. Es wurde dabei bewusst auf

⁴⁶ Die Beispiele wurden aufgrund der Ergebnisse der Pilotierung hinzugefügt, weil es sich dort zeigte, dass einige Schüler/innen trotz der Instruktion und dem Training immer noch nicht wussten, was von ihnen beim Lauten Denken eigentlich erwartet wird.

⁴⁷ Dieser Satz wurde im Hinblick auf die spätere Transkription und Datenanalyse hinzugefügt. Die Lernenden wurden aber mündlich informiert, dass es auch nichts ausmacht, wenn sie es nicht umsetzen, weil sie z.B. nicht daran denken oder es als unpraktisch erachten.

⁴⁸ Vgl. [https://de.wikipedia.org/wiki/Anno_Domini_\(Spiel\)](https://de.wikipedia.org/wiki/Anno_Domini_(Spiel)) (Abgerufen am 18.12.2020).

das Vorspielen eines aufgenommenen Beispiels verzichtet, was zwecks der höheren Standardisierung der Datenerhebungssituation z.B. von Heine und Schramm (2007, S. 178) empfohlen wird. Die Jugendlichen begegneten der Forscherin am Datenerhebungstag zum ersten Mal und sollten während der nächsten drei Lektionen in ihrer Anwesenheit intensiv arbeiten bzw. von ihr befragt werden. Es wurde aus diesem Grund als wichtig erachtet, dass die Lernenden während der begrenzten Trainingszeit vor der eigentlichen Untersuchung eine gute Beziehung mit der Forscherin aufbauen können und eine persönliche Demonstration wurde dafür als besser geeignet erachtet als das Zeigen eines Videobeispiels. Nach der Demonstration konnten die Lernenden das Spielen und Laute Denken selber ausprobieren.

Als zweite Trainingsaufgabe wurden zwei Lese- und zwei Hörverstehenstexte aus den IFB-Szenarien eingesetzt, die bei der qualitativen Haupterhebung nicht verwendet wurden. Dadurch sollten die Lernenden mit der Bedienung der computerbasierten Testaufgaben vertraut werden. Auch sollten sie dabei die Möglichkeit erhalten, das Laute Denken auch beim Lösen der Hörverstehensaufgaben auszuprobieren. Dadurch sollte vermieden werden, dass die Lernenden während der Haupterhebung beim Hören der Inputtexte gleichzeitig ihre Gedanken zu verbalisieren versuchen.

Als Erinnerung an das Laute Denken wurde den Schüler/innen ein Blatt mit der Anforderung zur Verbalisierung gezeigt (vgl. Abbildung 17). Die Lernenden wurden bereits beim Training mit dem Blatt und seiner Bedeutung bekannt gemacht.



Was geht dir durch den Kopf?

Abbildung 17: Blatt mit der Aufforderung zum Lauten Denken.⁴⁹

Die Forscherin gab den Schüler/innen während und nach dem Training Feedback zu ihrer Vorgehensweise und beantwortete bei Bedarf auch ihre Fragen dazu. Bevor die Lernenden gleich im Anschluss an das Training mit der Bearbeitung der eigentlichen Testaufgaben begannen, wurde ihnen zur Erinnerung die Instruktion zum Lauten Denken vorgelegt, mit der Bitte, diesen Text diesmal für sich still zu lesen. Nachdem gegebenenfalls letzte Fragen nach

⁴⁹ Bildquelle: <https://pixabay.com/de/illustrations/hatena-denken-sie-an-frage-die-m%C3%BChe-1184896/> (Abgerufen am 18.12.2020).

dem Training und Lesen der Instruktion beantwortet worden waren, begannen die Lernenden mit der Bearbeitung der Szenario- bzw. Off-Szenario-Aufgaben und dem gleichzeitigen Lauten Denken. Dieser Teil der Untersuchung dauerte ca. 30 Minuten pro Schüler/in. Die Forscherin sass etwa zwei Meter von dem/r Schüler/in entfernt. Um die Schüler/innen nicht abzulenken bzw. nicht zusätzlich zu stressen war ihr Stuhl so gerichtet, dass kein Augenkontakt zwischen ihr und den Schüler/innen bestand. Während die Schüler/innen den Test lösten, machte die Forscherin keine Kommentare zu ihrem Vorgehen und beantwortete bei Bedarf nur Fragen, die sich auf die technische Handhabung der computerbasierten Testaufgaben bezogen. Es lässt sich natürlich trotzdem nicht ausschliessen, dass die Anwesenheit der Forscherin die Vorgehensweise der Schüler/innen bei der Bewältigung der Testaufgaben beeinflusste. Auch die Tatsache, dass sie ihre Gedanken laut aussprechen mussten, hatte vermutlich einen Einfluss auf die Vorgehensweise der Schüler/innen (mehr dazu in Kap. 5.3.2).

Während der Datenerhebung stellte sich heraus (wie aufgrund der Erkenntnisse aus der Fachliteratur zu vermuten war), dass sich trotz des Trainings nicht alle Schüler/innen gleich gut an das Laute Denken gewöhnen konnten. Einige Lernende verbalisierten eher wenig und eine Schülerin wollte ihre Gedanken überhaupt nicht laut aussprechen, was dazu führte, dass sie aus der Stichprobe ausgeschlossen werden musste (vgl. Kap. 3.6). Auch machten einige Schüler/innen beim Lauten Denken metakognitive Kommentare, die sich – anders, als im Training und in der Instruktion kommuniziert – wohl nicht an sie selbst, sondern eher an die Forscherin richteten. So sagte beispielsweise Schülerin Cersei, dass sie beim Lösen einer Testaufgabe immer zuerst die auf Deutsch vorhandenen Antwortoptionen für sich auf Französisch zu übersetzen versucht, um danach im Inputtext suchen zu können (Pos. 65-68 im Transkript).

3.4.4.2 *STIMULATED-RECALL-INTERVIEWS*

Direkt nach der Bearbeitung der Testaufgaben und einer eventuellen kurzen Pause wurden die Lernenden zu ihrer Vorgehensweise beim Lösen der Aufgaben befragt. Als Stimulus dienten dabei die Bildschirmseiten, auf denen die Testaufgaben und die von Schüler/innen ausgewählten Antwortoptionen zu sehen waren.

In den *Stimulated-Recall*-Interviews wurden den Lernenden Fragen gestellt, die neben der Erhebung ihrer *recall thoughts* auch bzw. vor allem auf das Erfassen ihrer *hindsight reports* (Gass & Mackey, 2017) abzielten (vgl. Kap. 3.3). Laut Gass und Mackey (2017, S. 44–45, vgl. auch S. 50) eignen sich nur die *recall thoughts* für die Ermittlung der tatsächlichen Denkprozesse bei der Ausführung einer Handlung. In der vorliegenden Studie sollten jedoch die Forschungsteilnehmenden bereits beim simultanen Lauten Denken ihre ungefilterten

Gedanken verbalisieren, die sie beim Lösen der Testaufgaben hatten. Die *Stimulated-Recall*-Interviews dienten hier folglich zum einen der Validierung von Interpretationen aufgrund der Lautdenkprotokolle im Prozess der Datenanalyse und zum anderen der Erfassung ganz bestimmter, in den Lautdenkprotokollen wahrscheinlich nicht abgebildeter und auch postaktionaler Gedanken und Meinungen der Schüler/innen. Aus diesem Grund wurde in der vorliegenden Studie auch nicht angestrebt, in diesen Interviews vor allem die *recall thoughts* der Schüler/innen zu erheben. Man könnte deswegen auch argumentieren, dass die in dieser Studie durchgeführten Interviews gemäss dem Klassifizierungsvorschlag von Heine und Schramm (2016) eigentlich als „retrospektive Befragungen“ angesehen werden könnten. Da den Ausgangspunkt der Gespräche mit den Schüler/innen immer ein Stimulus in Form von bearbeiteten Testaufgaben (vgl. Gass & Mackey, 2017, S. 44) bildete und die Interviews immer direkt nach der Bearbeitung der Testaufgaben durchgeführt wurden, wurde jedoch entschieden, sie in dieser Studie trotzdem als *Stimulated-Recall*-Interviews zu bezeichnen.

Die Fragen, die den Schüler/innen in den Interviews gestellt wurden, dienten in erster Linie der Beantwortung der Forschungsfragen der vorliegenden Untersuchung (vgl. Kap. 3.2). Ein weiteres Ziel der Interviews war aber auch, zusätzliche und für das gesamte Projekt IFB interessante Daten zu erheben, z.B. zur Benutzerfreundlichkeit, zur Schwierigkeit der Testaufgaben und der Inputtexte, zur Einschätzung der Aufgabenformate durch die Schüler/innen, zum Einsatz von digitalen Textsorten in den Testaufgaben, zur Benutzung der Schulsprache in den Aufgabeninstruktionen oder auch zu den Themen, die in den Aufgaben vorkamen. Auf die Antworten der Lernenden auf diese zusätzlichen Fragen wird jedoch in dieser Arbeit nicht weiter eingegangen.

Als allgemeiner Einstieg in die Besprechung jeder Bildschirmseite mit den Testaufgaben wurden die Schüler/innen gefragt, was sie jeweils ganz genau gemacht hatten, um die Aufgabe zu lösen. Wenn nötig, d.h. wenn von den Lernenden nicht schon von alleine erwähnt, wurden ihnen weiterführende Fragen gestellt, z.B. dazu, wie sie beim Lesen oder Hören der Inputtexte vorgegangen waren, wo sie ihre Antwort gefunden hatten, ob sie evtl. andere Antworten in Betracht gezogen hatten, oder ob sie sicher waren, dass die von ihnen gewählte Antwort auch stimmt. Die Schüler/innen wurden dabei bewusst nicht explizit nach ihren Strategien beim Lösen der Aufgaben gefragt. Dadurch sollte vermieden werden, dass die Schüler/innen nur über das berichten, was sie persönlich unter Strategien verstehen und dabei evtl. auch andere interessante Phänomene unerwähnt lassen.

Die Schüler/innen, die das Szenario „Stadttrip“ bearbeiteten, wurden zusätzlich gefragt, wie sie die einleitenden Seiten am Anfang jedes Testteils fanden. Um einen Hinweis darauf zu bekommen, ob die Schüler/innen das Szenario als solches überhaupt bemerkten, wurden sie

am Ende der Besprechung gefragt, ob ihnen an diesem Test etwas Spezielles aufgefallen war. Wenn sie dies verneinten, wurden sie gefragt, ob ihnen bei diesen Aufgaben vom Thema her etwas Spezielles aufgefallen war. Wenn sie auch diese Frage negativ beantworteten, wurden sie in einem dritten Schritt explizit gefragt, ob ihnen aufgefallen war, dass es in allen Testaufgaben um die Planung eines Ausflugs nach Avignon ging. Die Schüler/innen, die Off-Szenario-Aufgaben lösten, wurden ihrerseits gefragt, wie sie es fanden, dass es in allen Aufgaben um ein anderes Thema ging.

Nachdem jede Bildschirmseite mit der gelösten Testaufgabe besprochen worden war, wurden die Schüler/innen in Anlehnung an Rossa (2012) gebeten, den von ihnen verstandenen Inhalt des Inputtextes in der Schulsprache bzw. in ihrem Dialekt zusammenzufassen. Dieser Schritt war den Lernenden vor der Bearbeitung der Testaufgaben nicht angekündigt worden, um ihr Testverhalten nicht zusätzlich zu beeinflussen. Für die mündliche Zusammenfassung des Inputtextes konnten die Schüler/innen den Lesetext nochmals lesen bzw. den Hörtext nochmals hören, wobei sie diesmal die Möglichkeit hatten, die Audioaufnahmen an einer beliebigen Stelle zu stoppen. Dieser Schritt wurde eingeplant, um die anschliessende Datenanalyse zu erleichtern. Die Erkenntnisse aus dem früheren KFM-Projekt „Task Lab“ zeigten nämlich, dass Schüler/innen mit einer eher schwach ausgeprägten Französischkompetenz bei der Rezeption französischer Inputtexte oft etwas anderes verstanden, als im Text zu lesen war. Aufgrund der Äusserungen der Schüler/innen konnte in solchen Fällen im Nachhinein nicht mit Sicherheit festgestellt werden, was die Lernenden im Text konkret verstanden hatten und wo evtl. Probleme des falschen Textverständnisses zu suchen sind (vgl. Barras, 2018).

Die *Stimulated-Recall*-Interviews wurden, wie bereits erwähnt, direkt nach dem Lösen der Testaufgaben durchgeführt. Alle Lernenden wurden nach dem Lautdenken-Teil gefragt, ob sie vor dem Interview eine kurze Pause einlegen möchten. Es zeigte sich, dass nicht alle Schüler/innen dies für nötig hielten. Da die gesamte Datenerhebung drei aufeinanderfolgende Lektionen dauerte und für die Lernenden wie auch für die Forscherin sehr intensiv war, wurden beide Parteien im Verlauf des *Stimulated-Recall*-Interviews jedoch irgendwann von Müdigkeit eingeholt. Es ist daher nicht auszuschliessen, dass dies einen unerwünschten Einfluss auf die Art und Weise der Gesprächsführung durch die Forscherin sowie auf die Aussagen der Schüler/innen hatte.

3.5 DATENAUFBEREITUNG

Die Lautdenkprotokolle und *Stimulated-Recall*-Interviews aller 30 Schüler/innen wurden vollständig und wie in Rossa (2012) nach den Regeln des GAT 2-Basistranskripts transkribiert.

Das Basistranskript stellt eine Erweiterung der Regeln des GAT 2-Minimaltranskripts dar (Selting et al., 2009). Die Regeln des Basistranskripts (vgl. ebd., S. 369–377) wurden um eine Ausnahme ergänzt: Wenn die Schüler/innen etwas im Dialekt sagten, wurden ihre Äusserungen in Standarddeutsch transkribiert und die betroffene Stelle wurde mit zwei Sternchen markiert, z.B. *das war cool* (und nicht etwa das isch kul gsi) (vgl. Anhang 4). Stellen, die evtl. auf die Identität der Schüler/innen schliessen lassen könnten, wurden bereits beim Transkribieren anonymisiert.

Als Transkriptionssoftware wurde MAXQDA (VERBI Software, 2017) eingesetzt. Die neuere Version von MAXQDA (VERBI Software, 2019) wurde für die spätere Analyse der Daten verwendet. Dies hat sich als äusserst vorteilhaft erwiesen, weil sich bei den in MAXQDA angefertigten Transkripten durch das Einfügen einer Zeitmarke eine Verbindung zwischen den Originaldaten (Video und Audio) und dem transkribierten Text herstellen lässt. Somit konnten die Transkripte bei Bedarf auch zu einem späteren Zeitpunkt, d.h. erst beim Codieren der Daten, wo nötig noch ergänzt und verfeinert werden.

Die Transkriptionen wurden von drei studentischen Hilfskräften angefertigt, die speziell für diese Aufgabe im Projekt IFB angestellt und aus den Projektgeldern entschädigt wurden. Alle drei studierten im Bachelor bzw. im Master das Fach Deutsch als Fremdsprache/Deutsch als Zweitsprache und sprachen als Erstsprache einen Schweizer Dialekt. Sie wurden vor dem Beginn der Transkriptionsarbeit mit den Transkriptionsrichtlinien bekannt gemacht, bekamen Feedback zu ihren ersten Transkriptionsversuchen und tauschten sich auch zu dritt regelmässig aus.

Alle Transkripte wurden anschliessend von der Forscherin überprüft und segmentiert. Die Segmentierung orientierte sich in Anlehnung an Rossa „an der Untersuchungseinheit einer ‚Proposition‘ (vgl. Kintsch & van Dijk, 1978)⁵⁰ und definierte als maximale Grösse eines Datensegments zwei Propositionen“ (2012, S. 106–107). Die Segmentierung hatte lediglich zum Ziel, die Lesbarkeit der Transkripte zu verbessern.⁵¹ Bei der Kontrolle der Transkripte wurde die Dauer der ungefüllten Pausen beim Lauten Denken sehr genau gemessen und ergänzt. Beim *Stimulated Recall* wurde auf die Länge der Pausen wegen dem anderen Charakter der erhobenen Daten (d.h. es handelte sich dabei um die Aussagen im Interview und nicht wie beim Lauten Denken um Verbalisierungen beim Lösen der Aufgaben, wo die

⁵⁰ Kintsch & van Dijk (1978, S. 367) definieren die Propositionen wie folgt: „Propositions are composed of *concepts* (...) each proposition must include first a predicate, or *relational concept*, and one or more *arguments*. The latter may be concepts or other embedded propositions. The arguments of a proposition fulfill different semantic functions, such as agent, object, and goal. Predicates may be realized in the surface structure as verbs, adjectives, adverbs, and sentence connectives“ (Hervorhebungen im Original).

⁵¹ Die Segmente sind daher nicht mit den Codiereinheiten gleichzusetzen, vgl. Kap. 3.7.2.

Pausen eine wichtigere Rolle spielen) und auch aus Ressourcengründen weniger geachtet, ausserdem konnte bei der Analyse der Interviews wenn nötig immer das Video konsultiert werden. Die Transkripte des Lauten Denkens wurden ausserdem bei der Kontrolle um Informationen darüber ergänzt, worauf die Lernenden zu diesem Zeitpunkt gerade geblickt bzw. was sie gerade gelesen hatten (Aufgabenstellung vs. Text).

Nach der Übertragung der Transkripte aus dem MAXQDA-Projekt in eine Word-Datei umfassten die 30 Transkripte stolze 1647 A4-Seiten (Schrift „Courier New“, Grösse 10). Es war klar, dass der Umfang der vertieft zu analysierenden Daten aus Ressourcengründen reduziert werden musste. Im nächsten Kapitel wird skizziert, wie dabei vorgegangen wurde.

3.6 DATENGRUNDLAGE FÜR DIE ANALYSE

Wie bereits oben erwähnt wurde schon bei der Planung der Datenerhebung damit gerechnet, dass die erhobenen Daten zwar sehr umfangreich ausfallen, aber ein Teil der erhobenen Daten für bestimmte Analyseteile unbrauchbar sein würde – sei es aus technischen Gründen oder weil sich nicht alle Untersuchungsteilnehmende für die Erhebung der Lautdenkdaten gut eignen würden. Diese Vermutung bestätigte sich auch tatsächlich. So weigerte sich die Schülerin Amy trotz Erklärungen der Forscherin und trotz Training, beim Lauten Denken ihre Gedanken zu verbalisieren. Auch traten bei einigen Datensitzungen Schwierigkeiten auf, die einen negativen Einfluss auf die Validität und Vergleichbarkeit der erhobenen Daten hatten. So musste bei den Schülerinnen Spencer und Sofia der *Stimulated-Recall*-Teil aus technischen Gründen (System-Absturz) mit einem „leeren“ Test ohne den vorher ausgewählten Antworten der Schülerinnen durchgeführt werden. Die Interview-Bedingungen waren hier demzufolge anders als bei den anderen Schüler/innen, die bei diesem Untersuchungsteil ihre Antworten sehen und auch kommentieren konnten. Ein weiterer Schüler (Billy) wurde während der Datenerhebung stark von seinen Mitschüler/innen gestört, die regelmässig in das Zimmer kamen, in dem die Untersuchung durchgeführt wurde. Aus diesen Gründen wurde entschieden, diese vier Schüler/innen aus den vertiefenden Analysen auszuschliessen.

In einem zweiten Schritt wurde entschieden, den nach wie vor sehr umfangreichen Datensatz weiter zu reduzieren und die eigentlichen Analysen mit den Daten von insgesamt 20 Schüler/innen durchzuführen. Die Entscheidung, welche sechs weiteren Schüler/innen aus den vertiefenden Analysen ausgeschlossen werden sollten, wurde prinzipiengeleitet getroffen. So wurde entschieden, dass die Hälfte der Lernenden (n=10) das Szenario und die andere Hälfte die Off-Szenario-Aufgaben bearbeitet haben sollte. Weiterhin sollten in beiden Unterstichproben je fünf Mädchen und je fünf Jungen vertreten sein. Nach der Unterteilung

der Stichprobe in diese Gruppen wurde anschliessend entschieden, diejenigen Schüler/innen auszuschliessen, die im Vergleich zu anderen Lernenden bei den Lautdenken-Sitzungen wenig verbalisiert hatten (Matthew, Tyler, Jon) oder von denen bestimmte Teile weiterer Tests fehlten (Willy, Hector). Schliesslich wurde eine weitere Schülerin (Chiara) ausgeschlossen, bei der aus technischen Gründen das Transkript mit den Zeitmarken, die die Transkriptstellen mit den Videoaufnahmen im Analyseprogramm MAXQDA verbinden, in zwei Dateien vorliegt, was sich bei der Codierung als äusserst unpraktisch erwies.

Die auf 20 Schüler/innen reduzierte Stichprobe setzt sich wie folgt zusammen:

Tabelle 6: Schüler/innen, die für die vertiefende Analyse ausgewählt wurden.

	Szenario „Stadttrip“	Off-Szenario
w	Alison Aria Arya Cersei Elena	Bonnie Dustin Loup Nancy Picard
m	Böbu Cyberman Howard Sheldon Toby	Eleven Ezra Mogli Omega Steve

Die Analyseeinheit (vgl. Kuckartz, 2018, S. 30–31) bildete das Transkript der Verbalprotokolle dieser 20 Schüler/innen.

Um diesen immer noch sehr umfangreichen Datensatz weiter zu reduzieren, wurde beschlossen, aus dem Szenario „Stadttrip“ nur die ersten zwei Bildschirmseiten pro Aufgabe zu analysieren. Dies lässt sich vor allem dadurch begründen, dass die Off-Szenario-Schüler/innen ebenfalls nur die ersten beiden Bildschirmseiten einer Aufgabe bearbeiteten. Weiter wurde entschieden, die Verbalprotokollteile zur jeweils letzten, noch experimentellen Aufgabe (Task D, vgl. Kap. 3.4.2) bei allen Schüler/innen aus dem Korpus auszuschliessen. Diese Aufgabe wurde auch deswegen ausgeschlossen, weil sie von den Schüler/innen zuletzt bearbeitet wurde und aus Zeitgründen beim *Stimulated-Recall*-Teil oft nur kurz und oberflächlich besprochen werden konnte.

Um das Korpus der Szenario- und Off-Szenario-Items in Bezug auf die Itemspezifikationen (d.h. intendierten Handlungsziele und Fertigkeiten) vergleichbar zu machen, wurde schliesslich auch die vorletzte Aufgabe (Szenario „Waitaha“, Task C) aus der Analyse der Off-Szenario-Items ausgeschlossen. Der so entstandene Datensatz, der die Grundlage für die vertiefenden Analysen der Strategien der 20 Schüler/innen darstellte, setzt sich damit aus den Verbalprotokollen zu folgenden Aufgaben zusammen:

Tabelle 7: Tasks, die für die vertiefende Analyse ausgewählt wurden.

Szenario „Stadttrip“	Task	Fertigkeit	Handlungsziel
	Task A (4 Items auf 2 Bildschirmseiten)	Leseverstehen	Einen schriftlichen Dialog mit Empfehlungen verstehen
	Task B (2 Items auf 2 Bildschirmseiten)	Hörverstehen	Einen längeren narrativen Text in groben Zügen verstehen
	Task C (6 Items auf 2 Bildschirmseiten)	Leseverstehen	Ein Suchresultat aufgrund von vorgegebenen Kriterien auswählen
Off-Szenario	Szenario „Walkies“, Task C (2 Items auf 2 Bildschirmseiten)	Leseverstehen	Einen schriftlichen Dialog mit Anweisungen verstehen
	Szenario „Stadttrip“, Task B (2 Items auf 2 Bildschirmseiten)	Hörverstehen	Einen längeren narrativen Text in groben Zügen verstehen
	Szenario „Berufswahl“, Task C (6 Items auf 2 Bildschirmseiten)	Leseverstehen	Ein Suchresultat aufgrund von vorgegebenen Kriterien auswählen

Die Reduktion der Daten bewirkte, dass anstelle von 1647 A4-Seiten, die die 30 Transkripte nach der Übertragung aus dem MAXQDA-Projekt in eine Word-Datei umfassten (Schrift „Courier New“, Grösse 10), neu 20 Transkripte im Umfang von 972 A4-Seiten analysiert werden mussten. Die 20 Transkripte wurden bei der Analyse vollständig zur Kenntnis genommen, um sicherzustellen, dass dabei alle wichtigen (allgemeinen oder auf die für die Analyse ausgewählten Tasks bezogenen) Informationen berücksichtigt werden. Nicht vertieft untersucht wurden aber dabei, wie oben geschildert, bei den Szenario-Aufgaben die letzten Bildschirmseiten zu Tasks A, B, und C sowie Task D und bei den Off-Szenario-Aufgaben die letzten Bildschirmseiten zu Tasks aus den Szenarien „Walkies“, „Stadttrip“ und „Berufswahl“ sowie Tasks aus den Szenarien „Waitaha“ und „Schulfest“.

3.7 DATENAUSWERTUNG

Für die Analyse der Lautdenkprotokolle und *Stimulated-Recall*-Daten können je nach Erkenntnisinteresse unterschiedliche Methoden eingesetzt werden (vgl. u.a. Aguado, 2013; Chi, 1997; Gass & Mackey, 2017; Alison Green, 1998; Y. Gu, 2014; Heine & Schramm, 2007). Alle Datenauswertungsverfahren, die für die Analyse von qualitativen Daten gewählt werden können, weisen Stärken und Schwächen auf, die es gemäss Aguado (2013, S. 120) jeweils im Hinblick auf ihre Gegenstandsangemessenheit zu reflektieren und abzuwägen gilt. Laut Y. Gu (2014) ist es beim Festlegen der Vorgehensweise für die Analyse von qualitativen Daten wichtig, das Gesamtdesign der Studie und die Art der Forschungsfragen im Auge zu behalten; auch sollen die Überlegungen zur Datenanalyse und Codierung des Materials mit der Datenerhebungsmethode und der erkenntnistheoretischen Perspektive der Studie übereinstimmen. In diesem Kapitel wird daher detailliert beleuchtet, welche Überlegungen zur Wahl der Auswertungsmethode in der vorliegenden Studie führten.

Im Fokus stehen Strategien, die von den Schüler/innen beim Lösen der szenariobasierten und nicht szenariobasierten Testaufgaben aus dem Projekt IFB eingesetzt wurden. Das Forschungsdesign wurde in Anlehnung an ähnliche Studien aus dem Bereich der Testforschung konzipiert (A. D. Cohen & Upton, 2006, 2007; Rossa, 2012). Es war daher naheliegend, auch bei der Datenauswertung kategorienbasiert in Anlehnung an diese und weitere ähnliche Studien vorzugehen, die in der Zwischenzeit erschienen waren (L. Gu & So, 2017; Suvorov, 2018). Ausserdem fielen die erhobenen Daten sehr umfangreich aus, was eine systematische und prinzipienbasierte Vorgehensweise bei ihrer Auswertung ohnehin zwingend machte. Da die Untersuchung einen exploratorischen Charakter hatte, wurde es hier weiter als wichtig erachtet, die Strategien der Schüler/innen nicht nur zu codieren und diese Codierungen evtl. zu quantifizieren, sondern ihnen bei der Analyse der Resultate und im Hinblick auf die Beantwortung der Forschungsfragen mehr Aufmerksamkeit zu schenken.

Als geeignetes Analyseverfahren, das die oben genannten Punkte zu erfüllen versprach, bot sich die qualitative Inhaltsanalyse an. Diese Form der Inhaltsanalyse bezweckt eine „systematische Reduktion von Komplexität“ (Aguado, 2013, S. 126), geht daher methodisch und kategorienbasiert vor (vgl. Kuckartz, 2018, S. 6) und „will sich nicht auf die oberflächlichen manifesten Textinhalte bei der Analyse beschränken, sondern bezieht auch die Elemente der Interpretation mit ein“ (Burwitz-Melzer & Steininger, 2016, S. 258; mit Verweis auf Kuckartz, 2012, S. 28). Bei der qualitativen Inhaltsanalyse wird also einem Phänomen, das in den Daten erkannt und regelgeleitet als kategorial relevant eingestuft wurde, eine Kategorie zugewiesen, was wiederum als Akt der Interpretation der Forschenden zu verstehen ist, weil diese Zuordnung eines Phänomens zu einer Kategorie aufgrund der Einschätzung des Datenmaterials durch die Forschenden geschieht (Burwitz-Melzer & Steininger, 2016, S. 264).

Die qualitative Inhaltsanalyse wurde im deutschsprachigen Raum in den 1980er Jahren von Mayring bekannt(er) gemacht (1983; seit der Erstpublikation erschienen zahlreiche Neuauflagen seines Buches) und von Kuckartz (1999, 2012, 2018) weiterentwickelt (vgl. auch Aguado, 2013; Burwitz-Melzer & Steininger, 2016). Für die Analyse der vorliegenden Verbalprotokolle wurden die inhaltlich strukturierende und die darauf aufbauende evaluative Variante der qualitativen Inhaltsanalyse nach Kuckartz (2018) gewählt. In Kap. 3.7.1 wird das gewählte Verfahren näher vorgestellt und die Wahl dieser Analysemethode weiter begründet.

Gemäss Galaczi (2014) soll jede inhaltsanalytische Studie auf methodischer Rigorosität (*methodological rigor*) basieren. Diese kann bei der Inhaltsanalyse erreicht werden „through valid, reliable, and dependable findings which are accurate and replicable; that is, researchers working at different points in time, using the same technique on the same data set, should get the same results“ (ebd., S. 1333). Die Reliabilität wird in der Inhaltsanalyse durch die

Codierenden, das Codiersystem und seine Anwendung auf die Daten erzielt. Die Resultate einer Inhaltsanalyse können dann als valide betrachtet werden, wenn in der Studie tatsächlich das Konzept bzw. Konstrukt untersucht wurde, das untersucht werden sollte. Dies sei wiederum abhängig von den Forschungsfragen, den gewählten Datenerhebungsverfahren, dem Codiersystem (d.h. der Sorgfalt bei der Entwicklung des Kategorienleitfadens sowie seiner transparenten und verlässlichen Anwendung beim Codieren), wie auch von der Angemessenheit der Schlussfolgerungen, die aufgrund der Daten gezogen werden, so Galaczi (ebd.). Sie plädiert daher für eine systematische und transparente Herangehensweise und Durchführung. In den folgenden Unterkapiteln wird aufgezeigt, was in der vorliegenden Untersuchung unternommen wurde, um den oben geschilderten Ansprüchen an die methodische Rigorosität einer inhaltsanalytischen Studie gerecht zu werden. Die Vorgehensweise bei der Entwicklung des Kategoriensystems, das bei der Codierung der Verbalprotokolle zum Einsatz kam, wird in Kap. 3.7.2 beschrieben. In Kap. 3.7.3 wird das Kategoriensystem vorgestellt. Zuletzt wird in Kap. 3.7.4 die Vorgehensweise bei der Datenauswertung kritisch reflektiert, indem zusammenfassend aufgezeigt wird, was in der Arbeit unternommen wurde, um die Gütekriterien der qualitativen Inhaltsanalyse einzuhalten. Dies wird in Anlehnung an die Checkliste für die interne Studiengüte gemacht, die von Kuckartz für die qualitative Inhaltsanalyse (2018, Kap. 9) entwickelt wurde und in der sich auch die von Galaczi (2014) genannten Punkte wiedererkennen lassen. Den Resultaten ist Kap. 4 dieser Arbeit gewidmet.

3.7.1 DIE INHALTLICH STRUKTURIERENDE UND EVALUATIVE QUALITATIVE INHALTSANALYSE NACH KUCKARTZ (2018) UND IHR EINSATZ IN DER VORLIEGENDEN STUDIE

In seinem Lehrbuch zur qualitativen Inhaltsanalyse knüpfte Kuckartz (2018) an den Ansatz von Mayring (u.a. 2015) an, sowie an verschiedene Arbeiten aus dem Bereich der klassischen Inhaltsanalyse und praktische Forschungsarbeiten, in denen qualitative Daten systematisch und kategorienbasiert analysiert wurden. Auf dieser Grundlage entwickelte er die qualitative Inhaltsanalyse weiter. Während es bei Mayring noch primär um das Bilden von Kategorien und das Auszählen ihrer Häufigkeiten ging, interessiert sich Kuckartz mehr für die Analyse *nach* der Codierphase, und zwar aus einer Position, die „stärker qualitativ und hermeneutisch akzentuiert ist“ und die auch „für die Berücksichtigung einer fallorientierten Perspektive plädiert“⁵² (2018, S. 6). Dies begründet Kuckartz damit, dass die Kategorienbildung und Codierung zwar selbstverständlich sehr wichtige Forschungsetappen darstellen, die Analyse

⁵² Dies stünde im Unterschied zum Ansatz von Mayring (2015), wo die kategorienorientierte und nicht die fallorientierte Perspektive eine Rolle spiele (vgl. Kuckartz, 2018, S. 49).

der codierten Daten aber erst danach ganz interessant werde und auch mehr als eine einfache, quantitative Auswertung der Häufigkeiten sein könne (ebd.). Die qualitative Inhaltsanalyse wird demnach in ihrem Ansatz als eine interpretative Form der Auswertung verstanden, in der „Codierungen aufgrund von Interpretation, Klassifikation und Bewertung vorgenommen [werden]; die Textauswertung ist hier also an eine menschliche Verstehens- und Interpretationsleistung geknüpft“ (ebd., S. 27). Die qualitative Inhaltsanalyse ist darüber hinaus nicht primär an der Quantifizierung der Codierungen interessiert, sondern am Text selbst, denn auch nach der Codierung bleiben die Inhalte der codierten Aussagen wichtig und der Wortlaut dieser Aussagen ist auch bei der Aufbereitung und Präsentation der Forschungsergebnisse von grosser Bedeutung (ebd., S. 48).

Kuckartz (ebd., Kap. 5-7) unterscheidet drei Basismethoden der qualitativen Inhaltsanalyse: 1) die inhaltlich strukturierende Inhaltsanalyse, bei der die Themen und Subthemen mittels Kategorien und Subkategorien identifiziert, systematisiert und analysiert werden; 2) die evaluative Inhaltsanalyse, die noch stärker hermeneutisch-interpretativ ausgerichtet ist und bei der die Inhalte von den Forschenden eingeschätzt, klassifiziert und bewertet werden⁵³; und 3) die typenbildende qualitative Inhaltsanalyse, die (häufig basierend auf der Vorarbeit einer inhaltlich strukturierenden oder evaluativen Codierung) nach mehrdimensionalen Mustern sucht, um in einer methodisch kontrollierten Form Typen zu bilden und eine Typologie zu entwickeln.

Für die Beantwortung der Forschungsfragen, die im Zentrum der vorliegenden Studie stehen, wurden introspektive Daten erhoben. Die Analyse dieser Art von Daten erfordert eine hohe Interpretationsleistung seitens der Codierenden, denn Lautdenkprotokolle und Protokolle der *Stimulated-Recall*-Interviews können nicht einfach wortbezogen bzw. automatisch ausgewertet und müssen daher von den Codierenden stets reflektiert werden. Auch müssen für die Codierung dieser Daten, wie auch sonst bei allen anderen Arten der Inhaltsanalyse, klare Regeln der systematischen Auswertung von Daten festgelegt und präzise, erschöpfende, schlüssige und trennscharfe Kategorien gebildet werden. Ausserdem sind in der vorliegenden Arbeit nicht nur die Analysen der Häufigkeiten der codierten Stellen von Bedeutung, sondern auch die konkreten Aussagen der Schüler/innen zu den untersuchten Phänomenen. Weil eine solche Herangehensweise an die Auswertung von qualitativen Daten

⁵³ Das qualitative Material wird bei der evaluativen Inhaltsanalyse mithilfe von Skalen eingeschätzt, die meist ordinaler Art sind, z.B. „hohe Motivation“, „mittelmässige Motivation“ und „niedrige Motivation“. Die evaluative Inhaltsanalyse bezieht sich dabei nicht auf die Bewertungen der einzelnen Textstellen, sondern des gesamten Falls. Es ist auch möglich, die evaluative Inhaltsanalyse – wie in der vorliegenden Studie – mit der inhaltlich strukturierenden Analyse zu kombinieren, indem bei der evaluativen Inhaltsanalyse auf der Vorarbeit aufgebaut wird, die bei der inhaltlich strukturierenden Codierung geleistet wurde (Kuckartz, 2018, Kap. 6).

im Vordergrund der von Kuckartz (2018) vorgeschlagenen Art der qualitativen Inhaltsanalyse steht, wurde entschieden, die erhobenen Daten nach den von ihm beschriebenen Prinzipien auszuwerten. Für die Codierung der Daten wurde die Basismethode der inhaltlich strukturierenden Inhaltsanalyse gewählt (vgl. ebd., Kap. 5). Aufbauend auf den Resultaten der inhaltlich strukturierenden Inhaltsanalyse wurden bei bestimmten Kategorien die vorgenommenen Codierungen zusätzlich nach den Prinzipien der evaluativen Inhaltsanalyse ausgewertet (vgl. ebd., Kap. 6). Dies wurde u.a. im Hinblick auf die Beantwortung der zweiten Forschungsfrage gemacht und wird in Kap. 3.7.3.2 detaillierter dargestellt. Auf die dritte, typenbildende Variante der qualitativen Inhaltsanalyse wurde hier verzichtet, weil die Typenbildung für die Beantwortung der Forschungsfragen nicht nötig war und auch nicht angestrebt wurde.

Im nächsten Kapitel (3.7.2) wird beschrieben, wie die Datenanalyse konkret vorgenommen und wie die dafür nötigen Instrumente entwickelt wurden. Die Begriffe „Kategorie“ und „Code“ werden dabei im Folgenden nach Kuckartz (2018, S. 36) synonym gebraucht. Auch weitere Begriffe zur qualitativen Inhaltsanalyse (Kategorienleitfaden, Kategorienhandbuch etc.) werden in dieser Arbeit in Anlehnung an Kuckartz (ebd.) verwendet.

3.7.2 VORGEHENSWEISE BEI DER DATENAUSWERTUNG

Die Auswertung der Verbalprotokolle der 20 Schüler/innen wurde computerunterstützt mithilfe des Analyseprogramms MAXQDA 2020 (VERBI Software, 2019) durchgeführt. Leitend waren dabei zunächst die im vorhergehenden Kapitel besprochenen Prinzipien der inhaltlich strukturierenden Inhaltsanalyse nach Kuckartz (2018): Mithilfe von Kategorien und Subkategorien wurde eine inhaltliche Strukturierung der erhobenen Daten vorgenommen (vgl. ebd., S. 101).

In der ersten Phase der Datenauswertung, die Kuckartz (2018) als „initiiierende Textarbeit“ bezeichnet, wurden die Transkripte der 20 ausgewählten Schüler/innen sorgfältig gelesen. Es wurde dabei versucht, sich im Hinblick auf die Forschungsfragen einen ersten Überblick über das jeweilige Transkript zu verschaffen. Bei dieser Gelegenheit wurden auch Auffälligkeiten in den Texten markiert und Ideen zur Auswertung festgehalten. Falls in einem Transkript Stellen entdeckt wurden, die schwer nachvollziehbar waren, wurde auf das Video zurückgegriffen, um die jeweiligen Passagen im Transkript um weitere Informationen zu ergänzen (z.B. Gestik oder Mimik der Lernenden). Es konnte dabei beobachtet werden, dass gewisse Lernende trotz der Hinweise in der Instruktion, beim Lauten Denken mit sich selbst zu sprechen, die Gedanken frei laufen zu lassen und nicht zu versuchen, jemandem etwas zu erklären, ihre Äusserungen eher an die Forscherin als an sich selbst gerichtet hatten. Weil

durch diese Äusserungen weitere spannende Einblicke in die Gedanken und Vorgehensweisen der Schüler/innen gewonnen werden konnten, die für die Beantwortung der Forschungsfragen ebenfalls von Bedeutung waren, wurde entschieden, auch diese metakognitiven Äusserungen bei der Datenauswertung mitzuberücksichtigen. Sie wurden während des Codierens nicht speziell gekennzeichnet, weil es im Hinblick auf die Beantwortung der Forschungsfragen als nicht nötig erachtet wurde.

Während dieser ersten Phase der Datenauswertung wurden in den Transkripten auch folgende allgemeine Informationen festgehalten und mit entsprechenden Codes versehen: 1) Identifikation des Untersuchungsteils (Lautes Denken oder *Stimulated Recall*) und des Aufgabenzusammenhangs (Szenario oder Off-Szenario), 2) Identifikation der aktuellen Bildschirmseite, auf die sich die Äusserungen im Transkript beziehen, 3) Bewertung der endgültigen Antwort, die von den Schüler/innen beim Lauten Denken angeklickt wurde, als richtig oder falsch. Die dafür erstellten Kategorien werden in Kap. 3.7.3.1 näher vorgestellt.

In der anschliessenden Phase wurde mit der Entwicklung der thematischen Kategorien für den ersten Codierprozess begonnen. Mithilfe von thematischen Kategorien wurden in den Daten Stellen erfasst, die sich auf ein für die Beantwortung der Forschungsfrage relevantes Thema bezogen (Kuckartz, 2018, S. 34). Die thematischen Kategorien spielen in der vorliegenden Studie eine strukturierende und systematisierende Rolle. Die verbalen Daten, die mithilfe der thematischen Kategorien codiert wurden, bleiben also hier, wie bereits erwähnt, im Gegensatz zur klassischen Inhaltsanalyse auch nach der Codierung von Bedeutung (vgl. ebd., S. 47), was in Kap. 4 deutlich wird.

Die Festlegung der möglichen thematischen Kategorien kann aufgrund der Daten, der Erkenntnisse aus ähnlichen Studien oder auch aufgrund von Theorien erfolgen (Galaczi, 2014; mit Verweis auf Zhang & Wildemuth, 2009). Die in dieser Untersuchung eingesetzten Kategorien wurden allen drei von Galaczi (ebd.) genannten Quellen entnommen, allerdings in chronologisch umgekehrter Reihenfolge. So wurden hier die thematischen Kategorien zuerst deduktiv aufgrund der Erkenntnisse aus der Fachliteratur entwickelt. Dafür wurden die theoretischen Arbeiten von A. D. Cohen beigezogen, der für die Unterteilung der Teststrategien (*test taking strategies*) einen einleuchtenden und in der Testforschung anerkannten theoretischen Rahmen vorschlägt (A. D. Cohen, 1997, 2007b, 2011b, 2012, 2014). Weiter wurden auch Studien zu Teststrategien konsultiert, die sich entweder an der Strategienklassifizierung von A. D. Cohen orientierten (A. D. Cohen & Upton, 2006, 2007; L. Gu & So, 2017; Suvorov, 2018) oder einen anderen Ansatz wählten (Nikolov, 2006; Rossa, 2012) (vgl. Kap. 2.2.4.2). Darauf basierend entstand eine erste Zusammenstellung der

möglichen Kategorien, die bei der Erprobung und Analyse des Materials sukzessiv überarbeitet und um weitere, am Material entstandene Kategorien ergänzt wurde.

Die drei strategienbezogenen Hauptkategorien bildeten in diesem Prozess die von A. D. Cohen vorgeschlagenen Lernerstrategien, Testmanagementstrategien und *test-wiseness*-Strategien. Diesen drei Hauptkategorien wurden nach und nach passende Subkategorien, also einzelne Lerner-, Testmanagement- und *test-wiseness*-Strategien, zugeordnet. Es handelte sich dabei um Strategien, deren Einsatz in den oben erwähnten Studien festgestellt wurde und/oder die in den vorliegenden Verbalprotokollen zu finden waren. Die Zuordnung der Subkategorien zu den Hauptkategorien wurde jeweils früh im Prozess der Codierung und nicht erst in einem weiteren Schritt vorgenommen, d.h. beim Festlegen einer neuen Kategorie wurde von Anfang an entschieden, ob es sich bei dieser Kategorie um eine (in Bezug auf das IFB-Konstrukt) konstruktrelevante Lernerstrategie, konstruktrelevante Testmanagementstrategie oder konstruktirrelevante *test-wiseness*-Strategie handelt. So wurde z.B. die Strategie „Kenntnisse über das Itemformat nutzen“ als konstruktrelevante Testmanagementstrategie definiert, weil in der Aufgabe, bei der diese Strategie vorkam (Szenario „Stadttrip“, Task A), ein Notizzettel mit vier Lücken mit zwei empfohlenen und zwei nicht empfohlenen Sachen vervollständigt werden sollte. Es war hier also wahrscheinlich dem Format geschuldet, dass manche Schüler/innen den Inputtext zu lesen aufhörten, nachdem sie zwei positive und zwei negative Punkte gefunden hatten: Auch im realen Leben würde man u.U. so vorgehen. Die Strategie „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“ wurde hingegen als eine konstruktirrelevante *test-wiseness*-Strategie betrachtet, weil die Schüler/innen, die sie einsetzten, bei der Auswahl der Antwortoption(en) die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigten bzw. verstanden und folglich nicht die kognitiven Prozesse durchliefen, die durch das Item vorgesehen waren. Beim Codieren der Daten wurden die entsprechenden Transkriptstellen, in denen „a specific and recognizable strategic choice made by the subject that is deliberate and purposeful and is intended to facilitate the reading or test-taking task“ (A. D. Cohen & Upton, 2006, S. 39) zu erkennen war, den jeweiligen Lerner-, Testmanagement- oder *test-wiseness*-Kategorien zugeordnet.

Die deduktiv und induktiv zusammengestellten Kategorien zu den Teststrategien wurden in einem weiteren Schritt um zusätzliche thematische Kategorien ergänzt, die während der initiierenden Textarbeit am Material (d.h. induktiv) entstanden und für die Beantwortung der Forschungsfragen oder auch die Reflexion der eingesetzten Forschungsmethoden wichtig waren. So wurden zu dieser Kategoriensammlung u.a. Kategorien zur Wahrnehmung des

Szenarios beim Test oder zum Einfluss des Lauten Denkens auf den Testbearbeitungsprozess der Schüler/innen hinzugefügt.

Für alle thematischen Kategorien wurden anschliessend Kategoriendefinitionen entworfen, die im Verlauf der Codierung allmählich präzisiert wurden. Auf der Grundlage der Kategoriendefinitionen entstand ein Kategorienhandbuch⁵⁴ mit den Kategorien und ihren Beschreibungen, das in einem nächsten Schritt mit Anweisungen und Hilfen zum Codierprozess versehen wurde. Der so entstandene Kategorienleitfaden wurde zuerst an zwei Transkripten erprobt und um weitere Anweisungen und Bemerkungen zum Codierprozess ergänzt. Bereits während dieser ersten Erprobung zeigte sich, dass der erste, noch sehr umfangreiche Entwurf des Kategorienleitfadens reduziert werden musste, weil gewisse darin enthaltene Kategorien nicht trennscharf genug oder auch zu feinkörnig waren und zusammengeführt werden mussten. Im gleichen Zuge wurden gewisse Kategoriendefinitionen präzisiert. So wurden beispielsweise nach der ersten Erprobung drei ursprüngliche Testmanagement-Kategorien „Antwortoption lesen und Anhaltspunkte für die Antwort im Inputtext suchen“, „Antwort im Inputtext lokalisieren“ und „Fokus auf relevante Textteile“ zu einer neu etwas breiter definierten Kategorie „Fokus auf relevante Textteile“ zusammengeführt. Die Erprobung zeigte nämlich, dass diese drei Strategien in den Daten kaum voneinander zu unterscheiden waren und immer zusammen codiert werden mussten.

Der überarbeitete Kategorienleitfaden wurde anschliessend an vier Transkripten erprobt, darunter an den zwei Transkripten, die bereits bei der ersten Erprobung gesichtet und nun mit den teilweise neuen Codes versehen wurden. Bei der zweiten Erprobung wurde auch der Kategorienleitfaden weiter gestrafft und präzisiert. Dem Kategorienleitfaden wurden auch Ankerbeispiele für die Codes angefügt, sofern diese in den vier analysierten Transkripten bereits gefunden werden konnten.

Um Hinweise zur intersubjektiven Nachvollziehbarkeit und zur Kohärenz des Kategorienleitfadens zu erhalten, wurde eine Datensitzung mit zwei Forschenden⁵⁵ durchgeführt. Es zeigte sich dabei, dass der Kategorienleitfaden und die damit vorgenommenen Codierungen insgesamt gut nachvollzogen und keine offensichtlichen Lücken oder Überschneidungen identifiziert werden konnten. Die Rückmeldungen zum Kategoriensystem verdeutlichten aber auch, dass gewisse Kategorien noch passender

⁵⁴ Der Kategorienleitfaden (auch „Codebuch“ oder „Codebook“ genannt) besteht aus dem Kategorienhandbuch (d.h. aus den Kategorien und ihren Beschreibungen) und aus den Anweisungen und Hilfestellungen für die Codierenden (Kuckartz, 2018, S. 39–40).

⁵⁵ Ich danke Thomas Studer und Katharina Karges für die spannende Diskussion und die wertvollen Rückmeldungen zum Kategoriensystem und zur Analyse der codierten Daten anlässlich dieser Datensitzung.

benannt werden mussten. So wurde beispielsweise die Strategie „Einsatz der Sprachkenntnisse...“ in „Einsatz von Verstandenenem...“ umbenannt. Andere Kategorien mussten wiederum präziser definiert werden, um sie von anderen, auf den ersten Blick ähnlichen Kategorien klarer abzugrenzen, z.B. die Testmanagement-Kategorie „Vorhersage machen“ vs. *test-wiseness*-Kategorie „Ausschluss oder Auswahl von Antwortoption(en) aufgrund des Hintergrundwissens“. Auch zeigte sich, dass gewisse Beispiele für die Codes ersetzt werden mussten, da sie nicht prototypisch genug waren: Es erwies sich als schwierig, die ursprünglichen Beispiele, die nur aus vier Transkripten stammten, ohne ihren breiteren Kontext nachzuvollziehen.

Mithilfe des überarbeiteten Codierleitfadens wurden von der Autorin der vorliegenden Studie die Transkriptteile der 20 Schüler/innen, die sich auf die zu analysierenden Items bezogen (vgl. Kap. 3.6), vollständig codiert. Aus Ressourcengründen konnten keine weiteren Codierenden beigezogen werden. Beim Codierprozess wurden zunächst die Transkripte aufmerksam gelesen und die Codiereinheiten⁵⁶ identifiziert. Wurde im Transkript eine relevante Textstelle gefunden, wurde sie mit dem einschlägigen Code verbunden. Dies geschah in Anlehnung an Kuckartz, der in seinem Methodenbuch die Konzeption vertritt, „dass im Rahmen der qualitativen Inhaltsanalyse in der Regel Sinneinheiten als Codiereinheiten gewählt werden und sich die codierten Segmente durchaus überlappen und ineinander verschachtelt sein können“ (2018, S. 43).

Die Grösse der codierten Textpassage variierte zwischen einigen Wörtern und mehreren Segmenten/Propositionen bzw. Zeilen (vgl. Kap. 3.5). Die Textpassage wurde jeweils so gewählt, dass sie auch ausserhalb des umgebenden Textes noch verständlich sein sollte. Passagen, die für die Beantwortung der Forschungsfragen nicht relevant waren (z.B. die für das Projekt IFB interessanten Diskussionen über die Schwierigkeit der Items, das Design oder die Usability der computerbasierten Tasks), wurden in den Transkripten nicht codiert. Besonders interessante Stellen, die im Rahmen von weiterführenden Untersuchungen analysiert werden könnten, wurden mit dem speziell dafür vorgesehenen Code „Sonstiges“ erfasst. Auf die mit dieser Kategorie erfassten Datenauszüge kann in dieser Arbeit jedoch nur teilweise eingegangen werden (vgl. Kap. 4.5).

Auch wenn die Strategien der Schüler/innen nur in den Transkriptteilen zu den ausgewählten Bildschirmseiten vertieft untersucht wurden (vgl. Kap. 3.6), wurden beim Codieren jeweils die vollständigen Transkripte dieser 20 Lernenden sorgfältig durchgelesen. Dadurch sollten in

⁵⁶ Unter Codiereinheit wird in der qualitativen Inhaltsanalyse „eine Textstelle verstanden, die mit einer bestimmten Kategorie, einem bestimmten Inhalt, z.B. einem Thema oder einem Unterthema, in Verbindung steht“ (Kuckartz, 2018, S. 41).

den Transkriptteilen zu den nicht untersuchten Testaufgaben eventuell wichtige Informationen zu den aufgabenübergreifenden Strategien der Schüler/innen oder zu ihrer Wahrnehmung der Untersuchungssituation identifiziert und bei der Analyse berücksichtigt werden (können). Dort, wo in den Transkripten zu den restlichen Bildschirmseiten Stellen gefunden wurden, die sich nicht nur auf die besprochene Bildschirmseite sondern auf allgemeine Verhaltensweisen der/des Lernenden beim Test bezogen, wurden diese mit einem passenden Code und einer speziellen Gewichtung versehen (vgl. Punkt 6 im folgenden Abschnitt)⁵⁷. So wurde beispielsweise im Transkript der Schülerin Elena folgende Aussage mit der Kategorie „Kognaten identifizieren und nutzen“ codiert, die sie beim Bearbeiten der dritten, aus der vertiefenden Analyse ausgeschlossenen Bildschirmseite zu Task A bei Szenario „Stadttrip“ machte:

*manchmal suche ich auch so parallelwörter (--) zum schauen (.) was es sein könnte;*⁵⁸ (Elena_Szenario, Pos. 36-37⁵⁹)

Der Codierprozess erwies sich als intensiv und anspruchsvoll, u.a. wegen des Umfangs der Transkripte. Auch waren die Lautdenkprotokolle wegen ihres fragmentarischen Charakters teilweise schwer zu interpretieren. Bei der Codierung mussten daher stets die jeweiligen Aufgaben und Inputtexte und an manchen Stellen zusätzlich auch die Videoaufnahmen konsultiert werden, um zu verstehen, worauf sich die Äusserungen der Schüler/innen bezogen. Für die Codierung der thematischen Kategorien wurden in jedem der 20 Transkripte im Sinne der *Intra-Rater*-Reliabilität (vgl. Knorr & Schramm, 2016, S. 93) mindestens sechs systematische Codierdurchgänge vorgenommen. Wurde bei der Überprüfung der Codierungen bei einem beliebigen Codierdurchgang eine Inkonsistenz in der Anwendung des Kategorienleitfadens festgestellt, wurde die Codierung der betroffenen Stelle angepasst. Auch wurden dabei, wo nötig, die Kategoriendefinitionen weiter präzisiert und zu gewissen Transkriptstellen ergänzende Informationen in Form von Memos festgehalten. Dies geschah z.B. an den Stellen in den Lautdenkprotokollen, die nicht eindeutig waren und für deren Interpretation eine entsprechende Passage aus dem Interview beigezogen werden musste (vgl. untenstehende Ausführungen zum zweiten Codierdurchgang). Wenn in den Transkripten der *Stimulated-Recall*-Interviews entdeckt wurde, dass nicht genügend Evidenz für das Vorkommen einer Strategie vorhanden ist oder die Forscherin eine Frage zu suggestiv stellte,

⁵⁷ Ein Überblick über alle definierten Gewichtungen ist ebenfalls im Anhang 5 zu finden.

⁵⁸ Zitateile, die zwischen Asterisken stehen, bedeuten, dass der/die Schüler/in diese Äusserung in seinem/ihrer Dialekt machten.

⁵⁹ Damit ist die Position im Transkript gemeint, gemäss der automatischen Nummerierung in MAXQDA (VERBI Software, 2019).

wurden die Antworten der Schüler/innen uncodiert gelassen. So wurde z.B. in der untenstehenden Aussage von Cersei die Strategie „Raten“ nicht codiert:

MB⁶⁰: gut (-) hm hm (-) warst du dir da SICHer bei äh (-) dieser option [oder hast du eher geraten;]

S⁶¹: [nein eben NICHT] weil ich (-) weil eben ALLes darin VORgekommen ist war ich (-) NICHT sicher; (Cersei_Szenario, Pos. 840-843)

Auch wenn die Schülerin bestätigte, dass sie sich unsicher war, ob ihre Antwort stimmte, kann hier nicht mit Sicherheit gesagt werden, dass sie ihre Antwort tatsächlich erraten musste.

Der Fokus der jeweiligen Codierdurchgänge bei der Hauptcodierung der 20 Transkripte wird im Folgenden näher erläutert:

1. Beim ersten Codierdurchgang wurden die Transkripte intensiv gelesen und mit den strategienbezogenen Kategorien (z.B. „Sorgfältiges Lesen“ oder „Kognaten identifizieren und nutzen“, vgl. Kap. 3.7.3.2) und mit weiteren thematischen Kategorien (z.B. „Szenario nach einem Hinweis wahrgenommen“ oder „Lautes Denken hat (eher) gestört“, vgl. Kap. 3.7.3.3) codiert. Dabei wurden auch unklare Stellen in den Transkripten markiert, die beim zweiten Durchgang intensiver angeschaut werden sollten.
2. Der zweite Durchgang diente der nochmaligen Überprüfung der vorgenommenen Codierungen. Dabei wurde vor allem auf die Übereinstimmung der Codierungen in den Transkriptteilen zum Lauten Denken und zum *Stimulated-Recall*-Interview geachtet.

Der Vergleich der beiden Untersuchungsteile zur gleichen Bildschirmseite erwies sich besonders dann als nützlich, wenn im Transkriptteil zum *Stimulated-Recall*-Interview festgestellt wurde, dass die Aussage der/des Jugendlichen nicht mit dem übereinstimmt, was in den Lautdenkprotokollen zur entsprechenden Stelle vorlag. Der Grund dafür war wohl, dass sich die Lernenden beim *Stimulated-Recall*-Interview nicht mehr richtig an die Testsituation erinnern konnten. In solchen Fällen wurde nach Konsultation der Lautdenkprotokolle die Codierung im Interviewteil entfernt, die beim ersten Durchgang lediglich aufgrund der Interviewaussage vorgenommen worden war. So behauptete beispielsweise Szenario-Schüler Böbu im Interview, er habe bei Task A, Seite 2 die Zuordnung der Antwortoption „Eine Bootsfahrt machen“ erraten. Diese Stelle wurde dann auch beim ersten Durchgang mit der entsprechenden Kategorie „Raten“ codiert.

⁶⁰ Interviewerin (Malgorzata Barras).

⁶¹ Schüler/in.

```
*eine bootsfahrt machen(-) <<verzieht das gesicht> nä::> (--)  
bin ich mir gerade gar nicht sicher; da habe ich glaube auch  
geraten bei dem;* (Boebu_Szenario, Pos. 813-815)
```

Der Vergleich dieser Äusserung mit der entsprechenden Transkriptstelle im Lautdenkprotokoll beim zweiten Codierdurchgang zeigte aber, dass der Schüler hier doch nicht auf das Raten zurückgegriffen, sondern seine Antwort aufgrund dessen ausgewählt hatte, was er vom Text verstanden hatte. Beim Bearbeiten dieser Aufgabe sagte er nämlich nach der Lektüre:

```
*etwas botsfahrt haben sie noch °h geschrieben-°h ((liest den  
text weiter)) (2.0) "ennuyeux"_ehm (---) ((schaut nach oben und  
dann auf die antwortoptionen)) ((zieht 4 nach A2w)) also sie  
empfehlen eine bo:tsfa:hrt zu machen`* (4.5) (Boebu_Szenario,  
Pos. 88-91)
```

Die Codierung der Strategie „Raten“ wurde folglich beim zweiten Codierdurchgang im *Stimulated-Recall*-Interview von Böbu entfernt und die Begründung dafür wurde in einem Memo festgehalten.

Ausserdem gab es im Transkriptteil zum Lauten Denken an manchen Stellen zwar Indizien für den Einsatz einer bestimmten Strategie, die aber alleine nicht stark oder nicht eindeutig genug waren. In diesen Fällen wurde im Transkript zum entsprechenden Interviewteil nach weiterer Evidenz gesucht. Falls die entsprechende Interviewpassage die Vermutung zum Einsatz einer Strategie beim Lauten Denken bestätigte, wurde die Stelle im Verbalprotokoll zum Lauten Denken nachcodiert. Die Erklärung dazu wurde dann auch in einem Memo festgehalten. Beim Schüler Ezra liess z.B. die unten zitierte Stelle aus dem Lautdenkprotokoll vermuten, dass er beim Ausschluss der Antwortoption „Eine Tabelle mit Vor- und Nachteilen zusammenstellen“ eine *test-wiseness*-Strategie einsetzte, für die die Kategorie „Ausschluss oder Auswahl einer Antwortoption bzw. der Antwortoptionen mit nicht rezipierten oder nicht verstandenen Wörtern“ vorgesehen war. Die Indizien für den Einsatz dieser Strategie wurden im untenstehenden Zitat fett markiert:

```
((schaut auf die aufgabenstellung)) und eine tabelle mit VOR und  
NACHteilen ((schaut zum text)) die habe ich jetzt ((cursor auf  
f)) (---) (m:) ((liest den text)) (4.2) nEIn; also ich denke EHer  
((schaut auf die aufgabenstellung)) einen text über die erFINDUNG  
(-) der wAlkies schreiben` (Ezra_Off, Pos. 62-65)
```

Der spätere Abgleich dieses Lautdenkprotokollausschnitts mit der entsprechenden Stelle im Interview zeigte, dass diese Vermutung richtig war und der Schüler beim Lösen dieser Testaufgabe tatsächlich auf die oben genannte Strategie zurückgegriffen hatte.

```
U:nd eine (tabelle) mit vor und nAchteilen erstellen habe ich
```

(--) Auch nicht richtig eh (---) (...) **das habe ich auch nirgends gefunden**; also einfach so: was pOsitiv und negativ sein könnte; **darum habe ich gedacht es muss das unterste sein**. (Ezra_Off, Pos. 482-488)

Die oben aufgeführte Stelle wurde daher im Lautdenkprotokoll von Ezra beim zweiten Codierdurchgang mit der erwähnten Kategorie nachcodiert und mit einem entsprechenden Memo versehen.

Konnte eine Stelle auch bei der zweiten Überprüfung des Transkripts und unter Zuhilfenahme des Videos nicht eindeutig interpretiert werden, wurde sie uncodiert gelassen. Dies geschah beispielsweise bei den Transkriptteilen zum Lauten Denken, in denen nicht genug Evidenz für den Einsatz bestimmter Strategien gefunden werden konnte. Auch wenn sich diese Strategien später aufgrund der Interviewaussagen der Schüler/innen rekonstruieren liessen, wurden sie nur im Interviewteil codiert und nicht in den Transkriptteilen zum Lauten Denken nachcodiert. Dies war u.a. beim Schüler Mogli der Fall, der beim Lösen der Aufgabe auf der zweiten Bildschirmseite zu Task B, Szenario „Stadttrip“, Folgendes verbalisierte:

°hh (3.2) *das gUte kann es NICht sein weil sie hat etwas SCHLECHtes gegessen- (---) °h danach ist sie krAnk gewesen- ((cursor auf Blb und dann auf Blc)) (2.0) dieser TAG ist (dann schön geworden) (.) also; hm_hm ((S klickt nach kurzem zögern Blc an)) (5.5) ((S klickt zweimal auf "weiter")) (6.4) (Mogli_Off, Pos. 76-81)

Aufgrund des Gesagten und des Videos konnte nicht ermittelt werden, *warum* der Schüler letztendlich die Antwortoption „Die schöne Aussicht vom Schloss“ und nicht etwa „Der starke Regen in der Stadt“ wählte, nachdem er die Option „Das gute Essen am zweiten Tag“ ausgeschlossen hatte. Daher wurden im Lautdenkprotokoll ausser der Strategie „Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“ keine weiteren Strategien codiert, auch wenn der Schüler im Interview ausführlicher Auskunft über seine Vorgehensweise und somit über die eingesetzten Strategien beim Bearbeiten dieser Aufgabe gab (was dort auch entsprechend codiert wurde):

das ESSEn am zweiten tag war eben schLECHt- also konnte es nicht gut sein; sie ist ja danach krANK geworden weil sie irgendetwas gegessen hat; (dann) habe ich gedacht DAS ist es SICHer nicht- und [...] dann habe ich gedacht also der REGen ist also nicht so faszinierend in einer stadt- sO (--) einfach mal so überLEGT- ich habe gedAcht ja die AUSSicht von einem schloss ist sicher noch (-) spAnnend. (Mogli_Off, Pos. 794-801)

3. Der dritte Codierdurchgang wurde durchgeführt, nachdem alle Transkripte zuerst der Szenario-Schüler/innen und dann der Off-Szenario-Schüler/innen codiert worden waren. Dies diente der Überprüfung, ob die Codierungen innerhalb der jeweiligen

Gruppen auch kongruent gehandhabt wurden, d.h. ob beispielsweise die gleichen Strategien bei den verschiedenen Schüler/innen auch jeweils mit den gleichen Codes erfasst wurden.

4. Beim vierten Codierdurchgang wurden die Codierungen aller 20 Transkripte nochmals auf ihre Übereinstimmung hin geprüft, diesmal auch gruppenübergreifend (Szenario- vs. Off-Szenario-Schüler/innen).
5. Danach wurden in MAXQDA (VERBI Software, 2019) systematisch alle codierten Textstellen zu allen Kategorien abgerufen und nacheinander durchgesehen. Dadurch sollte sichergestellt werden, dass die Codierungen innerhalb einer Kategorie ebenfalls überall gleich gehandhabt wurden.
6. In einem letzten, sechsten Schritt wurden nochmals alle codierten Textstellen zu jeder Kategorie durchgelesen und es wurde mithilfe der MAXQDA-Funktion „Gewichtung“ (vgl. VERBI Software, 2019) eine weitere Differenzierung der Kategorien zu Strategien vorgenommen, bei der nach den Prinzipien der evaluativen qualitativen Inhaltsanalyse vorgegangen wurde. Dies geschah u.a. im Hinblick auf die Quantifizierung der codierten Segmente für die Darstellung der Resultate. Die Gewichtungsfunktion mit ihrer Gewichtungsskala von 0 bis 100 erlaubt u.a. eine bessere Erfassung von Nuancen innerhalb einer Kategorie, wobei in MAXQDA „0“ als standardmässige Gewichtung eingestellt ist und zwischen 0 und 100 beliebige Zahlen für beliebige ergänzende Informationen vorgesehen werden können (Kuckartz & Rädiker, 2019, S. 79). Im vorliegenden Projekt wurde die Standardgewichtung „0“ überall dort beibehalten, wo sich eine bestimmte Äusserung auf das gerade besprochene Item bezog. Wenn der/die Schüler/in beim Besprechen einer Bildschirmseite eine wiederholte Aussage zu einer bestimmten Strategie, einer bestimmten Frage oder einer bestimmten Textstelle machte, die bei dieser Bildschirmseite oder bei diesem Item bereits besprochen und codiert worden war, wurde die erneut codierte Stelle jeweils mit „2“ gewichtet. So sagte z.B. Schülerin Loup bei der Besprechung eines bestimmten Items:

bei dem habe ich einfach geraten. (---) (Loup_Off, Pos. 731)

Da es sich dabei um erstmaliges Vorkommen dieser Strategie bei diesem Item handelte, wurde die Strategie mit der Standardgewichtung „0“ markiert. Im Verlauf des Gesprächs kam Loup auf das soeben erwähnte Item nochmals zu sprechen:

das habe ich eben NICHT verstanden dann habe ich geraten
(Loup_Off, Pos. 787)

Die zweite Erwähnung der Strategie „Raten“ zum gleichen Item wurde dann im Transkript mit Gewichtung „2“ gekennzeichnet.

Mit Gewichtung „5“ wurden weiter alle Codierungen zu Strategien versehen, bei denen aus dem Kontext hervorging, dass die Lernenden eine allgemeinere Aussage zu ihrem Verhalten beim Lösen der Testaufgaben machten. Diese Vorgehensweise war folglich item- oder textstellenübergreifend und betraf nicht nur das gerade besprochene Item oder die gerade besprochene Textstelle. Mit dieser Gewichtung wurde beispielsweise die untenstehende Aussage der Schülerin Picard gekennzeichnet, die als Strategie „Wortbedeutung identifizieren“ codiert wurde:

ein pAAr wörter habe ich nIcHT verstanden` dann habe ich halt
einfach (-) geschaut was: es hEIssen KÖNNTE vielleicht`mit dem
zusammenhang des textes- (Picard_Off, Pos. 236-238)

Um die Gewichtung „5“ vornehmen zu können, musste in manchen Fällen (wie im obigen Zitat) der breitere Kontext der codierten Aussage im Transkript konsultiert werden.

Bei gewissen Kategorien wurden mit der Gewichtungsfunktion nach den Prinzipien der evaluativen Inhaltsanalyse noch weitere, kategorienspezifische Differenzierungen vorgenommen. Diese werden in Kap. 3.7.3 im Detail vorgestellt.

In einem letzten Schritt wurden aus dem Kategorienleitfaden alle deduktiv erstellten Kategorien entfernt, für die in den vorliegenden Daten keine Evidenz gefunden werden konnte, und zwar weder in den Lautdenkprotokollen noch in den *Stimulated-Recall*-Interviews. Dies war z.B. bei der Kategorie „Selbstermutigung“ der Fall, die ursprünglich von Nikolov (2006, S. 16) übernommen wurde. Kategorien, für die es nur in einem der zwei Untersuchungsteile Evidenz gab, wurden hingegen im Kategorienleitfaden zu beiden Untersuchungsteilen beibehalten. Dies geschah u.a. im Hinblick auf die Diskussion der eingesetzten Datenerhebungsmethoden. So wurde z.B. die Lernerstrategie „Bisheriges Verständnis hinterfragen“ 4 Mal in Lautdenkprotokollen, aber in keinem *Stimulated-Recall*-Interview codiert. Sie bleibt zwar im Kategorienleitfaden zu beiden Untersuchungsteilen aufgelistet, im *Stimulated-Recall*-Teil des Kategorienleitfadens wird diese Strategie aber anstelle eines Beispiels mit der Bemerkung versehen, dass sie von den Schüler/innen während der Interviews nicht erwähnt wurde.

3.7.3 DAS KATEGORIENSYSTEM

Das im Hinblick auf die Beantwortung der Forschungsfragen entwickelte Kategoriensystem besteht aus drei Teilen. Der erste Teil beinhaltet Kategorien zur Erfassung der allgemeinen Informationen zu den beiden transkribierten Untersuchungsteilen (Lautes Denken oder *Stimulated Recall*, aktuell eingeblendete Bildschirmseite, Bewertung der endgültigen Antwort als „richtig“ oder „falsch“). Den zweiten Teil bilden thematische Kategorien zu den Strategien,

die in Anlehnung an A. D. Cohen (u.a. 2012, 2014) in Lernerstrategien, Testmanagementstrategien und *test-wiseness*-Strategien unterteilt wurden. Zum dritten Teil gehören weitere, nicht auf Strategien bezogene thematische Kategorien, die nicht nur für die Beantwortung der Forschungsfragen, sondern auch für die Reflexion der eingesetzten Datenerhebungsmethoden von Bedeutung sind. Dazu gehören u.a. Codes, mit denen die Wahrnehmung des Szenarios durch die Szenario-Schüler/innen oder die Rückmeldungen der Schüler/innen zum Lauten Denken während der Testbewältigung erfasst wurden.

Im Folgenden wird der Aufbau des Kategoriensystems vorgestellt. Der vollständige Kategorienleitfaden inkl. Kategoriendefinitionen und Beispielen befindet sich im Anhang 5. Die Ergebnisse der Analyse der codierten Datenauszüge werden in Kap. 4 vorgestellt.

3.7.3.1 CODIERUNG DER ALLGEMEINEN INFORMATIONEN ZU DEN ITEMS

Um in den Transkripten eine allgemeine Orientierung und spätere Analyse zu erleichtern, wurden in einem ersten Schritt die jeweiligen Untersuchungsteile mithilfe der entsprechenden Codes markiert. Gleichzeitig wurde auch der Zusammenhang der Aufgaben (Szenario oder Off-Szenario) kenntlich gemacht. Dafür wurden sechs Hauptkategorien vorgesehen:

- SC-LD: Szenario, Untersuchungsteil Lautes Denken
- SC-SR: Szenario, Untersuchungsteil *Stimulated-Recall*-Interview
- SC-Abschluss: Allgemeine, abschliessende Diskussion mit den Szenario-Schüler/innen nach dem *Stimulated-Recall*-Interview
- Off-LD: Off-Szenario, Untersuchungsteil Lautes Denken
- Off-SR: Off-Szenario, Untersuchungsteil *Stimulated-Recall*-Interview
- Off-Abschluss: Allgemeine, abschliessende Diskussion mit den Off-Szenario-Schüler/innen nach dem *Stimulated-Recall*-Interview

Als Subkategorien der Hauptkategorien SC-LD, SC-SR, Off-LD und Off-SR wurden die Bezeichnungen der jeweiligen Bildschirmseiten pro Aufgabe vorgesehen. So wurden beispielsweise in den Transkripten der Szenario-Schüler/innen die jeweiligen Teile der Lautdenkprotokolle zu Task A mithilfe folgender Subcodes erfasst:

- SC-LD\Intro: Bildschirmseite mit der Szenarioeinführung
- SC-LD\A-S1: Szenario, Lautes Denken, Task A (Bildschirm)Seite 1, inkl. Bildschirmseite mit dem Intro zu Task A⁶²

⁶² Diese einführenden Bildschirmseiten pro Aufgabe gehörten zum Szenario und wurden folglich nur von den Szenario-Schüler/innen gesehen. Weil zu diesen Bildschirmseiten in vielen Fällen sehr wenig verbalisiert wurde, aber auch um die Kategorien für die Szenario- und Off-Szenario-Aufgaben möglichst

- SC-LDVA-S2: Szenario, Lautes Denken, Task A Seite 2
- SC-LDVA-S3: Szenario, Lautes Denken, Task A Seite 3

Die entsprechenden Codes für das *Stimulated-Recall*-Interview zu diesen Bildschirmseiten lauten:

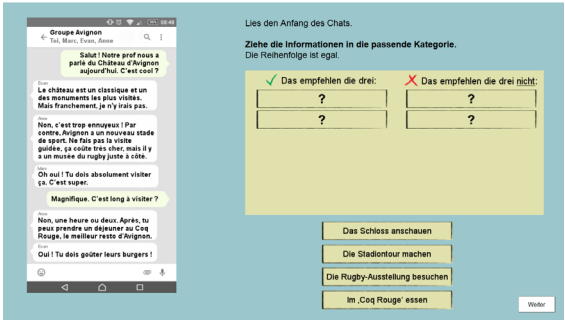
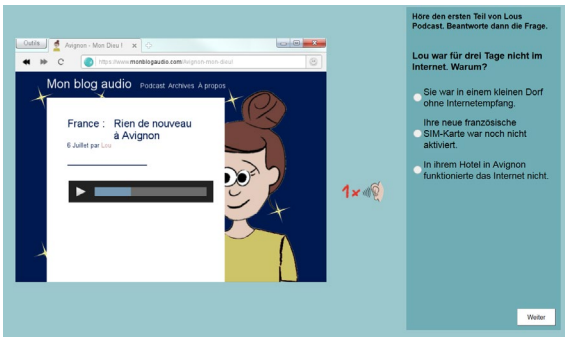
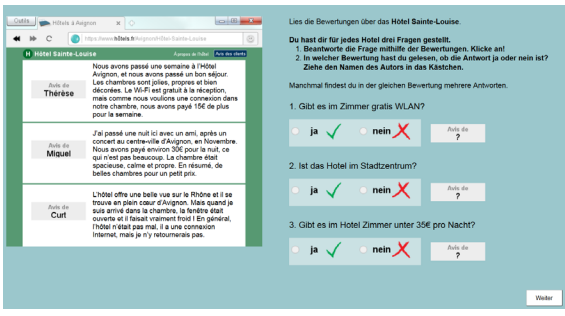
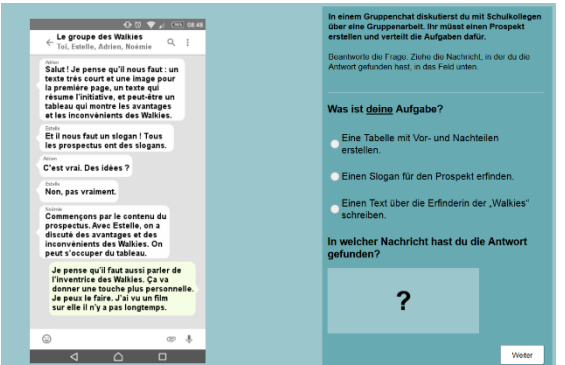
- SC-SR\Intro *Stimulated Recall*, Bildschirmseite mit der Einführung inkl. Diskussion am Anfang
- SC-SRVA-S1 Szenario, *Stimulated Recall*, Task A Seite 1, inkl. Bildschirmseite mit dem Intro zu Task A
- SC-SRVA-S2 Szenario, *Stimulated Recall*, Task A Seite 2
- SC-SRVA-S3 Szenario, *Stimulated Recall*, Task A Seite 3

In den Transkriptteilen zu den Bildschirmseiten, die für die vertiefende Analyse ausgewählt wurden, wurde in einem weiteren Schritt die Richtigkeit der Lernerantworten erfasst. Dafür wurden die Kategorien „R“ (= „richtig“) und „F“ (= „falsch“) vorgesehen.

Da es zu einigen, eher experimentellen Aufgabenformaten des Projekts IFB keine allgemeingültigen Vorgaben bezüglich der Auswertung richtiger und falscher Antworten gibt, wurden die Auswertungsrichtlinien vom Projektteam pro Aufgabenformat festgelegt, wie in Tabelle 8 ersichtlich ist. Diese Auswertungsrichtlinien wurden in der vorliegenden Studie übernommen.

parallel zu gestalten, wurden die Kommentare zu diesen Seiten gemeinsam mit der ersten Bildschirmseite zu der jeweiligen Aufgabe erfasst.

Tabelle 8: Codierung der Antworten als „richtig“ oder „falsch“.

Task(s) pro Aufgabenformat	Beispiel-Screenshot des Aufgabenformats	Codierung der Richtigkeit der Antwort(en)
1) Szenario „Stadtrip“, Task A		In dieser Aufgabe lagen für die vier Antwortmöglichkeiten auch vier Optionen vor. Es genügte folglich, nur zwei Antwortmöglichkeiten als „richtig“ oder „falsch“ zu erfassen. Es wurden hier nur die Antworten unter „Das empfehlen die drei“ gezählt (maximale Punktzahl: 2) und codiert. Die Reihenfolge der Antworten spielte dabei keine Rolle.
2) Szenario „Stadtrip“, Task B		Bei dieser Aufgabe konnten die Schüler/innen nur eine Antwort pro Bildschirmseite auswählen. Diese Auswahl wurde als richtig oder falsch codiert.
3) Szenario „Stadtrip“, Task C 4) Szenario „Berufswahl“, Task C		In dieser Aufgabe mussten die Schüler/innen die Fragen mit „ja“ oder „nein“ beantworten und jeweils auch angeben, welche Nachrichtautorin oder welcher Nachrichtautor eine Aussage dazu machte. Als richtig wurden nur diejenigen Antworten codiert, bei welchen beide Informationen korrekt waren (Aussage <u>und</u> Autor/in).
5) Szenario „Walkies“, Task C		Bei dieser Aufgabe konnten die Schüler/innen nur eine Antwort pro Bildschirmseite auswählen und mussten angeben, in welcher Nachricht sie diese Antwort fanden. Als richtig wurden nur diejenigen Antworten codiert, bei welchen beide Informationen korrekt waren (Antwort <u>und</u> entsprechende Nachricht).

Mit den Codes „R“ und „F“ wurden in MAXQDA (VERBI Software, 2019) pro Bildschirmseite nur die endgültigen Antworten der Schüler/innen erfasst. Als solche wurden die Antworten betrachtet, die die Lernenden beim Lauten Denken vor dem Anklicken der Taste „weiter“ zuletzt ausgewählt hatten.

3.7.3.2 DEFINITION UND CODIERUNG DER STRATEGIEN

Wie eingangs erwähnt, wurden alle Strategien beim Erstellen des Kategoriensystems als Subkategorien einer der drei Hauptkategorien zugeordnet. Diese Hauptkategorien wurden gemäss dem Systematisierungsvorschlag von A. D. Cohen (u.a. 2012, 2014) festgelegt: 1) Lernerstrategien, 2) Testmanagementstrategien und 3) *test-wiseness*-Strategien). Diese Hauptkategorien und ihre Subkategorien werden in diesem Kapitel nach einigen allgemeinen Bemerkungen zum Codierprozess vorgestellt.

Die zu codierenden Daten bestanden aus den Lautdenkprotokollen der 20 Schüler/innen beim Bearbeiten der Testaufgaben sowie den Äusserungen der gleichen Schüler/innen im anschliessenden *Stimulated-Recall*-Interview. Im Fokus der beiden Untersuchungsteile stand stets die Vorgehensweise der Lernenden bei der Testbewältigung. Die *Stimulated-Recall*-Interviews dienten hier primär der besseren Interpretation der Lautdenkprotokolle. Auch sollten dank der Interviews ergänzende Informationen zur Vorgehensweise der Lernenden beim Lösen der Testaufgaben und zu ihrer Wahrnehmung der Aufgaben und der Datenerhebungssituation gewonnen werden. Wie bei der Besprechung der beiden Datenerhebungsmethoden deutlich gemacht wurde (vgl. Kap. 3.3), weisen die erhobenen Daten unterschiedliche Merkmale auf. Im Hinblick auf die anschliessende Analyse war es folglich wichtig, alle codierten Segmente unmissverständlich und auf den ersten Blick dem jeweiligen Untersuchungsteil zuordnen zu können. Um das möglich zu machen, wurden alle strategienbezogenen Haupt- und Subkategorien nach ihrer Erstellung dupliziert. Sie wurden dann mit einem analogen Codenamen versehen, der ihre jeweilige Zugehörigkeit zum Lautdenkprotokoll bzw. zum *Stimulated-Recall*-Interview kennzeichnete. So wurde beispielsweise der Code LD-L01 für die Codierung der Lernerstrategie „Sorgfältiges Lesen“ im Lautdenkprotokoll benutzt, während der parallele Code SR-L01 bei der Codierung dieser Strategie im *Stimulated-Recall*-Interview zum Einsatz kam. Die Kategorien zu den beiden Untersuchungsteilen wurden teilweise anders definiert, um dem unterschiedlichen Charakter der Daten besser Rechnung zu tragen. Die Definition der oben erwähnten Strategie LD-L01 lautet beispielsweise: „Schüler/in liest den gesamten Inputtext sorgfältig durch“. Die entsprechende *Stimulated-Recall*-Kategorie wurde hingegen wie folgt definiert: „Es geht aus dem Interview hervor, dass der/die S den gesamten Inputtext sorgfältig durchgelesen hat“.

In den Transkriptteilen zum *Stimulated Recall* wurden ausserdem oft längere Passagen mit dem gleichen Code versehen, damit die codierten Segmente möglichst auch noch ausserhalb des Kontextes nachvollzogen werden können.

Am Ende der Besprechung einer Bildschirmseite im *Stimulated-Recall*-Teil wurden die Schüler/innen jeweils gebeten, kurz auf Deutsch zusammenzufassen, was sie in den französischen Inputtexten verstanden. Dank dieser mündlichen Zusammenfassungen sollten die Äusserungen der Lernenden in den Lautdenkprotokollen und im *Stimulated-Recall*-Interview besser nachvollzogen werden können. Um die Inputtexte zu rekapitulieren, setzten sich die Lernenden mit den Inputtexten nochmals intensiv auseinander und lasen bzw. hörten diese auch zum wiederholten Mal. Es liegt daher nahe, dass die Schüler/innen dabei möglicherweise anders als bei der ursprünglichen Testbewältigung vorgehen und die Texte u.U. auch aufmerksamer rezipierten. Aus diesem Grund wurden in den Transkriptteilen zu den mündlichen Zusammenfassungen diejenigen Stellen mit den passenden Kategorien codiert, an denen von den Lernenden ihre Vorgehensweise bei der Testbewältigung und nicht bei der mündlichen Zusammenfassung angesprochen wurde. Bei den mündlichen Zusammenfassungen wurden aber mithilfe eines speziell dafür vorgesehenen Codes Äusserungen der Schüler/innen erfasst, aus denen hervorgeht, dass sie ihre Antworten aufgrund der wiederholten Lektüre nun anpassen würden (Code SR-TM20, vgl. Tabelle 10). Mit der Kategorie „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ wurden in den mündlichen Zusammenfassungen jene Passagen in den Transkripten codiert, in denen die Schüler/innen ihr Unverständnis eines Wortes, eines Satzes oder eines grösseren Textteils andeuteten. Hintergrund ist, dass man vermuten kann, dass die Lernenden diese Textstellen auch schon bei der ursprünglichen Bearbeitung der Aufgabe nicht verstanden hatten. Wenn die Lernenden in einer längeren Passage nur einzelne Wörter verstanden, wurde die ganze Passage nur einmal mit diesem Code versehen. Aus diesem Grund ist die Quantifizierung der in mündlichen Zusammenfassungen mit diesem Code versehenen Stellen nicht als aussagekräftig zu betrachten.

Die Subkategorien zu den Strategien wurden, wie gesagt, den drei Hauptkategorien Lernerstrategien, Testmanagementstrategien und *test-wiseness*-Strategien zugeordnet. Im Folgenden wird näher auf diese Kategorien eingegangen.

Lernerstrategien

Als Lernerstrategien wurden in Anlehnung an A. D. Cohen (u.a. 2011b, 2014) die im Lautdenkprotokoll und/oder im *Stimulated-Recall*-Interview verbalisierten Gedanken und Handlungen codiert, auf die die Schüler/innen (mehr oder weniger) bewusst zurückgriffen, um die französischen Inputtexte zu verstehen. Diese Strategien sind nicht spezifisch für eine Testsituation, d.h. sie könnten (auch) bei der Rezeption eines beliebigen französischen Textes zum Einsatz kommen. In Tabelle 9 werden 18 Lernerstrategien aufgeführt, die in den vorliegenden Daten mindestens einmal beim Lauten Denken oder beim *Stimulated-Recall*-Interview codiert wurden. Die zwei letzten Spalten der Tabelle zeigen auf, wie die Kategorien für die beiden Untersuchungsteile definiert wurden.

Tabelle 9: Codierte Lernerstrategien.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Lernerstrategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-L01	SR-L01	Sorgfältiges Lesen	S ⁶³ liest den gesamten Inputtext sorgfältig durch.	Es geht aus dem Interview hervor, dass der/die S den gesamten Inputtext sorgfältig durchgelesen hat.
LD-L02	SR-L02	Schnelles Lesen	S überfliegt den gesamten Inputtext schnell. Er/Sie sucht dabei evtl. nach Schlüsselwörtern.	Es geht aus dem Interview hervor, dass der/die S den gesamten Inputtext schnell überflogen hat. Er/Sie hat dabei evtl. nach Schlüsselwörtern gesucht.
LD-L03	SR-L03	Einen Teil des Inputtextes (erneut bzw. genauer) lesen	S liest einen Teil des Inputtextes erneut bzw. genauer durch.	Es geht aus dem Interview hervor, dass der/die S einen Teil des Inputtextes erneut bzw. genauer durchgelesen hat.
LD-L04	SR-L04	Weiterlesen, um Nichtverstandenes zu verstehen	S verwendet andere Teile des Inputtextes, um das Verständnis eines bestimmten Teils zu erleichtern: Er/Sie sagt, dass er/sie weiterlesen will, um nach Informationen zu suchen, die zum Verständnis des bereits Gelesenen beitragen können. ⁶⁴	Es geht aus dem Interview hervor, dass der/die S Teile des Inputtextes verwendet hat, um das Verständnis eines bestimmten Teils zu erleichtern: Er/Sie sagt, dass er/sie weitergelesen hat, um nach Informationen zu suchen, die zum Verständnis des bereits Gelesenen beitragen können.
LD-L05	SR-L05	Bisheriges Verständnis hinterfragen	Nachdem er/sie mehr gelesen oder gehört hat, fragt sich der/die S, ob das bisherige Verständnis in Anbetracht der neuen Informationen noch zutrifft.	Es geht aus dem Interview hervor, dass der/die S sich fragte, nachdem er/sie mehr gelesen oder gehört hatte, ob das bisherige Verständnis in Anbetracht der neuen Informationen noch zuträfe.
LD-L06	SR-L06	Bisheriges Verständnis korrigieren	S korrigiert das Verständnis des Textes oder eines Teils davon, nachdem er/sie den Text erneut gelesen bzw. weitergelesen/weitergehört hat.	Es geht aus dem Interview hervor, dass der/die S das Verständnis des Textes oder eines Teils davon korrigierte, nachdem er/sie den Text erneut gelesen bzw. weitergelesen/weitergehört hatte.
LD-L07	SR-L07	Bisheriges Verständnis bestätigen	S bestätigt das aktuelle Verständnis des Inputtextes oder eines Teils davon, nachdem er/sie sich damit weiter und/oder vertiefter auseinandergesetzt hat.	Es geht aus dem Interview hervor, dass der/die S das endgültige Verständnis des Inputtextes oder eines Teils davon bestätigte, nachdem er/sie sich damit weiter und/oder vertiefter auseinandergesetzt hatte.

⁶³ Schüler/in.

⁶⁴ Strategien, die nur in einem Untersuchungsteil (Lautes Denken oder *Stimulated Recall*) vorkamen, werden im Kategorienleitfaden übersichtshalber in beiden Teilen aufgeführt. In dem Untersuchungsteil, in dem Sie nicht beobachtet werden konnten, wurden sie grau markiert.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Lernerstrategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-L08	SR-L08	Intensive Auseinandersetzung mit Teilen des Inputtextes	S liest vor, wiederholt, paraphrasiert bzw. übersetzt Wörter, Satzteile oder Sätze, oder fasst Teile des Inputtextes zusammen.	Es geht aus dem Interview hervor, dass der/die S den Inputtext vorgelesen, wiederholt, paraphrasiert, übersetzt oder zusammengefasst hat.
LD-L09	SR-L09	Informationen verknüpfen	S setzt Teile von gelesenen bzw. gehörten Informationen miteinander in Beziehung.	Es geht aus dem Interview hervor, dass der/die S Teile von gelesenen bzw. gehörten Informationen miteinander in Beziehung gesetzt hat.
LD-L10	SR-L10	Unbekannte Wörter, Satz- bzw. Textteile identifizieren	S identifiziert ein unbekanntes Wort, einen unbekannten Satz- bzw. Textteil oder er/sie zweifelt an seinem/ihrer Verständnis dieses Wortes oder dieser Satz- bzw. Textteile.	Es geht aus dem Interview hervor, dass der/die S ein unbekanntes Wort, einen unbekannten Satz- bzw. Textteil identifiziert hat oder dass er/sie an seinem/ihrer Verständnis dieses Wortes oder dieser Satz- bzw. Textteile gezweifelt hat.
LD-L11	SR-L11	Unbekannte Wörter ausblenden	S macht eine Äußerung dazu, dass er/sie sich nicht durch unbekannte Wörter ablenken lässt und diese ausblendet.	Es geht aus dem Interview hervor, dass der/die S sich nicht durch unbekannte Wörter ablenken liess und diese ausgeblendet hat.
LD-L12	SR-L12	Wortbedeutung identifizieren	S identifiziert die Bedeutung eines Wortes bzw. erinnert sich daran, z.B. aufgrund vom Kontext.	Es geht aus dem Interview hervor, dass der/die S die Bedeutung eines Wortes z.B. aufgrund des Kontextes identifiziert bzw. dass er/sie sich an die Wortbedeutung erinnert hat.
LD-L13	SR-L13	Morphosyntaktische Regeln anwenden	S nutzt explizit seine/ihre Kenntnisse der morphosyntaktischen Regeln.	Es geht aus dem Interview hervor, dass der/die S explizit seine/ihre Kenntnisse der morphosyntaktischen Regeln genutzt hat.
LD-L14	SR-L14	Interpunktionskennt nisse nutzen	S nutzt seine/ihre Kenntnisse von Interpunktionszeichen	Es geht aus dem Interview hervor, dass der/die S seine/ihre Kenntnisse von Interpunktionszeichen genutzt hat.
LD-L15	SR-L15	Kognaten identifizieren und nutzen	S identifiziert und nutzt (vermeintliche) Kognaten im Text.	Es geht aus dem Interview hervor, dass der/die S im Text (vermeintliche) Kognaten identifiziert und genutzt hat.
LD-L16	SR-L16	Prosodie beim Hören nutzen	S achtet beim Hören auf die Prosodie und nutzt diese Informationen.	Es geht aus dem Interview hervor, dass der/die S beim Hören auf die Prosodie geachtet und diese Informationen genutzt hat.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Lernerstrategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-L17	SR-L17	Relevantes Hintergrundwissen einbeziehen	S identifiziert und nutzt relevantes Hintergrundwissen in geeigneter Weise, d.h. er/sie greift <i>zusätzlich</i> zu den Informationen aus dem Text auf sein/ihr Weltwissen zurück, um die Textbedeutung zu verstehen.	Es geht aus dem Interview hervor, dass der/die S relevantes Hintergrundwissen identifiziert und es dann in geeigneter Weise genutzt hat, d.h. er/sie hat <i>zusätzlich</i> zu den Informationen aus dem Text auf sein/ihr Weltwissen zurückgegriffen, um die Textbedeutung zu verstehen.
LD-L18	SR-L18	Selbsteinschätzung	S macht beim Lauten Denken eine selbsteinschätzende (positive oder negative) Aussage in Bezug auf die eigenen Französischkenntnisse bzw. das Textverständnis.	Es geht aus dem Interview hervor, dass der/die S beim Lauten Denken eine selbsteinschätzende (positive oder negative) Aussage in Bezug auf die eigenen Französischkenntnisse bzw. das Textverständnis gemacht hat.

Bei zwei der oben aufgeführten Lernerstrategien wurden spezielle Gewichtungen „10“ und „25“ festgelegt, um bei der Analyse eine weitere Differenzierung der mit diesen Kategorien codierten Textsegmente zu ermöglichen. Dies betraf Strategie L10 „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ im *Stimulated-Recall*-Teil (SR-L10) und Strategie L15 „Kognaten identifizieren und nutzen“ in beiden Untersuchungsteilen (LD-L15 und SR-L15).

Mithilfe der Gewichtung „25“ wurden alle mit der Kategorie SR-L10 („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“) codierten Stellen gekennzeichnet, in denen die Schüler/innen bei der mündlichen Zusammenfassung der Inputtexte ein unbekanntes Wort oder eine unverständliche Textstelle vorfanden. Dank dieser Gewichtung konnte bei der anschliessenden Analyse erfasst werden, ob das codierte Segment im Gespräch mit der/dem Lernenden oder erst später bei der mündlichen Zusammenfassung des Lese- oder Hörtextes erwähnt wurde. Mit dem Code SR-L10 und seiner Gewichtung „25“ wurden bei den mündlichen Zusammenfassungen aus Praktikabilitätsgründen teilweise auch längere Äusserungen erfasst, z.B. wenn sich zeigte, dass die Schüler/innen immer wieder Teile eines grösseren Textfragments nicht verstanden. Beim Zusammenfassen der Texte wurden von gewissen Schüler/innen viele solche Wörter bzw. Stellen genannt, was damit zusammenhing, dass sie die Texte sehr gründlich ins Deutsche zu übersetzten versuchten. Andere Schüler/innen konzentrierten sich hingegen bei der mündlichen Zusammenfassung ausschliesslich auf das, was sie verstanden. Da sie dabei nicht erwähnten, was sie nicht verstanden, konnte bei ihnen der Code SR-L10 nicht angewandt werden.

Für die weitere Nuancierung der mit LD-L15 und SR-L15 („Kognaten identifizieren und nutzen“) codierten Segmente im Sinne der evaluativen qualitativen Inhaltsanalyse wurde die Gewichtung „10“ festgelegt. Diese Gewichtung wurde bei den codierten Stellen dann erfasst, wenn der Einsatz dieser Strategie zu Resultaten führte, die gemäss der Einschätzung der Forscherin als nicht zielführend betrachtet wurden. Dies betraf vor allem folgende Fälle:

- Ein Wort wurde von dem/der Lernenden fälschlicherweise als Kognat eingestuft.
- Die von dem/der Schüler/in ermittelte Bedeutung des (vermeintlichen) Kognaten stimmt im Textkontext nicht.
- Der/die Schüler/in vermutet, dass es sich bei einem Wort um einen Kognaten handelt. Er/sie kann aber nicht sagen, was das Wort auf Deutsch bedeutet.

Durch die vorgenommene Gewichtung konnte festgestellt werden, in wie vielen Fällen sich der Einsatz dieser Strategie als problematisch erwies (vgl. Kap. 4).

Testmanagementstrategien

Als Testmanagementstrategien wurden diejenigen Strategien der Lernenden codiert, die als sinnvolle, bewusste und zielorientierte Reaktionen auf die Testaufgabe(n) eingeschätzt wurden (vgl. u.a. A. D. Cohen, 2011b, 2014). Diese Strategien sind typisch und spezifisch für eine Testsituation und wurden in der vorliegenden Studie z.T. durch die Aufgabenformate ausgelöst, die im Projekt IFB zum Einsatz kamen. Tabelle 10 gibt einen systematischen Überblick über die Testmanagementstrategien, die in den Lautdenkprotokollen und/oder *Stimulated-Recall*-Interviews codiert wurden. Während die ersten 19 Strategien beide Untersuchungsteile betrafen, bezog sich die zuletzt aufgeführte Kategorie „Selbstkorrektur beim SR“ (SR-TM20) nur auf das *Stimulated-Recall*-Interview.

Tabelle 10: Codierte Testmanagementstrategien.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Testmanagement- strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-TM01	SR-TM01	Instruktionen und/oder Antwortoptionen (erneut) lesen	S ⁶⁵ liest die Instruktionen und/oder Antwortoptionen (erneut) durch.	Es geht aus dem Interview hervor, dass der/die S die Instruktionen und/oder Antwortoptionen (erneut) durchgelesen hat.
LD-TM02	SR-TM02	Paraphrasieren und/oder Übersetzen der Aufgabe⁶⁶	S paraphrasiert die Aufgabe oder übersetzt für sich die Frage, die Antwortoptionen oder Teile davon.	S sagt oder es geht aus dem Interview hervor, dass er/sie die Aufgabe paraphrasiert oder die Frage, die Antwortoptionen oder ihre Teile für sich übersetzt hat.
LD-TM03	SR-TM03	Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren⁶⁷	S liest die Aufgabenstellung bzw. Antwortoptionen und identifiziert das ihm/ihr darin unbekannte Vokabular.	Es geht aus dem Interview hervor, dass der/die S die Aufgabenstellung bzw. Antwortoptionen gelesen und das ihm/ihr darin unbekannte Vokabular identifiziert hat.
LD-TM04	SR-TM04	Vorhersage machen	S macht eine Vorhersage: Er/Sie überlegt sich vor der Rezeption des Inputtextes, welche Antwortoption(en) evtl. stimmen und welche ausgeschlossen werden könnten. Aus den weiteren Teilen des Transkripts geht aber hervor, dass er/sie diese Antwortoption(en) vor der Textrezeption noch nicht ausgeschlossen hat.	Es geht aus dem Interview hervor, dass der/die S eine Vorhersage gemacht hat. Er/Sie hat sich vor der Rezeption des Inputtextes überlegt, welche Antwortoption(en) evtl. stimmen und welche ausgeschlossen werden könnten, er/sie hat aber diese Antwortoption(en) vor der Textrezeption noch nicht ausgeschlossen.
LD-TM05	SR-TM05	Frage und Antwortoptionen beim Hören anschauen	S schaut sich die Frage und/oder Antwortoptionen beim Hören an.	Es geht aus dem Interview hervor, dass der/die S sich die Frage und/oder Antwortoptionen beim Hören angeschaut hat.

⁶⁵ Schüler/in.

⁶⁶ Die Aufgabenstellung, Fragen und Antwortoptionen wurden in der Schulsprache Deutsch formuliert. Bei der Off-Szenario-Aufgabe „Berufswahl“ kamen in der Aufgabenstellung auch französische Berufsbezeichnungen vor.

⁶⁷ Die Aufgabenstellung, Fragen und Antwortoptionen wurden in der Schulsprache Deutsch formuliert. Bei der Off-Szenario-Aufgabe „Berufswahl“ kamen in der Aufgabenstellung auch französische Berufsbezeichnungen vor.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Testmanagement- strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-TM06	SR-TM06	Fokus auf relevante Textteile	S liest die Frage und/oder Antwortoption(en). Er/Sie konzentriert sich beim Hören/Lesen auf die Teile des Inputtextes, die für die Beantwortung der Frage von Bedeutung sind und sucht, wo im Inputtext die entsprechenden Informationen (vermutlich) vorkommen. Evtl. ignoriert er/sie dabei die als nicht relevant eingestuft Teile des Inputtextes.	Es geht aus dem Interview hervor, dass der/die S die Frage und/oder Antwortoption(en) gelesen und sich beim Hören/Lesen auf die Teile des Inputtextes konzentriert hat, die für die Beantwortung der Frage von Bedeutung sind. Er/Sie hat gesucht, wo im Inputtext die entsprechenden Informationen (vermutlich) vorkommen und hat evtl. die als nicht relevant eingestuft Teile des Inputtextes ignoriert.
LD-TM07	SR-TM07	Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären	S klärt oder bestätigt die Bedeutung von bestimmten Teilen des französischen Inputtextes mithilfe der auf Deutsch vorhandenen Frage und/oder der Antwortoptionen.	Es geht aus dem Interview hervor, dass der/die S die Bedeutung von bestimmten Teilen des französischen Inputtextes mithilfe der auf Deutsch vorhandenen Frage bzw. der Antwortoptionen geklärt oder bestätigt hat.
LD-TM08	SR-TM08	Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)	S wählt oder schliesst eine Antwortoption bzw. -optionen aufgrund von etwas aus, was er/sie vom Inputtext verstanden hat.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption bzw. -optionen aufgrund von etwas ausgewählt oder ausgeschlossen hat, was er/sie vom Inputtext verstanden hat.
LD-TM09	SR-TM09	Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden	S liest bzw. hört den Inputtext nicht weiter, nachdem er/sie die (vermeintlich) für die Beantwortung der Frage benötigten Informationen im Text gefunden hat.	Es geht aus dem Interview hervor, dass der/die S den Inputtext nicht weiterlas bzw. -hörte, nachdem er/sie die (vermeintlich) für die Beantwortung der Frage benötigten Informationen im Text gefunden hatte.
LD-TM10	SR-TM10	Richtigkeit der Antwortoption überprüfen	S geht zur Frage und/oder zum Inputtext zurück, um zu prüfen, ob die in Erwägung gezogene Antwortoption korrekt ist oder nicht. Er/Sie macht das bevor oder nachdem er/sie eine Antwortoption wählt oder gewählt hat.	Es geht aus dem Interview hervor, dass der/die S zur Frage und/oder zum Inputtext zurückging, um zu prüfen, ob die in Erwägung gezogene Antwortoption korrekt wäre oder nicht. Er/Sie machte das, bevor oder nachdem er/sie eine Antwortoption wählte oder gewählt hatte.
LD-TM11	SR-TM11	Antwortoptionen miteinander vergleichen	S vergleicht die möglichen Antwortoptionen miteinander.	Es geht aus dem Interview hervor, dass der/die S die möglichen Antwortoptionen miteinander verglichen hat.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Testmanagement- strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-TM12	SR-TM12	Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen	S wählt oder schliesst eine Antwortoption aufgrund von Inferenzen aus. Kommt oft bei jenen Items vor, bei denen die Antwort nicht wortwörtlich im Text steht.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption auf der Grundlage von Inferenzen vom Inputtext gewählt hat. Kommt oft bei jenen Items vor, bei denen die Antwort nicht wortwörtlich im Text steht.
LD-TM13	SR-TM13	Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war	S wählt eine Antwortoption aus, die ein Wort bzw. Wörter enthält, weil er/sie im Inputtext korrekterweise ein passendes Wort oder passende Wörter zu den Schlüsselwörtern aus dieser Antwortoption identifiziert hat. Kommt oft bei Items vor, die auf selektives Lesen (<i>search reading</i>) abzielen.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, weil er/sie im Inputtext korrekterweise ein passendes Wort oder passende Wörter zu den Schlüsselwörtern aus dieser Antwortoption identifiziert hat. Kommt oft bei Items vor, die auf selektives Lesen (<i>search reading</i>) abzielen.
LD-TM14	SR-TM14	Auswahl einer Antwortoption, weil andere nicht plausibel sind ⁶⁸	S wählt eine Antwortoption bzw. -optionen aus, nachdem er/sie andere Antwortoptionen ausgeschlossen hat, die ihm/ihr basierend auf seinem/ihrem Verständnis des Inputtextes oder eines Teils davon als nicht plausibel erscheinen.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption bzw. -optionen auswählte, nachdem er/sie andere Antwortoptionen ausgeschlossen hatte, die ihm/ihr basierend auf seinem/ihrem Verständnis vom Inputtext oder eines Teils davon als nicht plausibel erschienen.
LD-TM15	SR-TM15	Kenntnisse über das Itemformat nutzen ⁶⁹	S nutzt bei Szenario „Stadttrip“, Task A die Kenntnisse über das Itemformat, um eine oder mehrere Antwortoptionen zu wählen. Wenn in Szenario „Stadttrip“, Task A z.B. zwei positive Antworten gefunden worden waren, dann wusste der/die S, dass die restlichen möglichen Antworten negativ sein müssen.	Es geht aus dem Interview hervor, dass der/die S beim Szenario „Stadttrip“, Task A die Kenntnisse über das Itemformat genutzt hat, um eine oder mehrere Antwortoptionen zu wählen. Wenn in Szenario „Stadttrip“, Task A z.B. zwei positive Antworten gefunden worden waren, dann wusste der/die S., dass die restlichen Antworten negativ sein müssen.

⁶⁸ Man könnte hier auch argumentieren, dass die Strategie konstruktirrelevant ist, weil die Entscheidung der Schülerin oder des Schülers nicht *für* eine Antwortoption fiel, sondern *gegen* die restlichen Antwortoptionen. Nach langem Überlegen wurde entschieden, diese Strategie doch als konstruktrelevant einzustufen, weil sie impliziert, dass die/der Lernende hier nicht einfach den Test „austricksen“ wollte, sondern die Frage aufgrund dessen zu beantworten versuchte, was sie/er vom Text verstand.

⁶⁹ Die Strategie wurde als konstruktrelevant definiert (d.h. als Testmanagement- und nicht als *test-wiseness*-Strategie), weil in der Aufgabe, bei der diese Strategie vorkam, ein Notizzettel mit vier Lücken mit zwei empfohlenen und zwei nicht empfohlenen Sachen vervollständigt werden sollte. Es ist also

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der Testmanagement- strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall-Interviews</i>
LD-TM16	SR-TM16	Zeitmanagement	S macht eine Aussage über ihr/sein Zeitmanagement. ⁷⁰	S macht eine Aussage über ihr/sein Zeitmanagement während der Testbewältigung.
LD-TM17	SR-TM17	Selbstkorrektur F-R	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere Antwort war falsch, die neu gewählte ist richtig.	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass die frühere Antwort falsch und die neu gewählte richtig war.
LD-TM18	SR-TM18	Selbstkorrektur R-F	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere Antwort war richtig, die neu gewählte ist falsch.	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass die frühere Antwort richtig und die neu gewählte falsch war.
LD-TM19	SR-TM19	Selbstkorrektur F-F	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere Antwort war falsch, die neugewählte ist ebenfalls falsch.	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass die frühere Antwort falsch und die neu gewählte ebenfalls falsch war.
-	SR-TM20	Selbstkorrektur beim SR	-	S merkt beim <i>Stimulated-Recall-Interview</i> , dass er/sie beim nochmaligen bzw. genaueren Lesen oder Hören nun mehr oder etwas anderes versteht. Er/Sie würde seine/ihre Antwort evtl. nachträglich korrigieren bzw. ist sich nicht mehr sicher, ob die früher gewählte Antwort tatsächlich stimmt. Es wird hier nicht festgehalten, ob die korrigierte Antwort richtig oder falsch wäre.

wahrscheinlich dem Format geschuldet, dass manche Schüler/innen zu lesen aufhörten, nachdem sie zwei positive oder zwei negative Punkte gefunden hatten. Auch im realen Leben würde man u.U. so vorgehen. Es ist aber ein Grenzfall: Man könnte diese Strategie auch als konstruktirrelevant einstufen.

⁷⁰ Strategien, die nur in einem Untersuchungsteil (Lautes Denken oder *Stimulated Recall*) vorkamen, werden im Kategorienleitfaden übersichtshalber in beiden Teilen aufgeführt. In dem Untersuchungsteil, in dem sie nicht beobachtet werden konnten, wurden sie grau markiert.

Die Codierungen von zwei der oben aufgeführten Testmanagementstrategien wurden in einem anschliessenden Schritt im Sinne der evaluativen qualitativen Inhaltsanalyse mithilfe der speziellen Gewichtungen „15“ und „20“ weiter ausdifferenziert. Es handelt sich dabei um die Strategien TM08 „Einsatz von Verstandenem bei Auswahl oder Ausschluss der Antwortoption(en)“ und TM12 „Auswahl oder Ausschluss der Antwortoption(en) aufgrund von Inferenzen“.

Mithilfe des Codes TM08 („Einsatz von Verstandenem bei Auswahl oder Ausschluss der Antwortoption(en)“) wurden in den Daten jene Passagen markiert, aus denen hervorgeht, dass die Lernenden bei der Auswahl oder auch beim Ausschluss bestimmter Antwortoptionen auf das zurückgegriffen haben, was sie vom Text verstanden haben. Während der Codierung konnte jedoch oft beobachtet werden, dass die Lernenden die entsprechenden Teile der Inputtexte entweder falsch oder auch offensichtlich nicht genügend verstanden, um darauf basierend die richtige Antwort zu wählen oder eine falsche Antwort aus dem richtigen Grund auszuschliessen. Beim letzten Codierdurchgang wurden daher alle Textpassagen, die von den Schüler/innen basierend auf ihren Verbalisierungen falsch oder nicht genügend verstanden wurden, mit der Gewichtung „15“ gekennzeichnet (bei den richtig verstandenen Passagen wurde die Standardgewichtung „0“ beibehalten). Es handelt sich in diesen Fällen um eine konstruktrelevante Strategie (also um eine Testmanagementstrategie und nicht um eine *test-wiseness*-Strategie), weil die Schüler/innen hier nicht versuchten, beim Lösen der Aufgabe den Inputtext aussen vor zu lassen. Im Gegenteil – sie bemühten sich darum, den französischen Text zu verstehen, wofür ihre Sprachkenntnisse in diesen Fällen aber noch nicht ausreichten.

Die Kategorie „Auswahl oder Ausschluss der Antwortoption(en) aufgrund von Inferenzen“ (LD-TM12 und SR-TM12) wurde codiert, wenn in den Daten beobachtet werden konnte, dass die Schüler/innen bei der Suche nach der richtigen Antwort auf Schlussfolgerungen zurückgriffen, die sie aufgrund des Inputtextes machten. Dies war bei einigen Items von den Testentwickler/innen auch so intendiert worden, d.h. die Antwort auf die gestellte Frage stand nicht wortwörtlich im Text und musste aus dem Kontext hergeleitet werden. In den mit TM12 codierten Segmenten konnte aber beobachtet werden, dass die Schüler/innen auch an anderen Stellen auf Inferenzen zurückgriffen, an denen dies unnötig war, weil die relevante Information explizit im Text formuliert gewesen wäre. Beim erneuten Codierdurchgang wurden daher alle mit dem Code TM12 hervorgehobenen Verbalisierungen der Schüler/innen sorgfältig mit dem Inputtext abgeglichen. Passagen, bei denen die Schüler/innen eine Antwort wegen ihrer Schlussfolgerungen wählten bzw. ausschlossen, welche aber nach dem Vergleich mit der entsprechenden Textstelle als unplausibel eingestuft wurden, weil es für

diese Deutung im Inputtext nicht genügend Evidenz gab, wurden mit der Gewichtung „20“ markiert. An dieser Stelle könnte die berechtigte Frage auftauchen, ob die mit Gewichtung „20“ markierten Ereignisse überhaupt konstruktrelevant oder eher konstruktirrelevant sind, d.h. ob diese Textsegmente nicht eher als eine separate *test-wiseness*-Strategie codiert werden sollten. Beim Einsatz dieser Strategie gingen die Schüler/innen jedoch immer vom Inputtext bzw. von einem bestimmten Fragment aus. Sie versuchten hier also nicht, den Inputtext zu ignorieren – im Gegenteil, sie setzten sich mit ihm z.T. intensiv auseinander. Es wurde daher entschieden, diese unplausiblen Schlussfolgerungen der Lernenden doch als eine (mit Gewichtung „20“ speziell gekennzeichnete) Variante der konstruktrelevanten Testmanagementstrategie TM12 zu codieren, und nicht als eine separate *test-wiseness*-Strategie.

Test-wiseness-Strategien

Als *test-wiseness*-Strategien wurden diejenigen Strategien der Lernenden codiert, die im Unterschied zu den Lernerstrategien und Testmanagementstrategien als konstruktirrelevant betrachtet wurden (vgl. A. D. Cohen, 2011b, 2014), und zwar im Hinblick auf das IFB-Testkonstrukt. Es handelt sich dabei um Strategien wie Raten oder Nutzen von Informationen über die Testformate. Wie auch in anderen Studien zu den Strategien bei der Testbewältigung (A. D. Cohen & Upton, 2006, 2007; L. Gu & So, 2017; Suvorov, 2018) fällt der Katalog der *test-wiseness*-Strategien dünner aus als bei den beiden konstruktrelevanten Strategien. In der vorliegenden Studie konnten insgesamt 7 *test-wiseness*-Strategien identifiziert und codiert werden (Tabelle 11).

Tabelle 11: Codierte *test-wiseness*-Strategien.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der <i>test-wiseness</i> - Strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall</i> -Interviews
LD-TW01	SR-TW01	Raten	S ⁷¹ rät (intelligent oder zufällig). Er/Sie muss aufgrund des Testformats eine Antwort wählen, ist sich aber nicht sicher, ob die gewählte Antwort auch stimmt. Entweder hat er/sie den Text nicht bzw. nicht aufmerksam genug rezipiert; oder aber nicht (genug) verstanden, um eine Antwort aufgrund von Verstandenenem zu wählen.	Es geht aus dem Interview hervor, dass der/die S geraten hat (intelligent oder zufällig). Er/Sie musste aufgrund des Testformats eine Antwort wählen, war sich aber nicht sicher, ob die gewählte Antwort auch stimmt. Entweder hat er/sie den Text nicht bzw. nicht aufmerksam genug rezipiert; oder aber nicht (genug) verstanden, um eine Antwort aufgrund von Verstandenenem zu wählen.
LD-TW02	SR-TW02	Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war	S wählt eine Antwort aus, weil ein Wort oder mehrere Wörter aus den Antwortoptionen auch im Inputtext in der gleichen oder ähnlichen Form vorkommen. Dabei wird aber die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigt bzw. verstanden. ⁷²	Es geht aus dem Interview hervor, dass der/die S eine Antwort ausgewählt hat, weil ein Wort oder mehrere Wörter aus den Antwortoptionen auch im Inputtext in der gleichen oder ähnlichen Form vorkamen. Dabei wurde aber die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigt bzw. verstanden.
LD-TW03	SR-TW03	Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern	S schliesst oder wählt eine oder mehrere Antwortoption(en) aus bzw. wählt bei „ja/nein-Fragen“ die negative Option aus, weil er/sie die entsprechenden Informationen nicht verstanden hat oder weil sie nach ihrem/seinem Verständnis im Inputtext nicht vorkamen. ⁷³	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption oder mehrere Antwortoptionen ausgeschlossen, ausgewählt bzw. bei „ja/nein-Fragen“ die negative Option angeklickt hat, weil er/sie die entsprechenden Informationen nicht verstanden hat oder weil sie nach ihrem/seinem Verständnis im Inputtext nicht vorkam(en).

⁷¹ Schüler/in.

⁷² Diese Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchläuft, die durch das Item vorgesehenen sind.

⁷³ Diese Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchläuft, die durch das Item vorgesehenen sind. Ausserdem waren die eingesetzten Aufgaben so konstruiert, dass in den Inputtexten jeweils Elemente aller Antwortoptionen vorkamen.

Code Lautes Denken	Code <i>Stimulated Recall</i>	Kurzbezeichnung der <i>test-wiseness</i> - Strategie	Definition der Strategie für die Codierung der Lautdenkprotokolle	Definition der Strategie für die Codierung der <i>Stimulated-Recall</i> -Interviews
LD-TW04	SR-TW04	Auswahl einer Antwortoption mit bekannten Wörtern	S wählt eine Antwortoption aus, die nur Wörter enthält, die er/sie im Inputtext verstehen konnte. ⁷⁴	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, die nur Wörter enthält, die er/sie im Inputtext verstanden hatte.
LD-TW05	SR-TW05	Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen	S schliesst nach der Rezeption des Inputtextes eine Antwortoption oder mehrere Antwortoptionen aus bzw. wählt eine Antwortoption, weil er/sie diese Antwortoption(en) aufgrund von eigenem Wissen, von eigenen Erfahrungen oder von eigenen Vorlieben (und nicht aufgrund des Inputtextes) für falsch oder für richtig hält.	Es geht aus dem Interview hervor, dass der/die S nach der Rezeption des Inputtextes eine Antwortoption oder mehrere Antwortoptionen ausgeschlossen bzw. eine Antwortoption ausgewählt hat, weil er/sie diese Antwortoption(en) aufgrund von eigenem Wissen, von eigenen Erfahrungen oder von eigenen Vorlieben (und nicht aufgrund des Inputtextes) für falsch oder für richtig hielt.
LD-TW06	SR-TW06	Auswahl einer anders aussehenden Antwortoption	S wählt eine Antwortoption, weil sie anders als die restlichen Antwortoptionen aussieht.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, weil sie anders als die restlichen Antwortoptionen aussah.
LD-TW07	SR-TW07	Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats	S wählt oder schliesst eine Antwortoption aus, weil er/sie bestimmte Vorstellungen davon hat, wie die (Sprach-)Tests normalerweise aufgebaut sein sollten bzw. weil er/sie die Aufgabeninstruktionen nicht aufmerksam genug gelesen hat.	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt oder ausgeschlossen hat, weil er/sie bestimmte Vorstellungen davon hatte, wie die (Sprach-)Tests normalerweise aufgebaut sein sollten bzw. weil er/sie die Aufgabeninstruktionen nicht aufmerksam genug gelesen hatte.

⁷⁴ Strategien, die nur in einem Untersuchungsteil (Lautes Denken oder *Stimulated Recall*) vorkamen, werden im Kategorienleitfaden übersichtshalber in beiden Teilen aufgeführt. In dem Untersuchungsteil, in dem Sie nicht beobachtet werden konnten, wurden sie grau markiert.

Für die *test-wiseness*-Strategien wurden ausser den für alle Strategie-Codes benutzten Gewichtungen „2“ (wiederholtes Nennen einer bereits codierten Strategie) und „5“ (allgemeine, itemübergreifende Aussage zum eigenen Testverhalten) keine speziellen Gewichtungen festgelegt.

Im Hinblick auf die Beantwortung der zweiten Forschungsfrage wurden im Sinne der evaluativen qualitativen Inhaltsanalyse alle vorgenommenen Codierungen der *test-wiseness*-Strategien gesichtet. Es wurde dabei pro Item, das unter Einbezug der codierten *test-wiseness*-Strategie(n) gelöst wurde, eingeschätzt, ob die Schüler/innen 1) aufgrund dieser Strategie(n) eine falsche Antwort wählten, 2) aufgrund dieser Strategie(n) die Frage richtig beantworten konnten, oder ob sie 3) zwar beim Bearbeiten dieses Items eine oder mehrere *test-wiseness*-Strategie(n) einsetzten, aber letztlich das Item aus einem konstruktrelevanten Grund richtig lösten (z.B. weil sie den Inputtext nochmals lasen und ihre Antwort korrigierten). Die Ergebnisse dieser vertiefenden Analyse werden in Kap. 4.3 vorgestellt.

Die in diesem Kapitel vorgestellten Kategorien zu den Strategien wurden für die Codierung von Transkripten sowohl der Szenario- wie auch der Off-Szenario-Schüler/innen eingesetzt.

3.7.3.3 CODIERUNG WEITERER THEMATISCHER KATEGORIEN

Neben den deduktiv und induktiv festgelegten Codes zu den Strategien wurden dem Kategorienleitfaden auch weitere thematische Codes hinzugefügt. Sie sind, wie bereits angedeutet, am Material und mit Blick auf die Beantwortung der Forschungsfragen bzw. auf die methodische Diskussion entstanden.

Die dritte Forschungsfrage betraf die eventuellen Unterschiede, die beim Einsatz der Strategien zum Lösen der szenariobasierten und nicht szenariobasierten Aufgaben beobachten werden konnten. Aus diesem Grund war es wichtig zu erfassen, ob die 10 Szenario-Schüler/innen beim Lösen des Tests die Einbettung der Aufgaben in ein übergreifendes Szenario überhaupt bemerkten. Für die Codierung der Lerneräusserungen in den Lautdenkprotokollen, die das Entdecken oder Wahrnehmen des Szenarios beim Lösen der Testaufgaben vermuten liessen, wurde die Kategorie „Entdeckung des Szenarios“ (LD-SC_Entd) vorgesehen. In den *Stimulated-Recall*-Interviews wurde in einem dreistufigen Verfahren zu eruieren versucht, ob die Schüler/innen die Szenarioeinbettung der Aufgaben während des Testbearbeitens überhaupt realisiert hatten. Sie wurden zuerst gefragt, 1) ob ihnen an den Aufgaben etwas Spezielles aufgefallen war. Wenn sie dabei die Szenarioeinbettung nicht erwähnten, wurden sie gefragt, 2) ob ihnen am Thema der Testaufgaben etwas aufgefallen war. Wenn sie dies verneinten, wurden sie letztlich gefragt,

3) ob ihnen aufgefallen war, dass es sich bei allen Aufgaben um die Planung einer Schulreise nach Avignon handelte. Mit der Kategorie „Szenario beim Test wahrgenommen“ (SR-SC_1) sollten ursprünglich alle Erwähnungen der Szenarioeinbettung in den Antworten auf die erste Frage codiert werden. Es zeigte sich aber, dass dies in keinem der 10 Interviews vorkam. Die Szenarioeinbettung wurde nur von einer Schülerin spontan beim *Stimulated Recall* angesprochen und diese Stelle wurde dann mit diesem Code gekennzeichnet. Für die Erfassung der Antworten auf die zweite und dritte Frage wurden die Kategorien „Szenario nach einem Hinweis wahrgenommen“ (SR-SC_2) und „Szenario überhaupt nicht wahrgenommen“ (SR-SC_3) bestimmt. Äusserungen, die zur letzten Kategorie passten, wurden in den Daten nicht gefunden und die Kategorie wurde am Schluss aus dem Kategorienleitfaden entfernt. Der Code „Szenario nach einem Hinweis wahrgenommen“ wurde hingegen vergeben, wenn die Schüler/innen 1) das Szenario (bzw. Avignon, Reisen etc.) nannten, nachdem sie gefragt wurden, ob ihnen thematisch etwas an den Aufgaben aufgefallen war; oder 2) wenn sie die dritte Frage der Forscherin bejahten (d.h. es fiel ihnen in Anschluss auf die Frage der Forscherin auf, dass die Aufgaben tatsächlich thematisch verbunden waren). Im zweiten Fall wurden die Äusserungen der Schüler/innen zusätzlich mit der Gewichtung „30“ versehen.

Die 10 Schüler/innen, die die Off-Szenario Aufgaben lösten, wurden in den Interviews gefragt, ob es ihnen etwas ausmachte, dass sich alle fünf Aufgaben auf ein anderes Thema bezogen. Um die Antworten der Schüler/innen zu erfassen, wurden die Kategorien „Übergreifendes Thema hat in Off-Szenario gefehlt“ und „Verschiedene Themen stören bei Off-Szenario nicht bzw. werden begrüsst“ (SR-Off-SC) vorgesehen. Da es sich zeigte, dass die Abwesenheit eines übergreifenden Themas offenbar niemandem gefehlt hatte und es dementsprechend auch nicht codiert wurde, wurde die erste Kategorie aus dem Kategorienleitfaden später entfernt.

Um für die Methodendiskussion Hinweise zum eventuellen Einfluss des Lauten Denkens auf die Testbewältigungsprozesse der Schüler/innen zu bekommen, wurden die Lernenden in den Interviews gefragt, ob es sie störte, dass sie beim Lösen der Aufgaben ihre Gedanken verbalisieren mussten. Die Antworten der Schüler/innen wurden mit den Codes „Lautes Denken hat (eher) gestört“ (SR-LD_1), „Lautes Denken hatte einen positiven Einfluss auf das Testverhalten“ (SR-LD_2) und „Lautes Denken war (eher) egal“ (SR-LD_3) kategorisiert.

Mit den Kategorien „LD-Sonstiges“ und „SR-Sonstiges“ wurden schliesslich in den Lautdenkprotokollen und *Stimulated-Recall*-Interviews interessante Äusserungen der Schüler/innen codiert, die eventuell im Rahmen anschliessender Untersuchungen analysiert werden könnten (z.B. Ansichten der Lernenden bezüglich der Sprachtests). Ausgewählte

Erkenntnisse, die aufgrund der mit dieser Codierung gekennzeichneten Stellen gewonnen wurden, werden in Kap. 4.5 dargestellt.

3.7.4 BERÜCKSICHTIGUNG DER GÜTEKRITERIEN FÜR DIE INTERNE STUDIENGÜTE BEI DER QUALITATIVEN INHALTSANALYSE

Bevor im nächsten Kapitel die Ergebnisse der Datenanalyse vorgestellt werden, soll an dieser Stelle nochmals ein Schritt zurückgegangen werden, um die Vorgehensweise bei der Datenauswertung kritisch zu betrachten. Da sich die Auswertung der Daten nach der von Kuckartz (2018) vorgeschlagenen Variante der qualitativen Inhaltsanalyse richtete, wird bei dieser Reflexion ebenfalls auf die von ihm formulierten Gütekriterien der qualitativen Inhaltsanalyse zurückgegriffen (ebd., Kap. 9).

Die Gütekriterien für die qualitative Forschung werden bereits seit den 1980er Jahren kontrovers diskutiert (ebd., S. 202; mit Verweis auf weiterführende Literatur). Kuckartz (ebd., S. 202–203) listet drei grundsätzliche Positionen auf, die innerhalb dieser Diskussion anzutreffen sind. Die erste postuliert die Universalität der Gütekriterien, d.h. spricht sich für gleiche Gütekriterien in der quantitativen und qualitativen Forschung aus. Von Vertreter/innen der zweiten Position wird die Spezifität der Kriterien der qualitativen Forschung und ihre Abgrenzung von den Gütekriterien der quantitativen Forschung gefordert. Die dritte Position, zu der sich auch Kuckartz in seinem Buch bekennt, schlägt einen „Weg zwischen strikter Ablehnung und bloßer Übertragbarkeit der Gütekriterien quantitativer Forschung“ ein (ebd., S. 203). Demnach sollten die Gütekriterien der qualitativen Forschung um- bzw. neu formuliert werden, was zur Festlegung spezifischer Standards führen sollte. Diese Standards sollten potenziell auch für forschungsfördernde Institutionen und die Bewertung von Forschungsanträgen relevant sein (ebd.).

Kuckartz schlägt vor, in Bezug auf die Gütekriterien der qualitativen Inhaltsanalyse zwischen der internen und externen Studiengüte zu unterscheiden (ebd., S. 203–218). Als Ausprägungen der internen Studiengüte listet er „Zuverlässigkeit, Verlässlichkeit, Auditierbarkeit, Regelgeleitetheit, intersubjektive Nachvollziehbarkeit, Glaubwürdigkeit etc.“ auf, zur externen Studiengüte zählen hingegen „die Fragen der Übertragbarkeit und Verallgemeinerbarkeit“ (ebd., S. 203). Er merkt weiter an, dass für die qualitative Inhaltsanalyse eher Kriterien interner Studiengüte zu formulieren seien, da es sich dabei um ein Verfahren zur *Auswertung* der qualitativen Daten handle; während sich die externe Studiengüte auf die gesamte Anlage der qualitativen Forschung beziehe (wofür die interne Studiengüte wiederum eine notwendige Voraussetzung sei).

Kuckartz schlägt für die Überprüfung der internen Studiengüte eine in Form von Fragen formulierte Checkliste vor (ebd., S. 204–205), die a) zuerst wichtige Punkte betreffend der Datenerfassung und Transkription auflistet⁷⁵ und b) danach auf Punkte hinweist, die bei der Durchführung der qualitativen Inhaltsanalyse im engeren Sinn als relevant zu erachten sind⁷⁶. Im Folgenden wird dargestellt, wie die Punkte, die Kuckartz unter a) und b) auflistet, in der hier vorliegenden Studie berücksichtigt bzw. umgesetzt wurden. Auf die Fragen der externen Studiengüte wird dann anlässlich der Diskussion der Resultate eingegangen (Kap. 5).

Die Datensitzungen zur Erhebung der Lautdenkprotokolle und *Stimulated-Recall*-Daten wurden von der Autorin der vorliegenden Studie unter Einbezug der relevanten Fachliteratur (u.a. Bowles, 2010; Gass & Mackey, 2017; Heine & Schramm, 2016; Rossa, 2012) konzipiert und durchgeführt (vgl. Kap. 3.3 und 3.4.4). Alle 30 Sitzungen wurden in Form von Audio- und Videoaufnahmen sowie ad hoc Notizen zu eventuellen Besonderheiten der Datenerhebungssituation dokumentiert. Sie wurden später von drei studentischen Hilfskräften transkribiert und im gleichen Zuge anonymisiert. Die drei Fremdsprachendidaktik-Studentinnen wurden zuvor im Rahmen eines Trainings mit den Transkriptionsrichtlinien bekannt gemacht. Als solche dienten die leicht angepassten Regeln des GAT 2-Basistranskripts (Selting et al., 2009; vgl. auch Schmidt et al., 2015). Die Transkripte wurden mithilfe des Programms MAXQDA 2018 angefertigt (VERBI Software, 2017), was die synchrone Arbeit mit der Videoaufnahme während des Transkriptions- und später auch

⁷⁵ „Wurden die Daten fixiert, z.B. bei Interviews in Form von Audio- oder Videoaufnahme? Wurde eine interviewbegleitende Dokumentation (Postskriptum) erstellt, in dem die Interviewsituation und Besonderheiten festgehalten wurden? Wann wurde das Postskriptum erstellt? Wurde eine vollständige Transkription des Interviews vorgenommen? Wurden Transkriptionsregeln benutzt und werden diese offengelegt? Wie sah der Transkriptionsprozess konkret aus? Wer hat transkribiert? Die Forschenden selbst? Wurde eine Transkriptionssoftware benutzt? Wurden die Daten anonymisiert? In welcher Weise? Ist das synchrone Arbeiten mit Audio-Aufnahme und Transkription möglich? Wurden die Transkriptionsregeln eingehalten und entspricht die verschriftlichte Fassung dem Gesagten?“ (Kuckartz, 2018, S. 204).

⁷⁶ „Ist die gewählte inhaltsanalytische Methode der Fragestellung angemessen? Wird die Wahl der Methode begründet? Wenn ja, wie? Wurde das jeweilige Verfahren in sich richtig angewendet? Wurde die Inhaltsanalyse computergestützt durchgeführt? Wurde das Material oder Teile desselben durch mehrere Codierende unabhängig voneinander bearbeitet? Wie wurde die Übereinstimmung der Codierenden ermittelt? Welches Vorgehen wurde bei Nicht-Übereinstimmung gewählt? Ist das Kategoriensystem in sich konsistent? Sind die Kategorien und Subkategorien gut ausgearbeitet? Wie präzise und ausführlich sind die Kategoriendefinitionen? Gibt es konkrete Beispiele (Zitate) als Illustration für die Bedeutung der Kategorien? Wurden alle erhobenen Daten bei der qualitativen Inhaltsanalyse berücksichtigt? Wie oft wurde das Material bis zur endgültigen Codierung durchlaufen? Ist Auditierbarkeit, d.h. unter anderem Nachvollziehbarkeit der Codierungen, gegeben? Wurden auch abweichende Fälle berücksichtigt? Wird auf Ausnahmefälle und Extremfälle hingewiesen und werden diese analysiert? Wurden im Verlauf der Inhaltsanalyse Memos geschrieben? Wann? Wie sehen beispielhafte Memos aus? Wurde mit Originalzitaten gearbeitet und nach welchen Kriterien wurden diese gewählt? Wurden nur Techniken selektiver Plausibilisierung angewendet oder wurde auch auf Gegenbeispiele und Widersprüche hingewiesen? Sind die gezogenen Schlussfolgerungen jeweils in den Daten begründet? Was wurde wie und in welcher Form dokumentiert und archiviert?“ (Kuckartz, 2018, S. 204–205).

während des Codierprozesses ermöglichte. Alle Transkripte wurden anschliessend von der Forscherin kontrolliert und wo nötig verfeinert. Aufgrund des Umfangs der Transkripte und einiger Probleme mit der Validität bestimmter Datensätze (z.B. bei wenigen oder keinen Verbalisierungen der Schüler/innen beim Lauten Denken) wurde das zu analysierende Korpus auf 20 Transkripte reduziert. In diesen Transkripten wurden jeweils Äusserungen der Schüler/innen zu ausgewählten Aufgaben analysiert. Die Auswahl der Transkripte und Aufgaben für die vertiefende Analyse wurde kriteriengeleitet getroffen und in Kap. 3.6 ausgewiesen.

Im Hinblick auf die Beantwortung der Forschungsfragen wurden die Datenauswertungsmethoden der inhaltlich strukturierenden und der evaluativen qualitativen Inhaltsanalyse gewählt, was in Kap. 3.7.1 begründet wurde. Die Analyse wurde computergestützt in MAXQDA 2020 (VERBI Software, 2019) durchgeführt. Die Festlegung der Kategorien für die Analyse wurde deduktiv, d.h. aufgrund der theoretischen Überlegungen und ähnlicher Studien aus dem Bereich der Testforschung, wie auch datengeleitet vorgenommen (vgl. Kap. 3.7.3.2) und die Vorgehensweise bei der Erarbeitung des Kategoriensystems und beim Codieren des Materials wurde detailliert dargelegt. Die Daten wurden ausschliesslich von der Studienautorin codiert, was nicht untypisch für eine Qualifikationsstudie mit limitierten Ressourcen ist (vgl. z.B. Kuckartz, 2018, S. 105). Dies bringt jedoch bestimmte Vor- und Nachteile mit sich, auf die hier etwas ausführlicher eingegangen werden soll.

Die qualitative Inhaltsanalyse ist ein hermeneutisches Verfahren und kann, wie Kuckartz anmerkt, als solches nicht ohne den *Verstehenden* existieren, „der immer schon ein Vorverständnis über den Gegenstand des Verstehens (...) besitzt“ (2018, S. 20). Für eine hermeneutische Deutung könne daher keine richtige oder falsche, sondern lediglich mehr oder weniger angemessene Interpretation postuliert werden (ebd.). Für die Arbeit nach den Prinzipien der qualitativen Inhaltsanalyse ist daher „unbedingt ein gewisses Mass an Interpretationskompetenz erforderlich, d.h. die Codierer müssen über die Fragestellung, die theoretischen Konstrukte und die Bedeutung der Kategorien gut informiert werden“ (ebd., S. 44). Es hat sich deswegen für die vorliegende Studie als vorteilhaft erwiesen, dass sie von der gleichen Forscherin geplant wurde, die die Daten später auch erhoben und mit dem ebenfalls von ihr ausgearbeiteten Codiersystem ausgewertet hat. Gleichzeitig führte es aber dazu, dass der Forschungsprozess als subjektiv zu betrachten ist, was wiederum reflektiert werden muss (vgl. u.a. Demirkaya, 2014; Galaczi, 2014).

Die Subjektivität bei der Analyse von qualitativen Daten lässt sich gemäss Holliday (2015) nicht vermeiden. Vielmehr werde aber die Validität und Glaubwürdigkeit der qualitativen Forschung dadurch bestimmt, wie mit dieser Subjektivität in einer Studie umgegangen wird.

Dies hängt laut Holliday (ebd.) wiederum von drei Prinzipien ab. Das erste Prinzip betrifft die transparente Beschreibung des Forschungsprozesses (*transparency of method*), was in der vorliegenden Studie stets angestrebt wurde. Zur besseren Nachvollziehbarkeit der Vorgehensweise bei der Datenauswertung und der Resultate soll auch die umfangreiche und gut zugängliche Dokumentation beitragen: Der vollständige Kategorienleitfaden kann im Anhang 5 nachgeschlagen und alle codierten (und uncodierten) Transkripte können von Interessierten im Forschungsdatenarchiv des KFM⁷⁷ konsultiert werden. Das zweite Prinzip, das Holliday (ebd.) als *submission to the data* bezeichnet, betrifft die Offenheit der Forschenden gegenüber unerwarteten Phänomenen, die in den Daten gefunden werden und die ursprüngliche Forschungsrichtung evtl. auch verändern können. Diesem Prinzip wurde bei der Analyse der Daten entsprochen, indem die Kategorien nicht nur deduktiv, sondern auch induktiv erstellt wurden. Mit induktiv erstellten Kategorien konnten folglich auch neue Phänomene in den Daten erfasst werden. Das dritte Prinzip betrifft die Angemessenheit der Behauptungen, die aufgrund der qualitativen Forschung formuliert werden (*making appropriate claims*). Das Ziel der qualitativen Forschung bestehe laut Holliday (ebd.), wie seinem Zitat am Anfang der vorliegenden Studie zu entnehmen ist, nicht darin, etwas zu beweisen oder die Resultate von anderen Studien zu bestätigen bzw. zu widerlegen. Vielmehr gehe es hier um den Gewinn von neuen Informationen, die zum weiteren Nachdenken über die beforschten Phänomene anregen sollten. Bei der Darstellung und Diskussion der Resultate in Kap. 5 wird darauf geachtet, dieses zentrale Prinzip einzuhalten.

Die Subjektivität beim Interpretieren der Lautdenkprotokolle und Codieren der Strategien hat laut Y. Gu (2014, S. 78–79) auch damit zu tun, dass dies typischerweise aus der Perspektive der Forschenden und weniger aus der Perspektive der Beforschten geschehe. Y. Gu spricht sich daher dafür aus, bei der Codierung der Strategien auch die Perspektive der Lernenden im Auge zu behalten. Diesem Postulat wurde in der vorliegenden Studie bereits bei der Datenerhebung entsprochen, indem neben der Erhebung von Lautdenkprotokollen auch ergänzende *Stimulated-Recall*-Interviews durchgeführt wurden. Dank diesen Interviews konnte die Perspektive der Lernenden besser erfasst und bei der Datenanalyse mitberücksichtigt werden.

Bei der qualitativen Inhaltsanalyse handelt es sich um ein Auswertungsverfahren, das auf unterschiedliche Arten von Daten angewandt werden kann (Kuckartz, 2018), darunter auf Daten, die mit introspektiven Methoden wie Lautes Denken oder *Stimulated Recall* erhoben wurden (vgl. dazu Beiträge im Sammelband von Aguado et al., 2013). Generell empfiehlt auch

⁷⁷ <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000012367>

Kuckartz (2018, S. 105), Daten von mehreren Codierenden auswerten zu lassen, die unabhängig voneinander arbeiten sollten. Es geht aber bei der qualitativen Inhaltsanalyse nicht um die Ermittlung der aus den quantitativen Studien bekannten Codiererübereinstimmung, sondern um die anschließende Diskussion und Konsensfindung zwischen den Codierenden (auch als „konsensuelles Codieren“ bekannt). Diese Vorgehensweise ermögliche präzisere Kategoriendefinitionen und eine zuverlässigere Zuordnung der Daten zu Kategorien. Im Fall der vorliegenden Studie war der Einbezug von weiteren Codierenden, wie bereits erwähnt, aufgrund der beschränkten Ressourcen nicht möglich. Obwohl die von Kuckartz vertretene Position, die qualitativen Daten möglichst immer durch mehrere Codierende auswerten zu lassen, durchaus einleuchtet, deuten die Ausführungen von Y. Gu (2014) darauf hin, dass sich die Codierung der Lautdenkprotokolle durch andere, am Projekt nicht intensiv beteiligte Personen als schwierig bzw. sogar kontraproduktiv gestalten könnte. In der Studie von Y. Gu, Hu und Zhang (2005) wurde in den Lautdenkprotokollen von Primarschüler/innen aus Singapur nach Hinweisen zu ihrem Strategieeinsatz beim Hörverstehen fremdsprachlicher Texte gesucht. Beim Codieren der Daten haben Y. Gu et al. versucht, auf die Unterstützung einer Forschungsassistentin zurückzugreifen. Dies erwies sich aber nicht als zielführend, wie Y. Gu (2014, S. 76–77) berichtet:

[C]oding was such an intensely iterative and dynamic process of subjective meaning-making that we found it hard to involve research assistants. We did involve one research assistant for the coding of our data at the last stage, after we had already finalised the coding scheme. However, we had to abandon the data coded by this research assistant in the final analysis, because her interpretations were considerably different from ours even after extensive training. A number of researchers (Bong, 2002; Humble, 2012) have emphasised the agency, thoughtfulness and reflexivity on the part of coding researchers. Our experiences lend support to these observations, in that the iterative and reflexive processes of coding and analysis reflect so much of the researcher, and determine what is coded, what is not, how things are categorised, and what insights and relationships are seen among those categories, so much so that simply looking at the code patterns provided by research assistants would become almost meaningless.

Aus diesem Grund wird es von Y. Gu (2014, S. 80) auch nicht empfohlen, das Codieren von Lautdenkprotokollen Forschungsassistent/innen zu überlassen.

In den Fällen, wo eine Codierung durch mehrere Personen nicht möglich ist, wird man laut Kuckartz (2018, S. 105) „nicht umhin kommen, sich mit diesem Mangel zu arrangieren und

selbst darauf zu achten, bei Zweifelsfällen die expliziten Kategoriendefinitionen zu verbessern und konkrete Beispiele festzuhalten“. Burwitz-Melzer und Steininger empfehlen hierfür „entweder Kollegen und Kolleginnen zu Rate [zu] ziehen oder besonders sorgfältig [vorzugehen] und mehrere Korrekturschleifen [einzuplanen]“ (2016, S. 260). In der vorliegenden Studie wurde beides unternommen: Es wurden zum einen im Sinne der *Intra-Rater*-Reliabilität (vgl. Knorr & Schramm, 2016, S. 93) insgesamt sechs Codierdurchgänge des Materials vorgenommen, was jeweils zur Verfeinerung des Kategorienleitfadens führte. Um Hinweise zur möglichen Präzisierung der Kategorien und Zuverlässigkeit der Codierungen zu bekommen, wurde zum anderen nach der Erprobungsphase des Kategoriensystems auch eine Sitzung mit zwei Forschenden organisiert. Diese waren mit den Forschungsfragen und dem Forschungsdesign gut vertraut, machten sich vor der Datensitzung auch mit dem Kategorienleitfaden bekannt und versuchten, die bereits codierten Transkriptauszüge nachzuvollziehen. Ihre Anmerkungen waren sehr hilfreich und trugen wesentlich zur Verbesserung der Konsistenz und der intersubjektiven Nachvollziehbarkeit des Kategoriensystems bei (vgl. Kap. 3.7.2).

3.8 ZUSAMMENFASSUNG

In diesem Teil wurde das Vorgehen bei der empirischen Untersuchung der Strategien, die Schüler/innen beim Lösen der Testaufgaben aus dem Projekt IFB einsetzten, detailliert dargestellt und begründet. Zuerst wurde der Forschungskontext beschrieben, d.h. die Datenerhebung im Projekt IFB und die Testaufgaben, die in diesem Projekt entwickelt wurden. Das Projekt IFB hatte u.a. zum Ziel, prototypische, szenariobasierte und am Computer lösbare Testaufgaben zum Lese- und Hörverstehen in den Fremdsprachen Englisch und Französisch zu entwickeln und zu erproben, die evtl. künftig im Rahmen eines Bildungsmonitorings eingesetzt werden könnten. In der vorliegenden Studie wurden ausgewählte Aufgaben aus dem Projekt IFB vertieft untersucht.

Anschliessend wurden in diesem Teil die Forschungsfragen, das Forschungsdesign und die eingesetzten Datenerhebungsmethoden erläutert. Um zu untersuchen, 1) welche Strategien bei Schüler/innen beim Lösen der IFB-Testaufgaben beobachtet werden können, 2) ob sich zwischen konstruktirrelevanten Strategien und erfolgreichen bzw. nicht erfolgreichen Aufgabenlösungen Zusammenhänge erkennen lassen sowie 3) ob sich Unterschiede beim Einsatz der Strategien zeigen, wenn szenariobasierte und nicht szenariobasierte Testaufgaben gelöst werden, und, falls ja, worauf dies zurückgeführt werden könnte, wurde in diesem Projekt auf zwei introspektive Datenerhebungsmethoden zurückgegriffen: Lautes Denken und *Stimulated Recall*. Die Vorgehensweise bei der Datenerhebung wurde zuerst in individuellen Sitzungen mit 6 Schüler/innen pilotiert (Dauer: 3 Lektionen à 45 Minuten pro

Schüler/in). Während der Haupterhebung wurden die Daten von insgesamt 30 Lernenden in individuellen Sitzungen erhoben, die ebenfalls je 3 Lektionen à 45 Minuten dauerten. Etwa die Hälfte der Schüler/innen bearbeitete dabei französische Versionen der IFB-Testaufgaben, die in das Szenario „Stadttrip“ eingebettet waren (n=17), währenddessen die restlichen Schüler/innen (n=13) vergleichbare Aufgaben aus dem Projekt IFB lösten, die verschiedenen Szenarien entnommen waren. Dies wurde gemacht, um Indizien zum eventuellen Einfluss der Szenarioeinbettung der Testaufgaben auf die Vorgehensweise der Schüler/innen bei ihrer Bearbeitung gewinnen zu können. Die erhobenen Daten wurden anschliessend transkribiert und aufbereitet. Da sie sehr umfangreich ausfielen, wurde vor der Analyse eine kriterienbasierte Auswahl der vertieft zu analysierenden Daten getroffen, was in Kap. 3.6 geschildert wurde. Es wurde entschieden, die Verbalprotokolle von 20 Schüler/innen zu ausgewählten Aufgabenteilen (d.h. Bildschirmseiten) zu analysieren.

Das letzte, umfangreichste Kap. 3.7 ist der Datenauswertung gewidmet. Hier wurden die bei der Analyse der vorliegenden Daten eingesetzten Datenauswertungsverfahren vorgestellt: Die inhaltlich strukturierende und die evaluierende qualitative Inhaltsanalyse nach Kuckartz (2018). Es wurde weiter aufgezeigt, wie diese Verfahren in der vorliegenden Studie operationalisiert wurden. Auch wurde hier das Kategoriensystem präsentiert, das im Rahmen dieser Untersuchung entwickelt und eingesetzt wurde. Die Codierung der Strategien wurde hier nach dem Klassifizierungsvorschlag der testbezogenen Strategien von A. D. Cohen vorgenommen, d.h. die in den Verbalprotokollen beobachteten Strategien wurden jeweils als konstruktrelevante Lernerstrategien, konstruktrelevante Testmanagementstrategien oder konstruktirrelevante *test-wiseness*-Strategien gekennzeichnet.

Im letzten Teil dieses Kapitels wurde schliesslich die Vorgehensweise bei der Datenauswertung kritisch reflektiert: Dort wurde zusammenfassend aufgezeigt, was in der vorliegenden Arbeit durch transparente und ausführliche Darstellung der erhobenen Daten und der Vorgehensweise bei ihrer Auswertung unternommen wurde, um den Gütekriterien der qualitativen Inhaltsanalyse gerecht zu werden.

4 ERGEBNISSE

Im Sinne der qualitativen Forschung widmet sich ein wichtiger Teil dieses Kapitels den inhaltlichen Ergebnissen der kategorienbasierten Datenauswertung, die in erster Linie mit einem Fokus auf die inhaltliche Bedeutung und Interpretation der Daten präsentiert und illustriert werden. Die Datengrundlage ist sehr umfassend. Sie beinhaltet: 1) die Transkripte der Verbalprotokolle von 20 für die vertiefende Analyse ausgewählten Schüler/innen beim Lösen der Testaufgaben, 2) die Transkripte der *Stimulated-Recall*-Interviews, die mit diesen Schüler/innen im Anschluss an den Test durchgeführt wurden, sowie 3) die Testergebnisse dieser 20 Schüler/innen, die mit den Testergebnissen der IFB-Stichprobe für Französisch in Bezug gebracht werden.

Seale postuliert, dass es sich auch in der qualitativen Forschung lohnt, das Zählbare zu zählen, denn „[c]ounting events that are well defined and illustrated can increase the credibility of claims made by qualitative researchers, by guarding against the charge of anecdotalism“ (1999, S. 121). Aus diesem Grund wird in der Darstellung der Ergebnisse auch auf die Anzahl der codierten Segmente eingegangen. Bei der Rezeption dieser quantitativen Angaben ist aber stets zu beachten, dass den Daten – wie bei qualitativen Studien üblich – eine kleine Stichprobe zugrunde liegt. Vor allem aber muss dabei die Natur der erhobenen Daten im Blick behalten werden: In den Lautdenkprotokollen und in den *Stimulated-Recall*-Interviews konnte nur das codiert und gezählt werden, was die Schüler/innen verbalisierten bzw. bei einigen Kategorien das, was von der Forscherin aufgrund des Videos und des Transkripts beobachtet werden konnte. Hinzu kommt, dass die Schüler/innen unterschiedlich mitteilungsfreudig waren: Einige verbalisierten viel und gern, andere wiederum eher sparsam. Es ist ferner davon auszugehen, dass die Schüler/innen nicht alle Gedanken laut aussprachen, die ihnen bei der Bearbeitung der Aufgaben und im Verlauf des Interviews durch den Kopf gingen, z.B. wegen der Parallelität und/oder Automatisierung bestimmter Denkprozesse, oder auch weil ihnen nicht alle Gedanken in verbaler Form vorlagen (vgl. Heine & Schramm, 2016; Knorr & Schramm, 2012). Auch kann nicht ausgeschlossen werden, dass eine andere Codiererin oder ein anderer Codierer zu einer anderen Anzahl codierter Segmente kommen würde. Die Zahlen, die im Folgenden im Zusammenhang mit den Codierungen dargestellt werden, sind daher mit Vorsicht zu interpretieren: Sie dienen hier ausschliesslich dazu, „die in den Daten vorhandenen Informationen voll auszuschöpfen und hierzu gehört auch die Information, ob ein Phänomen oder ein Zusammenhang von Phänomenen häufig vorkommt oder nur als Einzelfall aufgetreten ist“ (Kuckartz, 2018, S. 116; vgl. auch Y. Gu, 2014).

Bevor in Kap. 4.2 bis 4.4 über die Ergebnisse berichtet wird, die in Bezug auf die Forschungsfragen ermittelt wurden, wird in Kap. 4.1 dargestellt, welche Resultate die

20 Lernenden, deren Testbewältigungsstrategien vertieft analysiert wurden, bei den untersuchten Testaufgaben erzielten. Zusätzlich werden die Resultate der 20 Lernenden mit den empirischen Itemschwierigkeitswerten aus der quantitativen Hauptstudie IFB abgeglichen. Die Ergebnisse werden in Kap. 5 zusammengefasst und diskutiert.

4.1 AUSWERTUNG DER UNTERSUCHTEN TESTAUFGABEN

Die beiden untenstehenden Tabellen (Tabelle 12 und Tabelle 13) zeigen die Testresultate der Testteilnehmenden zu den ausgewählten Testaufgaben. Die Vorgehensweise beim Lösen dieser Aufgaben wurde in den Transkripten von 10 ausgewählten Szenario- und 10 ausgewählten Off-Szenario-Schüler/innen untersucht (vgl. Kap. 3.6). Die Szenario-Schüler/innen konnten in diesen Aufgaben maximal 12 und die Off-Szenario-Schüler/innen maximal 10 korrekte Antworten geben. In den Tabellen wurden die richtigen Antworten mit 1 und die falschen mit 0 erfasst (vgl. Kap. 3.7.3.1).

Neben der Anzahl richtiger Antworten pro Testteilnehmer/in (Zeilen) und pro Testitem (Spalten) wird auch die empirische Itemschwierigkeit (*Item Difficulty*) ausgewiesen. Die Itemschwierigkeit wurde im quantitativen Teil des Projekts IFB aufgrund der Resultate aller Schüler/innen berechnet, die die französische Version dieser Aufgaben lösten ($n=303$). Die Itemschwierigkeit wurde in den bereits in Kap. 3.4.3.2 vorgestellten 2-PL-Modellen der IRT (Birnbaum, 1968) geschätzt, wobei Hörverstehens- und Leseverstehensaufgaben getrennt modelliert wurden. Werte um 0 situieren die Schwierigkeit der Aufgaben in der Mitte der getesteten Stichprobe, negative Werte im Bereich der einfacheren und positive im Bereich der schwierigeren Aufgaben. Items mit einem Wert um 0 konnten durchschnittlich gute Schüler/innen aus der Stichprobe also mit einer Wahrscheinlichkeit von 50% richtig oder falsch lösen. Itemschwierigkeiten über 0.5 bedeuten, dass durchschnittliche Lernende diese Items mit einer Wahrscheinlichkeit von 38% lösen konnten und Itemschwierigkeiten über 2 bedeuten, dass diese Items sehr schwierig waren und durchschnittliche Lernende sie nur noch mit einer Wahrscheinlichkeit von 12% und weniger lösen konnten. Itemschwierigkeiten unter -0.5 konnten durchschnittliche Lernende hingegen mit einer Wahrscheinlichkeit von 62% und unter -2 mit einer Wahrscheinlichkeit von 88% und mehr korrekt lösen.

Für das Item „Stadtzentrum“ (Szenario „Stadttrip“, Task C, Seite 2) wird keine Itemschwierigkeit angegeben, da dieses Item bei der Berechnung des Modells entfernt wurde. Grund dafür war, dass das Item schlecht differenzierte, d.h. starke wie schwache Lernende ähnliche Chancen hatten, das Item korrekt zu lösen. Dies verstößt gegen die Annahmen des psychometrischen Modells.

Insgesamt ist zu beachten, dass die empirischen Itemschwierigkeiten bei den Off-Szenario-Aufgaben für die originalen Aufgabenversionen berechnet wurden, die in Szenarien eingebettet waren. Zwar wurden an den Items selbst kaum Veränderungen vorgenommen, die Vorinformationen, die sich aus dem Szenario für die Lernenden ergaben, unterschieden sich aber geringfügig.

Tabelle 12: Übersicht über richtige und falsche Antworten der Szenario-Schüler/innen zu den Tasks aus den vertiefenden Analysen (inkl. empirische Itemschwierigkeit aus der quantitativen Hauptstudie).

Szenario „Stadttrip“													
Aufgabe	Task A	Task A	Task A	Task A	Task B	Task B	Task C	Task C	Task C	Task C	Task C	Task C	Anzahl korrekter Antworten pro Schüler/in (Max. 12)
Bildschirm-seite	Seite 1	Seite 1	Seite 2	Seite 2	Seite 1	Seite 2	Seite 1	Seite 1	Seite 1	Seite 2	Seite 2	Seite 2	
Item	Rugby-Ausstellung	Coq Rouge	Carte Avignon	Stadt-rund-gang	Internet	beein-druckt in Avignon	WLAN	Stadt-zentrum	preis-wert	WLAN	Stadt-zentrum	preis-wert	
Alison	0	1	1	1	0	0	0	1	1	1	1	1	8
Aria	1	1	1	0	0	0	1	1	1	1	0	1	8
Arya	1	1	0	1	0	0	0	1	1	1	1	1	8
Cersei	1	1	0	0	1	0	0	0	1	0	0	1	5
Elena	1	0	0	1	1	0	0	0	1	0	0	0	4
Böbu	0	1	1	0	0	0	0	0	0	1	0	0	3
Cyberman	1	1	1	1	0	0	1	1	1	1	0	1	9
Howard	1	1	0	1	1	0	1	0	1	1	1	1	9
Sheldon	0	1	1	0	1	0	1	1	1	1	0	1	8
Toby	1	1	1	1	1	1	1	1	1	1	1	1	12
Anzahl korrekter Antworten pro Item (Max. 10)	7	9	6	6	5	1	5	6	9	8	4	8	
Item Difficulty (2-PL) ⁷⁸	-0.409	-1.328	-0.608	-0.464	-0.11	1.328	0.042	0.948	-1.002	0.731	(kein Wert)	-0.037	

⁷⁸ Einfache Items mit einem Wert unter -0.5 wurden in den letzten Zeilen der Tabelle grün, schwierige (ab 0.5) rot und die mittelschweren grau markiert.

Tabelle 13: Übersicht über richtige und falsche Antworten der Off-Szenario-Schüler/innen zu den Tasks aus den vertiefenden Analysen (inkl. empirische Itemschwierigkeit aus der quantitativen Hauptstudie).

Szenario	"Walkies"		"Stadttrip"		"Berufswahl"						Anzahl korrekter Antworten pro Schüler/in (Max. 10)
Aufgabe	Task C	Task C	Task B	Task B	Task C	Task C	Task C	Task C	Task C	Task C	
Bildschirm-seite	Seite 1	Seite 2	Seite 1	Seite 2	Seite 1	Seite 1	Seite 1	Seite 2	Seite 2	Seite 2	
Item	<i>deine Aufgabe</i>	<i>Esther</i>	<i>Internet</i>	<i>beein-druckt in Avignon</i>	<i>selbst Produkte testen</i>	<i>mehrere Sprachen sprechen</i>	<i>viel reisen</i>	<i>selbst Produkte testen</i>	<i>mehrere Sprachen sprechen</i>	<i>viel reisen</i>	
Bonnie	1	0	1	1	1	1	0	0	1	1	7
Dustin	1	1	1	0	1	0	0	0	1	0	5
Loup	0	0	0	0	1	0	0	0	1	1	3
Nancy	1	0	1	0	0	0	0	0	1	1	4
Picard	1	0	0	0	0	0	0	0	0	0	1
Eleven	0	0	0	0	0	0	1	0	0	1	2
Ezra	1	0	1	1	0	0	0	1	1	0	5
Mogli	0	0	1	0	1	0	0	0	1	0	3
Omega	1	0	1	0	1	0	1	0	1	1	6
Steve	0	0	0	1	0	0	0	0	0	0	1
Anzahl korrekter Antworten pro Item (Max. 10)	6	1	6	3	5	1	2	1	7	5	
Item Difficulty (2-PL) ⁷⁹	1.545	1.439	-0.11	1.328	0.843	1.858	1.51	1.254	-0.113	1.239	

⁷⁹ Die schwierigen Items ab dem Wert 0.5 wurden in den letzten Zeilen der Tabelle rot und die mittelschweren grau markiert.

In Tabelle 12 sieht man, dass die Szenario-Schüler/innen die Testaufgaben unterschiedlich gut bewältigten: Während Tobi alle 12 Items richtig löste, hatte Böbu nur 3, Elena nur 4 und Cersei nur 5 Antworten richtig. Die Mehrheit der Szenario-Schüler/innen war aber beim Bearbeiten der Aufgaben recht erfolgreich (Alison, Aria, Arya, Sheldon mit je 8 und Cyberman und Howard mit je 9 von 12 Punkten). Die einfachsten Items für diese 10 Szenario-Schüler/innen waren die Items „Coq Rouge“ und „preiswert“ (Seite 1), „WLAN“ (Seite 2), „preiswert“ (Seite 2) sowie „Rugby-Ausstellung“. Der Vergleich dieser Items mit den Itemschwierigkeitswerten aus der quantitativen Datenerhebung zeigt, dass die Items „Coq Rouge“ sowie „preiswert“ (Seite 1) auch für alle Lernenden der Französischstichprobe im Projekt IFB sehr einfach waren. Das Item „preiswert“ (Seite 2) war in der Hauptstudie hingegen mittelschwer und das Item „WLAN“ (Seite 2) war für die Schüler/innen der IFB-Französischstichprobe sogar eher schwierig.

Auch die Off-Szenario-Schüler/innen (vgl. Tabelle 13) lösten die Testaufgaben mit unterschiedlich gutem Resultat. Auffallend ist, dass keine Schülerin bzw. kein Schüler alle Items richtig löste und die besten Lernenden hier lediglich 7 (Bonnie), 6 (Omega) oder 5 (Dustin, Ezra) der 10 möglichen Punkte erreichten. Die schwächsten Lernenden konnten hier nur 1 (Steve, Picard) oder 2 Punkte (Eleven) holen. Zu den einfachsten Items gehörten für diese Schüler/innen-Gruppe „mehrere Sprachen sprechen“ (Seite 2) mit 7 korrekten Antworten und „deine Aufgabe“ und „Internet“ mit je 6 korrekten Antworten. Die richtige Lösung zu den Items „mehrere Sprachen sprechen“ (Seite 1), „Esther“ und „selbst Produkte testen“ (Seite 2) konnte hingegen von jeweils nur je einer Schülerin oder einem Schüler gefunden werden. Wenn man die Schwierigkeitswerte der 10 von den Off-Szenario-Schüler/innen gelösten Items betrachtet, fällt sofort auf, dass die Mehrheit dieser Items für alle Schüler/innen der IFB-Französischstichprobe sehr schwierig war: 7 der 10 Items hatten einen Schwierigkeitswert von mehr als 1, ein weiteres einen Wert von 0.843. Nur zwei Items befinden sich eher im mittleren Bereich der Itemschwierigkeit („Internet“ und „mehrere Sprachen sprechen“, Seite 2). Diese beiden Items wurden auch von den Lernenden in der qualitativen Studie sichtlich häufiger korrekt gelöst.

Es zeigt sich also, dass die Szenario-Schüler/innen mehr einfache und sehr einfache Items als die Off-Szenario-Schüler/innen lösen mussten, was bei der Auswahl der Aufgaben nicht beabsichtigt war. Diese wichtige Erkenntnis, die erst nach der Datenerhebung gewonnen werden konnte, muss bei der Interpretation der Resultate der vorliegenden Studie mitberücksichtigt werden. Es kann nämlich nicht ausgeschlossen werden, dass die unterschiedliche Itemschwierigkeit der Szenario- und der Off-Szenario-Items einen Einfluss auf die Testbewältigungsstrategien der hier untersuchten Schüler/innen gehabt haben könnte.

4.2 ZU FORSCHUNGSFRAGE 1: STRATEGIEN DER SCHÜLER/INNEN BEIM LÖSEN DER TESTAUFGABEN

Die erste Forschungsfrage sollte Antworten darauf liefern, welche Strategien bei Schüler/innen beim Lösen der IFB-Testaufgaben beobachtet werden können. In diesem Teil werden die Strategien dargestellt, die in den Transkripten der 20 Lernenden gefunden und codiert wurden. Auch wenn der Fokus dieser Studie nicht primär auf der Quantifizierung der Codierungen liegt, wird im Weiteren zuerst über die (wie erwähnt mit Vorsicht zu interpretierende) Häufigkeit der vergebenen Codes berichtet. Ziel ist es, jeweils zuerst einen schnellen Überblick über alle codierten Strategien zu ermöglichen. In einem weiteren Schritt werden die einzelnen Strategien dann aus dem qualitativen Blickwinkel ausführlicher beleuchtet.

Insgesamt wurden in den Lautdenkprotokollen 1333 Strategien und in den *Stimulated-Recall*-Interviews 1210 Strategien codiert. Nach Abzug der speziellen Codierungen mit Gewichtung „2“ (mit der bei allen Kategorien wiederholte Erwähnungen des gleichen Phänomens an einer anderen Stelle im Transkript bei der Bearbeitung einer Bildschirmseite gekennzeichnet wurden) sowie der Codierungen mit der Gewichtung „25“ (d.h. der Codierungen der Kategorie „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ bei der mündlichen Zusammenfassung der Inputtexte in *Stimulated-Recall*-Interviews) beträgt die Anzahl der codierten Strategien im Lautdenken-Teil 1290 und in den *Stimulated-Recall*-Interviews 945. Der Abzug der Codierungen mit den zwei speziellen Gewichtungen wurde vorgenommen, um ein realistischeres Bild der Häufigkeiten der vorkommenden Strategien zu gewinnen. Bei der Besprechung der Häufigkeit der Codierungen von einzelnen Strategien wird im weiteren Teil dieser Arbeit überall dort, wo es nicht anders gekennzeichnet wurde, auf die Zahlen nach Abzug der Gewichtungen „2“ und „25“ Bezug genommen.

Der Vergleich der Anzahl Codierungen der gleichen Strategie im Lautdenken- und *Stimulated-Recall*-Untersuchungsteil zeigt teilweise erhebliche Abweichungen zwischen den beiden Untersuchungsteilen. So wurde z.B. die Strategie „Intensive Auseinandersetzung mit Teilen des Inputtextes“ in den Lautdenkprotokollen 124 Mal und in den *Stimulated-Recall*-Interviews lediglich 14 Mal codiert. Diese Diskrepanzen sind in erster Linie auf die unterschiedlichen Eigenschaften der erhobenen Daten zurückzuführen: Die Lautdenkprotokolle und die darin vorgenommenen Codierungen widerspiegeln gewisse Strategien, die bei der Bewältigung der Testaufgaben eingesetzt wurden, viel genauer als die entsprechenden Äusserungen der Lernenden im anschließenden *Stimulated-Recall*-Interview. Wenn beispielsweise eine Schülerin bei der Rezeption einer Aufgabe/Bildschirmseite sich mehrmals an

unterschiedlichen Stellen mit den Teilen des Inputtextes intensiv auseinandersetzte, wurde die Strategie „Intensive Auseinandersetzung mit Teilen des Inputtextes“ im Lautdenken-Teil auch mehrmals codiert. Im *Stimulated-Recall*-Interview wurde hingegen über diese Strategie von der gleichen Schülerin bei der Besprechung der gleichen Aufgabe nur 1 Mal berichtet und die Strategie wurde dort deswegen auch nur 1 Mal codiert. Bestimmte Strategien wurden darüber hinaus nur in einem Untersuchungsteil ermittelt, so z.B. die Strategie „Prosodie beim Hören nutzen“. Sie wurde in keinem der Lautdenkprotokolle, dafür aber in den *Stimulated-Recall*-Interviews von zwei Lernenden erwähnt und codiert. Dass diese Strategie beim Lauten Denken unerwähnt blieb, hatte laut einer dieser zwei Lernenden damit zu tun, dass sie diese Strategie gleichzeitig mit dem Hören des Inputtextes angewandt hatte, während sie schnell an mehrere Sachen gleichzeitig gedacht hatte, und deswegen nichts dazu verbalisierte.

Nachfolgend werden die Ergebnisse zu den Codierungen der Lerner- und Teststrategien (d.h. der Testmanagement- und *test-wiseness*-Strategien) vorgestellt. Um eine bessere Übersicht über die vorgenommenen Codierungen pro Strategie zu ermöglichen, wird am Anfang jedes Unterkapitels zunächst in tabellarischer Form über die Häufigkeiten der strategienbezogenen Codes der jeweiligen Strategiegruppe pro Untersuchungsteil (Lautes Denken und *Stimulated Recall*) berichtet. Den Tabellen ist jeweils die Anzahl der speziell festgelegten Gewichtungen pro Strategie zu entnehmen, die bei der Interpretation der Zahlen zu berücksichtigen sind. Dies betrifft besonders die bereits erwähnten Gewichtungen „2“ (wiederholte Erwähnungen des gleichen Phänomens an einer anderen Stelle im Transkript bei der Bearbeitung einer Bildschirmseite) und „25“ (Codierungen bei der mündlichen Zusammenfassung der Inputtexte bei der Kategorie „SR-L10: Unbekannte Wörter, Satz- bzw. Textteile identifizieren“), die von der Anzahl der total codierten Strategien überall dort abgezogen und in einer speziell dafür vorgesehenen Spalte aufgelistet wurden, wo sie auch vorkamen. Auch wird in den Tabellen die Anzahl der Codierungen mit Gewichtung „5“ aufgeführt, mit der allgemeine(re) Aussagen der Schüler/innen zu ihrem Verhalten beim Lösen der Testaufgaben erfasst wurden. Solche Aussagen waren aufgaben-, item- oder textstellenübergreifend und betrafen nicht (nur) das gerade besprochene Item oder die betroffene Aufgabe oder Textstelle. Die Codierungen mit dieser Gewichtung wurden von der Gesamtanzahl der codierten Strategien nicht abgezogen.

Im Anschluss an die tabellarische Übersicht werden die codierten Strategien aus qualitativer Sicht im Detail dargestellt (vgl. dazu auch den Kategorienleitfaden im Anhang 5). Für diesen Zweck wurden die Lerner- und Testmanagementstrategien – ähnlich wie in A. D. Cohen und Upton (2006) und Suvorov (2018) – in thematisch zusammenhängende Gruppen eingeteilt.

Auf den unterschiedlichen Gebrauch der häufig codierten Strategien durch die einzelnen Schüler/innen wird in Kap. 4.2.3 näher eingegangen. Der Vergleich der Strategien, die von

den Gruppen der Szenario- und der Off-Szenario-Schüler/innen häufig eingesetzt wurden, bildet den Gegenstand des Kap. 4.4.⁸⁰

4.2.1 LERNERSTRATEGIEN

In den Verbalprotokollen wurden 18 Lernerstrategien ermittelt, die in den Lautdenkprotokollen und/oder in den *Stimulated-Recall*-Interviews vorkamen. In den Lautdenkprotokollen konnten insgesamt 450 und in den *Stimulated-Recall*-Interviews 277 Verwendungen von Lernerstrategien beobachtet und codiert werden, was knapp 35% der in den Lautdenkprotokollen und ca. 24% der in den Interviews gefundenen Strategien entspricht. Tabelle 14 zeigt auf, wie oft welche Strategien in den beiden Untersuchungsteilen codiert und wie oft spezielle Gewichtungen der codierten Segmente vorgenommen wurden.

⁸⁰ Weitere, vertiefende Informationen dazu können auch der Projektdokumentation im Forschungsdatenarchiv entnommen werden (<https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000012367>).

Tabelle 14: Häufigkeit der codierten Lernerstrategien inkl. Gewichtungen.

Kurzbezeichnung Code	Code LD	Anzahl Codierungen in LD-Protokollen (n=20) ⁸¹	Gewichtung „2“ ⁸²	Gewichtung „5“ ⁸³	Code SR	Anzahl Codierungen in SR-Interviews (n=20)	Gewichtung „2“	Gewichtung „5“	Anzahl Codierungen in SR-Interviews nach Abzug Codierungen mit Gew. „2“ und Gew. „25“
Sorgfältiges Lesen	LD-L01	86	0	0	SR-L01	18	0	5	18
Schnelles Lesen	LD-L02	47	0	0	SR-L02	42	5	2	37
Einen Teil des Inputtextes (erneut bzw. genauer) lesen	LD-L03	96	0	0	SR-L03	44	7	22	37
Weiterlesen, um Nichtverstandenes zu verstehen	LD-L04	0	0	0	SR-L04	6	0	4	6
Bisheriges Verständnis hinterfragen	LD-L05	4	0	0	SR-L05	0	0	0	0
Bisheriges Verständnis korrigieren	LD-L06	11	0	0	SR-L06	3	0	0	3
Bisheriges Verständnis bestätigen	LD-L07	1	0	0	SR-L07	0	0	0	0
Intensive Auseinandersetzung mit Teilen des Inputtextes	LD-L08	124	0	0	SR-L08	14	0	7	14
Informationen verknüpfen	LD-L09	3	0	0	SR-L09	17	1	6	16
Unbekannte Wörter, Satz- bzw. Textteile identifizieren	LD-L10	41	0	0	SR-L10	213 ⁸⁴	0	6	33
Unbekannte Wörter ausblenden	LD-L11	1	0	0	SR-L11	16	0	5	16
Wortbedeutung identifizieren	LD-L12	7	0	0	SR-L12	6	1	5	5
Morphosyntaktische Regeln anwenden	LD-L13	3	0	0	SR-L13	11	0	3	11
Interpunktionskenntnisse nutzen	LD-L14	1	0	0	SR-L14	1	0	0	1

⁸¹ Weil es unter den codierten Lernerstrategien keine Codierungen mit Gewichtung „2“ gab, musste von den Zahlen in dieser Spalte nichts abgezogen werden und die Spalte „Anzahl Codierungen in den LD-Protokollen nach Abzug der Codierungen mit Gewichtung „2““ erübrigte sich hier.

⁸² Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

⁸³ Eine allgemeine(re) Aussage zum eigenen Verhalten beim Lösen der Testaufgaben, die aufgaben-, item- oder textstellenübergreifend ist und nicht nur das gerade besprochene Item oder die betroffene Aufgabe oder Textstelle betrifft.

⁸⁴ 180 von diesen 213 Codierungen hatten die spezielle Gewichtung „25“, d.h. es handelte sich dabei um Codierungen, die bei den mündlichen Zusammenfassungen der Inputtexte vorgenommen wurden.

Kurzbezeichnung Code	Code LD	Anzahl Codierungen in LD-Protokollen (n=20) ⁸¹	Gewichtung „2“ ⁸²	Gewichtung „5“ ⁸³	Code SR	Anzahl Codierungen in SR-Interviews (n=20)	Gewichtung „2“	Gewichtung „5“	Anzahl Codierungen in SR-Interviews nach Abzug Codierungen mit Gew. „2“ und Gew. „25“
Kognaten identifizieren und nutzen	LD-L15	6 ⁸⁵	0	2	SR-L15	28	2	12	26 ⁸⁶
Prosodie beim Hören nutzen	LD-L16	0	0	0	SR-L16	4	2	0	2
Relevantes Hintergrundwissen einbeziehen	LD-L17	3	0	0	SR-L17	2	0	0	2
Selbsteinschätzung	LD-L18	16	0	0	SR-L18	0	0	0	0
Total		450	0	2		425	18	77	227

⁸⁵ 2 von diesen 6 Codierungen hatten die spezielle Gewichtung „10“, d.h. der oder die Schüler/in hat das Wort fälschlicherweise als Kognat eingestuft bzw. hat die Bedeutung des Wortes im vorliegenden Kontext falsch ermittelt; oder er/sie vermutete, dass es sich bei einem Wort um einen Kognaten handelt, er/sie wusste aber nicht, was das Wort auf Deutsch bedeutet.

⁸⁶ 6 von diesen Codierungen hatten die spezielle Gewichtung „10“ (vgl. vorangegangene Fussnote).

Die in den Daten identifizierten und codierten Lernerstrategien wurden in drei Gruppen eingeteilt, die als thematisch zusammenhängend betrachtet werden können. Zu diesen Gruppen gehören: A) Strategien zur Herangehensweise an die Lektüre des Inputtextes bzw. seiner Teile, B) Strategien zur Nutzung des Textabschnitts, der Hauptideen und des Hintergrundwissens als Verständnishilfe, und C) Strategien zur Nutzung des linguistischen Wissens. Die inhaltliche Darstellung der einzelnen Strategien folgt dieser Einteilung.

Gruppe A) Strategien zur Herangehensweise an die Lektüre des Inputtextes bzw. seiner Teile

Als Strategien, die mit den Herangehensweisen der Lernenden an die Lektüre der französischen Inputtexte oder seiner Teile zu tun haben, werden in dieser Arbeit die Strategien „Sorgfältiges Lesen“, „Schnelles Lesen“, „Einen Teil des Inputtextes (erneut bzw. genauer) lesen“, „Intensive Auseinandersetzung mit Teilen des Inputtextes“, „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“, „Unbekannte Wörter ausblenden“ und „Selbsteinschätzung“ betrachtet. Sie werden im Folgenden näher beleuchtet.

Strategie „Sorgfältiges Lesen“ (L01)

Die Strategie „Sorgfältiges Lesen“ wurde in den Lautdenkprotokollen aufgrund der Videoaufnahmen codiert, und zwar dann, wenn die Schüler/innen einen längeren Inputtext mindestens eine Minute lang bzw. die kurzen Abschnitte zu Internetreviews in den Szenarien „Stadtrip“ und „Berufswahl“ mindestens 20 Sekunden lang lasen. Manche Lernende lasen die Texte still und verbalisierten kaum etwas dabei, andere lasen Teile des Textes vor, wieder andere übersetzten fast jedes verstandene Wort ins Deutsche und/oder verbalisierten ihre Gedanken dazu mehr oder weniger ausführlich.

Diese Strategie wurde in den *Stimulated-Recall*-Interviews aufgrund der Bemerkungen der Lernenden immer dann codiert, wenn aus dem Gesagten hervorging, dass sie den ganzen Text lasen und möglichst viel davon zu verstehen versuchten, wie z.B. Mogli:

ich habe eigentlich (-) (vorzu) gelesen satz für satz und dann habe ich noch einmal schnell überLEGT was verstehe ich was verstehe ich nICHt- (Mogli_Off, Pos. 304-306)

Diese Strategie wurde in den Lautdenkprotokollen 86 und in den *Stimulated-Recall*-Interviews 18 Mal codiert (davon 5 Mal itemunabhängig, d.h. es handelte sich hier um eine allgemeine(re) Aussage die nicht nur das gerade besprochene Item oder die betroffene Aufgabe oder Textstelle betraf, was mit Gewichtung „5“ markiert wurde).

Strategie „Schnelles Lesen“ (L02)

Die Codierung dieser Strategie wurde in den Lautdenkprotokollen (wie auch die Strategie „Sorgfältiges Lesen“) aufgrund der Videoaufnahmen vorgenommen. Die Strategie wurde codiert, wenn die Lernenden einen längeren Inputtext innert weniger als einer Minute bzw. die kurzen Abschnitte zu Internetreviews in Task C, Szenario „Stadttrip“ und in der Off-Szenario-Aufgabe „Berufswahl“ innert weniger als 20 Sekunden lang rezipierten. Wie auch bei der Strategie „Sorgfältiges Lesen“ konnte dabei beobachtet werden, dass gewisse Lernende es bevorzugten, den Text still zu lesen, während andere ihre Gedanken dazu verbalisierten, den Text vorlasen und/oder das Verstandene auf Deutsch zusammenfassten.

Der Code „Schnelles Lesen“ wurde im *Stimulated-Recall*-Interview einem Transkriptteil zugewiesen, wenn die Lernenden berichteten, dass sie den Text bzw. Teile davon nur schnell zur Kenntnis genommen, z.B. weil sie Ausschau nach bestimmten Schlüsselwörtern gehalten hatten:

`ich glaube ich habe eher nach entsprechendem gesucht ich habe mehr
überFLOGEN` (Omega_Off, Pos. 584-585)`

„Schnelles Lesen“ wurde in den Lautdenkprotokollen 47 Mal und in den *Stimulated-Recall*-Interviews 37 Mal (davon 2 Mal itemunabhängig) codiert.

Nicht immer liess sich aufgrund der *Stimulated-Recall*-Transkripte ermitteln, ob die Schüler/innen den Text schnell oder sorgfältig gelesen hatten, wie z.B. in dieser Aussage von Aria:

`*also ich habe zuerst geLESen`* (Aria_Szenario, Pos. 488)`

In solchen Fällen wurde die Stelle uncodiert gelassen und ein entsprechendes Memo angefertigt.

Strategie „Einen Teil des Inputtextes (erneut bzw. genauer) lesen“ (L03)

Die Codierungen dieser Strategie wurden vorgenommen, wenn aus der Videoaufnahme und/oder aus den Verbalisierungen der Schüler/innen hervorging, dass sie einen Textteil erneut bzw. genauer lasen oder gelesen hatten, wie z.B. der Schüler Böbu:

`°hhh ((verzieht das gesicht)) *hää? (-) ich muss nochmal lesen;*
(Boebu_Szenario, Pos. 53-54)`

In den Lautdenkprotokollen konnte diese Strategie 96 Mal beobachtet werden.

In den *Stimulated-Recall*-Interviews wurde die Strategie 37 Mal, davon 22 Mal itemunabhängig codiert. Sie wurde in diesem Untersuchungsteil dann codiert, wenn aus den Äusserungen der Lernenden hervorging, dass sie eine bestimmte Textstelle genauer unter die Lupe genommen hatten. Diese Strategie kam auch oft im Zusammenhang mit der Strategie „Schnelles Lesen“ vor. So berichtete beispielsweise Dustin, dass sie die Texte zuerst überflogen hatte, um dann die Stellen genauer zu lesen, in denen sie ein Schlüsselwort fand:

das (-) einfach durchgelesen und dann überflogen und wenn ich etwas (-) gefunden habe was sein könnte (-) genauer gelesen. (Dustin_Off, Pos. 695-697)

Strategie „Intensive Auseinandersetzung mit Teilen des Inputtextes“ (L08)

Die Strategie „Intensive Auseinandersetzung mit Teilen des Inputtextes“ kam in den Lautdenkprotokollen am häufigsten vor (124 Mal). Sie wurde codiert, wenn die Schüler/innen Teile des Inputtextes vorlasen, wiederholten oder paraphasierten; Wörter, Satzteile oder Sätze übersetzten oder auch Teile des Inputtextes zusammenfassten. Einige Lernende verbalisierten dabei nur einige Stichwörter und lasen den Text danach einen Moment lang still weiter:

im tourISmusbüro (1.0) ist eine BUSkarte- ((S liest den inputtext)) (28.3) (Alison_Szenario, Pos. 38-39)

Andere wiederum lasen den Inputtext vor und übersetzten für sich alles, was sie davon verstanden, wie z.B. Aria:

((liest den ersten satz aus dem ersten text vor)) (9.0) (irgendetwas) ist gut (--) zimmer sind schön (--) *sauber* und (.) ((schüttelt mit dem kopf)) gut dekoriert (4.0)*das w-lan (-) ist (---) bei der rezeption *hh aber-* (-) nous (5.0) le connection (--) nous avon pay- (---) ahm fünfzehn euro (.) fünfzehn euro (-) zu zahlen wenn man das will* (Aria_Szenario, Pos. 150-155).

In den *Stimulated-Recall*-Interviews wurde diese Strategie 14 Mal codiert. Bei sieben Fällen wurde sie bei der Besprechung einer bestimmten Stelle genannt:

dann habe ich es durchgelesen` u:nd ÜBErsetzt (Aria_Szenario, Pos. 479-480)

In den restlichen sieben Fällen handelte es sich um eine allgemeine Erwähnung dieser Strategie, die von den Lernenden mehrmals eingesetzt worden war, u.a. von Alison:

also einfach habe ich eigentlich IMMER wieder ein wenig überLEGt- für mich immer wieder ein wenig übersEtzt- ja; ((S lächelt)) ja. (Alison_Szenario, Pos. 1377-1378)

Strategie „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ (L10)

Die Strategie „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ gehört zu den in den Lautdenkprotokollen eher häufig codierten Strategien (41 Codierungen). Eine Transkriptstelle wurde mit dieser Strategie immer dann codiert, wenn festgestellt wurde, dass die Lernenden ein unbekanntes Wort, einen unbekannten Satz- bzw. Textteil identifizierten bzw. wenn sie Zweifel an ihrem Verständnis dieses Wortes oder dieser Satz- bzw. Textteile äusserten, wie z.B. im Transkript von Arya:

```
*es gibt verschiedene (--) "monuMEnts" (-) ich weiss nicht mehr genau  
was das HEISST-* (Arya_Szenario, Pos. 18-19)
```

Die unbekannten Wörter, Sätze und Textteile wurden von den Lernenden auch häufig in den *Stimulated-Recall*-Interviews angesprochen, und zwar insgesamt 33 Mal⁸⁷. Manche Schüler/innen, darunter Alison, berichteten in diesem Zusammenhang auch, dass sie sich durch ihr lückenhaftes Textverständnis unsicher (gewesen) waren, ob die von ihnen gewählte Antwort richtig war:

```
dieses wort "EnnuyEUx" ich weiss nicht das verstEHe ich eben nicht-  
und da war ich mir dann nicht genau sIcher- (Alison_Szenario,  
Pos. 533-535)
```

Strategie „Unbekannte Wörter ausblenden“ (L11)

Wenn ein unbekanntes Wort angetroffen wurde und dieses nicht entschlüsselt werden konnte, setzten manche Schüler/innen die Strategie „Unbekannte Wörter ausblenden“ ein. Diese Strategie wurde nur in einem Lautdenkprotokoll beobachtet, dem von Alison. Sie entschied sich bei Szenario „Stadttrip“, Task C Seite 2 für die Antwort „im Zimmer gibt es gratis WLAN“, obwohl sie nicht wusste, was der Ausdruck „sans fil“ („drahtlos“) im Inputtext bedeutet:

```
("SANS fil") (--) weiss ich jetzt nicht genAu aber dann (.) denke ich  
es ist gRAtis- (Alison_Szenario, Pos. 127)
```

In den *Stimulated-Recall*-Interviews wurde hingegen über den Gebrauch dieser Strategie häufiger, insgesamt 16 Mal berichtet. So erklärte z.B. Arya, wie sie vorgegangen war, als sie den ihr ebenfalls unbekannten Ausdruck „sans fil“ an der gleichen Stelle wie Alison angetroffen hatte:

```
hier steht eben (1.5) "il y avait une connexion internet sans fil";  
ich habe einfach wirklich NICHT verstanden was jetzt "sans fil" heisst  
(.) heisst OHne nönönö und em ich habe jetzt trotzdem mal auf JA  
geklickt weil ich dachte (1.5) e:: "il y avait une connexion internet"
```

⁸⁷ Die Codierung dieser Strategie wurde darüber hinaus 180 Mal in den mündlichen Zusammenfassungen der Inputtexte vorgenommen und anschliessend mit Gewichtung „25“ speziell gekennzeichnet.

und das sagt eigentlich es hat eine (-) verbindung zum internet ohne etwas (.) aber es HAT eine- also habe ich einfach mal ja geklickt- (Arya_Szenario, Pos. 1453-1459)

Strategie „Selbsteinschätzung“ (L18)

Die Strategie „Selbsteinschätzung“ wurde in den Lautdenkprotokollen 16 Mal gefunden. Sie wurde codiert, wenn die Lernenden eine selbsteinschätzende Aussage in Bezug auf die eigenen Französischkenntnisse bzw. das eigene Textverständnis machten. Die nähere Betrachtung der codierten Segmente zeigte, dass die mit diesem Code erfassten Äusserungen der Schüler/innen ausschliesslich selbstkritisch waren, z.B. bei Dustin:

okey irgendwie habe ich da NICHT so viel verstanden (--) (Dustin_Off, Pos. 60)

Bei manchen Codierungen entsteht auch der Eindruck, dass diese Äusserungen nicht (wie beim Lauten Denken vorgesehen) selbstadressiert waren, sondern sich eher an die Forscherin richteten:

ich brauche immer lange bei französisch texten (---) das lesen; (Elena_Szenario, Pos. 86)

In den *Stimulated-Recall*-Interviews wurde diese Strategie nicht codiert, weil es dafür in den Transkripten keine Belege gab.

Gruppe B) Strategien zur Nutzung des Textabschnitts, der Hauptideen und des Hintergrundwissens als Verständnishilfe

Zu den Strategien, bei denen ein Textabschnitt, die Hauptideen oder das Hintergrundwissen der Lernenden eingesetzt werden, um das Textverständnis zu erleichtern, werden hier die Strategien „Weiterlesen, um Nichtverstandenes zu verstehen“, „Bisheriges Verständnis hinterfragen“, „Bisheriges Verständnis korrigieren“, „Bisheriges Verständnis bestätigen“, „Informationen verknüpfen“ und „Relevantes Hintergrundwissen einbeziehen“ gezählt.

Strategie „Weiterlesen, um Nichtverstandenes zu verstehen“ (L04)

Die Strategie „Weiterlesen, um Nichtverstandenes zu verstehen“ wurde in keinem der Lautdenkprotokolle codiert, weil kein/e Schüler/in beim Lauten Denken erwähnte, dass er oder sie den Text weiterlesen wollte, um nach Informationen zu suchen, die zum Verständnis des bereits Gelesenen beitragen könnten. Hingegen wurde diese Strategie von sechs Lernenden in den *Stimulated-Recall*-Interviews angesprochen. Zwei von ihnen berichteten über diese Strategie bei der Besprechung einer bestimmten Textstelle; vier weitere Schüler/innen

machten hingegen eine allgemeinere Aussage, die vermuten lässt, dass sie mehrmals so vorgegangen waren, wie z.B. Alison:

```
wenn ich mir nicht sicher wa:r (-) habe ich es dAnn (--) zuerst noch  
den rEst gelesen` und dAnn erst eingefügt (Alison_Szenario,  
Pos. 517-519)
```

Strategie „Bisheriges Verständnis hinterfragen“ (L05)

In den Lautdenkprotokollen wurde die Strategie „Bisheriges Verständnis hinterfragen“ 4 Mal codiert, und zwar immer dann, wenn sich die Lernenden fragten, ob ihr bisheriges Textverständnis noch zutrifft, nachdem sie mehr vom Text gelesen oder gehört haben. Dies führte in allen vier Fällen dazu, dass die Lernenden den jeweiligen Textteil nochmals lasen, wie z.B. Howard (vgl. markierte Stellen):

```
ANNe und marc sagen (-) sie soll (-) STAdion tour machen, ((will 2 nach  
Alx ziehen, lässt es aber sein)) (1.5) ODer äh (-) rugby schauen.  
((liest den text)) (5.7) (Howard_Szenario, Pos. 18-21)
```

Diese Strategie wurde in den *Stimulated-Recall*-Interviews von keiner Schülerin und keinem Schüler erwähnt.

Strategie „Bisheriges Verständnis korrigieren“ (L06)

Die Strategie „Bisheriges Verständnis korrigieren“ wurde in den Lautdenkprotokollen dann codiert, wenn aus den Verbalisierungen der Lernenden hervorging, dass sie das Verständnis des Textes oder eines Teils davon korrigierten, nachdem sie den Text erneut (weiter)gelesen oder weitergehört hatten. Sie wurde in den Lautdenkprotokollen 11 Mal beobachtet. So dachte beispielsweise Howard zuerst, dass im Text der Besuch des Schlosses und der Rugbysausstellung empfohlen wird. Nachdem er aber weitergelesen hatte, änderte er seine Meinung:

```
also wollen sie die rugby [...] (2.0) ausstellung; ((liest den text))  
(14.6) [...] also sie hat jetzt DOCH (-) äh je n y irais PAS (--) also  
ich glaube jetzt dass es DOCH nicht so gut ist (Howard_Szenario,  
Pos. 23-30)
```

In den *Stimulated-Recall*-Interviews wurde diese Strategie 3 Mal erwähnt, darunter auch von Howard, der sich im Gespräch an die oben zitierte Stelle erinnerte:

```
das SCHLOSS habe ich zuerst gedacht dann sagt ja je n y irais PAS (-)  
also (-) ne pas heisst ja (-) NICHT; (--) also (-) hab ich jetzt mal  
ANgenommen dass es NICHT gut ist (Howard_Szenario, Pos. 329-333)
```

Strategie „Bisheriges Verständnis bestätigen“ (L07)

Nur in einem Lautdenkprotokoll und in keinem *Stimulated-Recall*-Interview wurde die Strategie „Bisheriges Verständnis bestätigen“ gefunden. Lediglich beim Schüler Böbu konnte beobachtet werden, wie er das aktuelle Verständnis des Inputtextes oder eines Teils davon bestätigte, nachdem er sich damit weiter bzw. vertiefter auseinandergesetzt hatte:

```
*also ganz am anfang haben sie etwas gesagt von (-)* "carte (.) avignon"  
°hh (2.0) *und (--) die kostet zwanzig euro <<tuschelnd> (unv.)> (2.2)  
glaube ich-* h°° (--) nein (.) *die karte ist NICHT (--) ah sie ist  
nicht (.) zum haben;* °h (--) ehm ((hält den zeigefinger unter die nase  
und liest den text)) (4.5) *gut sie empfehlen es SCHON aber (-) man  
kann sie ja nicht haben-* (Boebu_Szenario, Pos. 79-84)
```

Aus dem Video und aus dem Transkript geht hervor, dass Böbu zuerst meinte, dass die „Carte Avignon“ empfohlen wird. Dann las er weiter und meinte kurz, dass die Karte nicht zu haben sei (also wahrscheinlich, dass sie auch nicht empfohlen wird), um dann im Endeffekt doch zum (richtigen) Schluss zu kommen, dass die Karte empfohlen wird.

Strategie „Informationen verknüpfen“ (L09)

In den Lautdenkprotokollen wurden diejenigen Stellen im Transkript als Strategie „Informationen verknüpfen“ codiert, denen zu entnehmen war, dass die Lernenden Teile von gelesenen bzw. gehörten und von ihnen verstandenen Informationen miteinander in Beziehung setzten. Dies konnte insgesamt 3 Mal beobachtet werden, u.a. bei Aria, die im Hörtext (Szenario Stadttrip, Task B, Bildschirmseite 2) die Informationen über den Regen (fälschlicherweise) mit der ebenfalls erwähnten Krankheit der Sprecherin assoziierte:

```
*starker regen (--) ((blickt auf die seite und dann wieder zu den  
antwortoptionen)) (das ist ja etwas wo sie) ist krank geworden*  
(Aria_Szenario, Pos. 129-130)
```

In den *Stimulated-Recall*-Interviews wurde diese Strategie hingegen deutlich öfters erwähnt, und zwar 16 Mal. Davon wurde sie 10 Mal in Bezug auf eine bestimmte Textstelle angesprochen, wie z.B. von Bonnie, die die Wörter „cassée“ („kaputt“) und „Hotel“ in Verbindung gebracht und sich deswegen beim Task B, Szenario „Stadttrip“, Bildschirmseite 1 für die (richtige) Antwortoption „In ihrem Hotel in Avignon funktionierte das Internet nicht“ entschieden hatte:

```
*ich habe etwas mit also (-) eben cassée und so gehört (-) also kaputt,  
(--) dann hatte ich eben das gefühl das könnte ja noch sein (-) und  
dann (-) sie haben noch etwas vom HOTel geredet und dann habe ich das  
gefühl gehabt ja es könnte kaputt sein (--) aber ich habe es AUCH etwas  
zusammengereimt.* (Bonnie_Off, Pos. 566-572)
```

In 6 Fällen handelte es sich um eine allgemeine, aufgabenübergreifende Erwähnung dieser Strategie im Interview, wie z.B. ebenfalls bei Bonnie:

ich habe vor allem nach wörtern gesucht die ich verSTEHe (-) und dann halt den rest je nach dem zusammengereimt oder (-) halt geschaut ja (-) DIE wörter verstehe ich das könnte etwa DAS sein. (Bonnie_Off, Pos. 351-354)

Strategie „Relevantes Hintergrundwissen einbeziehen“ (L17)

Die Strategie „Relevantes Hintergrundwissen einbeziehen“ wurde immer dann codiert, wenn aus den Verbalisierungen der Lernenden beim Lauten Denken bzw. beim *Stimulated-Recall*-Interview hervorging, dass sie *zusätzlich* zu den Informationen aus dem Text auf ihr Welt- bzw. Hintergrundwissen zurückgriffen, um die Bedeutung des Textes zu verstehen. Diese Strategie wurde in den Lautdenkprotokollen 3 Mal codiert, u.a. bei Sheldon, der bei Szenario „Stadttrip“, Task A, Bildschirmseite 2 in Bezug auf die „Carte Avignon“ sagte:

bei (.) dieser K:A:rte kann man glaub vielleicht mit zÜgen oder (-) bÜssen fahren einfach *mit allen* gratis [...] also es kostet halt VORher aber nachher kann man mit allen gratis fahren (.) so vielleicht für einen tag oder so; (Sheldon_Szenario, Pos. 43-46)

Die Strategie „Relevantes Hintergrundwissen einbeziehen“ wurde im obigen Beispiel codiert, weil im Inputtext lediglich stand: „Pour avoir des entrées réduites dans les musées et le bus gratuit, il y a la Carte Avignon. Elle coûte 20€. Tu la trouves à l’office du tourisme.“ („Für einen ermässigten Eintritt in die Museen und kostenloses Busfahren gibt es die Carte Avignon. Sie kostet 20€. Du findest sie im Tourismusbüro.“). Es ist daher davon auszugehen, dass Sheldon die weitere und auch korrekte Interpretation der Textpassage (so vielleicht für einen tag oder so) aufgrund seines Hintergrundwissens vornahm.

Diese Strategie wurde in den *Stimulated-Recall*-Interviews 2 Mal codiert, u.a. bei Sheldon, der im Interview Bezug auf die oben beschriebene Stelle im Inputtext nahm:

am ANfang wurde halt gerade die: (--) cArte Avignon (-) erwÄhnt` und ich dEnke das ist so eine halt (-) wie eine TAGESkarte wo man mit allem fahren kann- (Sheldon_Szenario, Pos. 371-374)

Mit dieser Aussage bestätigte Sheldon die aufgrund seines Lautdenkprotokolls naheliegende Vermutung, dass ihm das Konzept einer Tageskarte bekannt war und dass er bei der Lektüre des Inputtextes davon ausgegangen war, dass es sich bei der „Carte Avignon“ um eine solche Karte handelte.

Gruppe C) Strategien zur Nutzung des linguistischen Wissens

Als Strategien, die dazu dienen, das linguistische Wissen zu nutzen, um die Bedeutung des Inputtextes zu verstehen, werden in dieser Arbeit die Strategien „Wortbedeutung identifizieren“, „Kognaten identifizieren und nutzen“, „Morphosyntaktische Regeln anwenden“, „Interpunktionskenntnisse nutzen“ und „Prosodie beim Hören nutzen“ betrachtet.

Strategie „Wortbedeutung identifizieren“ (L12)

In den Lautdenkprotokollen konnte die Strategie „Wortbedeutung identifizieren“ 7 Mal beobachtet werden. Sie wurde jeweils dann codiert, wenn die Lernenden die Bedeutung eines Wortes entdecken (z.B. Aria) bzw. sich daran erinnern konnten (z.B. Loup):

```
das personal ist (2.0) ((hebt die hände hoch)) serviable ((verzieht  
das gesicht)) (1.5) serviert gut? (Aria_Szenario, Pos. 191-192)  
  
c est vrai ((schaut nach oben)) heisst c est vrai heisst ähm (--)  
RICHTig ja (-) das heisst RICHTig weil (-) faux heisst falsch (.) ja;  
(Loup_Off, Pos. 15-16)
```

Die entsprechende Strategie wurde in den *Stimulated-Recall*-Interviews 5 Mal codiert. Es handelte sich dabei jeweils um allgemeine Erwähnungen dieser Strategie und nicht um eine konkrete Textstelle, an der diese Strategie angewandt worden war. So berichtete z.B. die Schülerin Picard, dass sie versucht hatte, die Bedeutung der unbekannten Wörter aus dem Zusammenhang abzuleiten:

```
ein pAAr wörter habe ich nICHT verstanden` dann habe ich halt einfach  
(-) geschaut was: es hEIssen KÖNNTE vielleicht` mit dem zusammenhang  
des textes- (Picard_Off, Pos. 236-238)88
```

Strategie „Kognaten identifizieren und nutzen“ (L15)

In den Lautdenkprotokollen wurden aufgrund der Bemerkungen der Schüler/innen insgesamt 6 Codierungen der Strategie „Kognaten identifizieren und nutzen“ vorgenommen, darunter bei Ezra:

```
"LE tester" denke ich das hat eine verbindung mit (-) mit dem dEU/ also  
ein parallelwort mit tEsten- (Ezra_Off, Pos. 173-175)
```

In zwei Fällen handelte es sich um eine allgemeine und aufgabenübergreifende Nennung dieser Strategie durch die Schülerin Elena, die 2 Mal solche – wohl eher an die Forscherin als an sich selbst gerichtete – Anmerkungen machte:

⁸⁸ Diese Transkriptstelle wurde gleichzeitig auch mit der Strategie L10 codiert („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“).

manchmal suche ich auch so parallelwörter (--) zum schauen (.) was es sein könnte; (Elena_Szenario, Pos. 36-37)

also ich suche parallelwörter (-) manchmal (Elena_Szenario, Pos. 83)

Die Analyse der *Stimulated-Recall*-Interviews zeigte, dass die Lernenden 26 Mal vom Einsatz dieser Strategie berichteten (darunter 6 Mal itemübergreifend) und nach Wörtern gesucht hatten, die denen aus dem Inputtext ähnlich sind, nicht nur im Deutschen, sondern in allen ihnen bekannten Sprachen. So hatten z.B. Omega auf seine Englischkenntnisse und Mogli auf seine Italienischkenntnisse zurückgriffen, um die Bedeutung der französischen Wörter zu entziffern:

dann hat (-) [...] die andere gesagt (2) "la carte n'est pas vala(.)ble" habe ich wegen dem (-) englischen available` ja (-) also (.) ist nicht (4) wie sagt man- (5) die ist einfach nicht (.) gültig (Cybermen_Szenario, Pos. 442-445)

S:"je pEnse" *dAs kann man jetzt mit dem italienischen (-) "pEnser" glaube ich-*

MB: und was heisst das`

S: *DENKe;* also ich denke- (Mogli_Off, Pos. 460-462)

Bei der genaueren Betrachtung der mit der Strategie „Kognaten identifizieren und nutzen“ gekennzeichneten Stellen konnte festgestellt werden, dass die Schüler/innen gewisse Wörter fälschlicherweise als Kognaten einstufen bzw. die Bedeutung des Wortes im vorliegenden Kontext nicht verstanden. Auch stellte sich heraus, dass die Lernenden manchmal zwar vermuteten, dass es sich bei einem bestimmten Wort um einen Kognaten handelte, aber nicht wussten, was dieses Wort auf Deutsch bedeutet. Solche Fälle wurden in beiden Untersuchungsteilen mit der speziellen Gewichtung „10“ versehen. In den Lautdenkprotokollen wurden zwei solche Stellen beobachtet. So meinte z.B. Elena, dass in Szenario „Stadttrip“, Task A, Bildschirmseite 1 die Antwortoption „Das Schloss anschauen“ zu den Empfehlungen gehört, weil im Text stand „Tu dois absolument goûter leurs burgers!“ („Du musst unbedingt ihre Burger probieren!“) und sie das französische Wort „burgers“ als Kognaten zum deutschen Wort „Burg“ identifizierte:

das schloss anschauen ((zieht 1 nach Alw)) (2.0) *weil ich denke dass (-) hier STEHT ja (-) burg ((zeigt mit der maus auf "leurs burgers")) (-) das ist ja schloss.* (Elena_Szenario, Pos. 18-19)

In den *Stimulated-Recall*-Interviews kam es in 6 von 26 Fällen vor, dass die Lernenden die Bedeutung eines Kognaten falsch ermittelten oder nicht erklären konnten. Dies war z.B. bei Ezra zu beobachten, der das französische Wort „formation“ („Ausbildung“) für einen Kognaten hielt und darunter „Aufstellung einer Fussballmannschaft“ verstand:

S: und "formation" heisst formation; hm [...]

MB: was IST formation? wie verstehst du das`

S: eine formation also zum beispiel beim fußball eine: formaTION
 MB: also wie eine gruppierung`
 S: also (-) zum beispiel beim fußball dass es immer vier verteidiger
 drei mittelfeld und drei stürmer hat- habe ich gedacht;
 MB: Ah und das ist wie: die formation der mannschaft-
 S: ja. [...] also habe ich gedacht das ist ja ein parallelwort- (Ezra_Off,
 Pos. 1180-1194)

Beim Schüler Eleven stellte sich hingegen im Gespräch heraus, dass er zwar vermutete, dass es sich beim französischen Wort „formation“ um einen Kognaten handelt. Er konnte aber nicht erklären, was das Wort auf Deutsch bedeutet:

S: [...] apres je fini ma formation *er hat etwas seine formation beendet
 irgendetwas so?`
 MB: wie verSTEHST du formation?
 S: *formation das ist (-) ich glaube das ist einfach so ein parallelwort
 für in für das DEUTSCHE;`
 MB: [...] was heisst das im deutschen,
 S: *formation also (-) eben (-) ich weiss nicht ganz; (--) also ich
 (-) also ich verstehe es nicht HUNDert prozent aber ich DENKe es jetzt
 einfach (-) JA_a (-) irgendetwas er ist (-) äh er ist dann irgendetwas
 FERTig mit etwas;` (--) (Eleven_Off, Pos. 943-951)

Strategie „Morphosyntaktische Regeln anwenden“ (L13)

Die Strategie „Morphosyntaktische Regeln anwenden“ konnte in den Lautdenkprotokollen 3 Mal beobachtet werden, davon 2 Mal beim Off-Szenario-Schüler Omega. Im ersten Fall geht aus dem Transkript hervor, dass er den Inputtext von Mélanie in Aufgabe „Berufswahl“, Bildschirmseite 1 nicht ganz verstand. Er identifizierte zwar das Wort „voyager“ richtig als „reisen“, schien aber etwas aus dem Konzept gebracht zu sein, weil das Verb im Text (in seinen Augen) nicht konjugiert zu stehen schien:

*AH ich habe schon jetzt keine ahnung was das heisst- (-) ("voyager")
 reisen; aber es ist nicht konjugiert* (--) (Omega_Off, Pos. 90-92)

Er wählte bei dieser Aufgabe trotzdem die richtige Antwort, die Strategie schien ihn hier aber mehr zu verunsichern als ihm geholfen zu haben. Im zweiten Fall bemerkte Omega, dass es sich bei „voulais“ um die Vergangenheitsform von „wollen“ handelt:

"je voul/" A:H; (2.0) also (---) "vouler" w_wollen; (-) wollte; (-)
 vergangenheit; ((liest den text)) (5.0) (Omega_Off, Pos. 140-142)

Auch hier schien die Strategie nicht wirklich hilfreich zu sein: Die weiteren Transkriptteile zeigen, dass er den Text nicht verstand und deswegen bei der Auswahl seiner Antwort auf zwei *test-wiseness*-Strategien (TW02 und TW07) zurückgriff. Im dritten beobachteten Fall wurde diese Strategie von Aria bei Szenario „Stadttrip“, Task A Bildschirmseite 1 angewandt. Aria scheint nicht verstanden zu haben, was in der folgenden Stelle aus dem Inputtext unter „je n’y irais pas“ zu verstehen ist: „Le château est un classique et un des monuments les plus

visités. Mais franchement, je n'y irais pas.“ („Das Schloss ist ein Klassiker und eine der am häufigsten besuchten Sehenswürdigkeiten. Aber ehrlich gesagt würde ich dort nicht hingehen.“). Sie meint (fälschlicherweise), dass es sich hier um eine Vergangenheitsform handelt:

```
das schloss ist ein klassisch_es monument immer (unv.) oder so ne faire  
aber *irgendetwas was nicht ist (--) ist nie (---) ((verzieht das  
gesicht)) es interessiert mich nicht so? (--) keine ahnung* (3.0)  
(c'est imparfait) ((lacht unsicher)) (Aria_Szenario, Pos. 25-29)
```

Diese Strategie scheint auch Aria nicht geholfen zu haben, die nicht verstandenen Teile des Inputtextes besser zu begreifen.

In den *Stimulated-Recall*-Interviews berichteten die Schüler/innen 11 Mal über den Einsatz der Strategie „Morphosyntaktische Regeln anwenden“. Interessanterweise wurde in 10 Fällen (davon 3 Mal aufgabenübergreifend) erwähnt, dass sie bei der Lektüre der Texte auf die Verneinung geachtet hätten (die im Französischen mit „non“ oder „ne...pas“ markiert wird), was in den Lautdenkprotokollen nicht beobachtet werden konnte. Der Fokus auf die Verneinung schien den Lernenden geholfen zu haben, die allgemeine Textaussage zu begreifen, auch wenn sie nicht in der Lage waren, alle Schlüsselwörter im Text zu verstehen. So berichtet z.B. Cersei darüber:

```
MB: warum MEINtest du dass sie die karte NICHT äh (--) [empfehlen?]  
S: [weil sie] zwanzig euro kostet (-) le carte n est pas vala/ (-)  
valable (-) und da dachte ich wenn es schon (-) NE pas ist (-)  
dass es sie nicht gut ist [für (-) das ja;]  
MB: [hm_hm (-) ah okay] (--) verSTEHST du das wort valable;  
S: nein, (Cersei_Szenario, Pos. 532-537)
```

Die einzige Bemerkung aus den *Stimulated-Recall*-Interviews, die sich im Zusammenhang mit dieser Strategie nicht auf die Verneinung bezog, stammt vom Schüler Ezra:

```
ich mAChe die: (1.5) also ich habe (unv.) es ist MEHRzahl etwas-  
(Ezra_Off, Pos. 1322-1323)
```

Es war Ezra zwar aufgefallen, dass ein Wort im Inputtext im Plural stand, da er aber das Wort nicht kannte, half ihm die Strategie nicht weiter, diesen Textteil zu verstehen.

Strategie „Interpunktionskenntnisse nutzen“ (L14)

Die Strategie „Interpunktionskenntnisse nutzen“ konnte nur im Lautdenkprotokoll des Schülers Eleven beobachtet werden. Bei seiner Auseinandersetzung mit dem Text von Etienne in der Aufgabe „Berufswahl“, Bildschirmseite 1, machte er folgende Bemerkung:

```
nein ((wählt C1_1b)) (--) *hier, ((zieht 3 nach C1_1rev)) °h weil hier  
((meint den text von etienne)) steht irgendwie nouveau produit (-) sie
```

haben ein neues (-) es heisst a_VANT* (1.0) hat es neue produkt. testen
(.) ((schaut auf den text)) man sieht nichts mit fragezeichen*
(Eleven_Off, Pos. 110-113)

Es bleibt allerdings unklar, warum er in diesem Inputtext nach einem Fragezeichen suchte.

Im *Stimulated-Recall*-Teil wurde diese Strategie nur von Alison angesprochen. Sie meinte, dass für sie ein Ausrufezeichen ein zusätzlich bestätigendes Signal dafür gewesen war, dass etwas im Text positiv dargestellt und empfohlen wurde:

zum beispiel "OUI" dann (mit einem) ausrufezeichen (-) kann man sich
schon dEnken dass es dann (ein) (--) pOsitiv ist` also dass sie das
empfEHLEn- (Alison_Szenario, Pos. 331-333)

Strategie „Prosodie beim Hören nutzen“ (L16)

Die Strategie „Prosodie beim Hören nutzen“ liess sich in keinem der Lautdenkprotokolle beobachten. Hingegen wurde sie von Arya und Böbu in den *Stimulated-Recall*-Interviews erwähnt, und zwar in beiden Fällen in Bezug auf die Hörverstehensaufgabe im Szenario „Stadttrip“ (Task B, Bildschirmseite 2). Die Frage in dieser Multiple-Choice-Aufgabe lautete: „Was hat Lou in Avignon besonders beeindruckt?“ und den Lernenden standen drei Antwortoptionen zur Wahl: „Das gute Essen am zweiten Tag“, „Der starke Regen in der Stadt“ und „Die schöne Aussicht vom Schloss.“ Böbu überlegte sich dazu zuerst, dass sich wohl niemand über den starken Regen in der Stadt freut. Er hörte dann den Inputtext, verstand die entsprechende Stelle nicht ganz und schien durch die von ihm eingesetzte Strategie „Prosodie beim Hören nutzen“ etwas verunsichert zu sein, denn die Bloggerin Lou hatte seiner Ansicht nach „recht fröhlich“ geklungen, als sie über den Regen erzählte:

*sie hat irgendetwas gesagt vom REGEN und (-) (unv.) (-) etwas habe
ich nicht so ganz verstanden` (-) sie klang zwar recht FRÖhlich aber
(-) °h war dann trotzdem nicht so sicher-* (Boebu_Szenario,
Pos. 1290-1292)

Der Schülerin Arya war hingegen bereits beim Lesen der auf Deutsch verfassten Antwortoptionen aufgefallen, dass sie das französische Wort für „Regen“ nicht kannte. Sie hatte deswegen ähnlich wie Böbu überlegt, dass der Regen während des Urlaubs für die meisten Menschen wohl etwas Negatives wäre und gehofft, dass sie vielleicht mit etwas Glück der Stimme der Sprecherin entnehmen kann, wann diese über den Regen berichtete:

dann wusste ich okay REgen; (---) ich weiss nicht was regen IST aber
für die meisten menschen ist es ja etwas NEGAtives` em: vor allem wenn
man in den ferien ist ist es ein wenig (.) schADe wenn (es) nur regnet;
und em dann wusste ich okay (---) es ist etwas NEGAtives und vielleicht
(-) hört man das dann auch an ihrer stimme also oder an ihrem AUSdruck-
(Arya_Szenario, Pos. 1103-1110)

Weiter erklärt Arya auch, warum sie diese Strategie beim Lauten Denken nicht verbalisiert hat. Sie behauptet, dass ihr viele Gedanken zum einen sehr schnell und zum anderen auch während des Hörens durch den Kopf gegangen sind, sodass sie keine Zeit bzw. Möglichkeit gehabt hatte, diese laut auszusprechen:

```
ich habe mir ein bisschen das überlegt- also einfach sehr schnell [...]
ich erzähle es jetzt so LANge aber es war so einfach ((schnippst mit
den fingern)) so- und dann habe ich es mir versucht anzuhören; ja; und
viele der überlegungen habe ich dann auch WÄHrend dem hören em (--)
geHABT; (unv.) wie zum beispiel wie das NEGAtive (--) dass vielleicht
die stimme sich verändert habe ich mir (-) während dem HÖREN überlegt
(Arya_Szenario, Pos. 1111-1116)
```

Resümee

In diesem Kapitel wurden die Lernerstrategien inkl. der Anzahl ihrer Codierungen in den jeweiligen Untersuchungsteilen ausführlich dargestellt. In den Daten konnten insgesamt 18 Lernerstrategien identifiziert werden, die sich in folgende thematisch zusammenhängende Gruppen einteilen lassen: A) Strategien zur Herangehensweise an die Lektüre des Inputtextes bzw. seiner Teile, B) Strategien zur Nutzung des Textabschnitts, der Hauptideen und des Hintergrundwissens als Verständnishilfe, und C) Strategien zur Nutzung des linguistischen Wissens.

Die Lernerstrategien machten fast 35% der in Lautdenkprotokollen und etwa 24% der in *Stimulated-Recall*-Interviews codierten Strategien aus. Auch wenn es sich dabei um konstruktrelevante Strategien handelte, führte ihr Einsatz nicht automatisch zum Erfolg beim Lösen der Testaufgabe. Dies war u.a. bei den Strategien „Kognaten identifizieren und nutzen“ oder „Morphosyntaktische Regeln anwenden“ deutlich zu erkennen, die an sich zwar konstruktrelevant sind, aber den Schüler/innen in den Fällen wenig genützt hatten, in denen sie von den Inputtexten allgemein zu wenig verstanden, um die Testfragen auf dieser Grundlage richtig zu beantworten.

Das folgende Kapitel ist den Teststrategien gewidmet, zu denen die Testmanagement- und die *test-wiseness*-Strategien zählen. Diese Strategien werden in analoger Weise wie die oben behandelten Lernerstrategien vorgestellt.

4.2.2 TESTSTRATEGIEN

Neben den Lernerstrategien, die auch bei der testunabhängigen Rezeption der Inputtexte eingesetzt werden können, wurden in den Daten auch Strategien gefunden, auf die die Schüler/innen ausschliesslich aufgrund der Testsituation zurückgegriffen hatten. Insgesamt

konnten in der Studie 27 Teststrategien identifiziert werden, darunter 19 bzw. 20⁸⁹ konstruktrelevante Testmanagementstrategien und 7 konstruktirrelevante *test-wiseness*-Strategien.

4.2.2.1 TESTMANAGEMENTSTRATEGIEN

Nach Abzug der Codierungen mit Gewichtung „2“ (wiederholte Nennung der gleichen Strategie an einer weiteren Stelle im Transkript) konnten in den Lautdenkprotokollen zusammengekommen 731 und in den *Stimulated-Recall*-Interviews 571 Codierungen der Testmanagementstrategien gezählt werden. Dies entspricht etwa 57% bzw. 60% aller Strategien, die in den 20 Transkripten codiert wurden. Tabelle 15 zeigt die Häufigkeit der in beiden Untersuchungsteilen codierten Strategien inkl. ihrer Gewichtungen auf.

⁸⁹ Die Strategie TM20 („Selbstkorrektur beim *Stimulated Recall*“) wurde ausschliesslich für den *Stimulated-Recall*-Teil vorgesehen.

Tabelle 15: Häufigkeit der codierten Testmanagementstrategien inkl. Gewichtungen.

Kurzbezeichnung Code	Code LD	Anzahl Codierungen in LD-Protokollen (n=20)	Gewichtung „2“ ⁹⁰	Gewichtung „5“ ⁹¹	Anzahl Codierungen in LD-Protokollen nach Abzug Codierungen mit Gew. „2“	Code SR	Anzahl Codierungen in SR-Interviews (n=20)	Gewichtung „2“	Gewichtung „5“	Anzahl Codierungen in SR-Interviews nach Abzug Codierungen mit Gew. „2“
Instruktionen und/oder Antwortoptionen (erneut) lesen	LD-TM01	291	0	0	291	SR-TM01	134	0	15	134
Paraphrasieren und/oder Übersetzen der Aufgabe	LD-TM02	32	0	1	32	SR-TM02	16	0	2	16
Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren	LD-TM03	2	0	0	2	SR-TM03	15	4	0	11
Vorhersage machen	LD-TM04	2	0	0	2	SR-TM04	5	0	1	5
Frage und Antwortoptionen beim Hören anschauen	LD-TM05	39	0	0	39	SR-TM05	4	1	1	3
Fokus auf relevante Textteile	LD-TM06	93	0	1	93	SR-TM06	145	37	23	108
Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären	LD-TM07	5	0	0	5	SR-TM07	4	1	1	3
Einsatz von Verstandem bei Auswahl/Ausschluss der Antwortoption(en)	LD-TM08	122	2	0	120 ⁹²	SR-TM08	126	6	13	120 ⁹³
Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden	LD-TM09	3	0	0	3	SR-TM09	15	0	6	15
Richtigkeit der Antwortoption überprüfen	LD-TM10	14	0	0	14	SR-TM10	4	0	3	4
Antwortoptionen miteinander vergleichen	LD-TM11	39	0	0	39	SR-TM11	34	1	0	33

⁹⁰ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

⁹¹ Eine allgemeine(re) Aussage zum eigenen Verhalten beim Lösen der Testaufgaben, die aufgaben-, item- oder textstellenübergreifend ist und nicht nur das gerade besprochene Item oder die betroffene Aufgabe oder Textstelle betrifft.

⁹² 33 von 120 Codierungen dieser Strategie im Lautdenk-Teil hatten die spezielle Gewichtung „15“. Sie wurde codiert, wenn die mit TM08 codierte Passage falsch oder nicht genügend verstanden wurde, um die richtige Antwort zu wählen.

⁹³ 26 von 120 Codierungen dieser Strategie in den *Stimulated-Recall*-Interviews hatten die spezielle Gewichtung „15“ (vgl. vorangegangene Fussnote).

Kurzbezeichnung Code	Code LD	Anzahl Codierungen in LD-Protokollen (n=20)	Gewichtung „2“ ⁹⁰	Gewichtung „5“ ⁹¹	Anzahl Codierungen in LD-Protokollen nach Abzug Codierungen mit Gew. „2“	Code SR	Anzahl Codierungen in SR-Interviews (n=20)	Gewichtung „2“	Gewichtung „5“	Anzahl Codierungen in SR-Interviews nach Abzug Codierungen mit Gew. „2“
Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen	LD-TM12	26	0	0	26 ⁹⁴	SR-TM12	34	1	0	33 ⁹⁵
Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war	LD-TM13	26	0	0	26	SR-TM13	22	0	0	22
Auswahl einer Antwortoption, weil andere nicht plausibel sind	LD-TM14	14	0	0	14	SR-TM14	21	0	0	21
Kenntnisse über das Itemformat nutzen	LD-TM15	6	0	0	6	SR-TM15	13	5	7	8
Zeitmanagement	LD-TM16	0	0	0	0	SR-TM16	3	0	2	3
Selbstkorrektur F-R	LD-TM17	11	0	0	11	SR-TM17	3	0	0	3
Selbstkorrektur R-F	LD-TM18	5	0	0	5	SR-TM18	0	0	0	0
Selbstkorrektur F-F	LD-TM19	3	0	0	3	SR-TM19	0	0	0	0
Selbstkorrektur beim SR	-	-	-	-	0	SR-TM20	32	3	0	29
Total		733	2	2	731		630	59	74	571

⁹⁴ 11 von 26 Codierungen dieser Strategie im Lautdenk-Teil hatten die spezielle Gewichtung „20“. Sie wurde codiert, wenn die mit TM12 codierte Passage Schlussfolgerungen beinhaltete, die nach dem Vergleich mit der entsprechenden Textstelle als unplausibel eingestuft wurden, weil es für diese Deutung im Inputtext keine bzw. nicht genügend Evidenz gab.

⁹⁵ 16 von 33 Codierungen dieser Strategie in den *Stimulated-Recall*-Interviews hatten die spezielle Gewichtung „20“ (vgl. vorangegangene Fussnote).

Die codierten Testmanagementstrategien werden im Weiteren entlang drei thematisch zusammenhängender Gruppen dargestellt. Zu diesen zählen: A) Strategien zur Auseinandersetzung mit den Instruktionen und/oder Itemelementen, B) Strategien zur Auseinandersetzung mit dem Inputtext, sowie C) Strategien zur Vorgehensweise bei der Wahl der Antwortoption(en).

Gruppe A) Strategien zur Auseinandersetzung mit den Instruktionen und/oder Itemelementen

Zu dieser Gruppe werden die Testmanagementstrategien gezählt, auf die die Lernenden bei der Beschäftigung mit den Instruktionen und/oder Elementen der Items zurückgriffen. Als solche wurden hier die Strategien „Instruktionen und/oder Antwortoptionen (erneut) lesen“, „Paraphrasieren und/oder Übersetzen der Aufgabe“, „Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren“, „Vorhersage machen“ und „Frage und Antwortoptionen beim Hören anschauen“ festgelegt.

Strategie „Instruktionen und/oder Antwortoptionen (erneut) lesen“ (TM01)

Die Strategie „Instruktionen und/oder Antwortoptionen (erneut) lesen“ konnte in den Lautdenkprotokollen mit Abstand am häufigsten beobachtet werden (291 Codierungen). Sie wurde aufgrund des Videos (z.B. bei Alison) und/oder der entsprechenden Bemerkung(en) der Lernenden codiert (z.B. bei Omega).

```
((S liest die instruktionen)) (5.6) (Alison_Szenario, Pos. 15)
```

```
((S liest die aufgabenstellung und spricht leise mit)) (17.0) <<runzelt  
die stirn> WA:S in welchem text (--) hast du gelesen (ob es) JA oder  
NEIN ist-> (---) (Omega_Off, Pos. 84-85)
```

Sie wurde auch codiert, wenn die Lernenden beim Lesen des Inputtextes die Aufgabenstellung und/oder die Antwortoptionen zum wiederholten Mal konsultierten, wie beispielsweise Arya:

```
*°hh MArc sagt (auch) das sei gut- ((schaut zu den  
antwortmöglichkeiten)) ja. (das heisst es ist) positiv.*  
(Arya_Szenario, Pos. 31-32)
```

Auch in den *Stimulated-Recall*-Interviews wurde diese Strategie sehr oft codiert, und zwar 134 Mal, davon 15 Mal aufgabenunabhängig. So berichtete u.a. Bonnie, dass sie beim Lösen der Testaufgaben immer zuerst die Fragen und Antwortoptionen zur Kenntnis genommen hat, um zu wissen, was in der Aufgabe bzw. im Inputtext auf sie zukommt:

ich mache eigentlich immer das gleiche system (-) dann halt die FRAGE und dann (-) die ANTWORTEN (-) dann habe ich (-) zuerst noch einmal überlegt auf was ich mich quasi befassen muss (Bonnie_Off, Pos. 542-544)

Strategie „Paraphrasieren und/oder Übersetzen der Aufgabe“ (TM02)

Diese Strategie wurde in den Lautdenkprotokollen 32 Mal codiert. In 31 Fällen konnte beobachtet werden, dass die Lernenden die Aufgabe oder Teile davon nicht nur lasen, sondern auch wiederholten oder in eigenen Worten zusammenfassten. So ging z.B. Bonnie beim Szenario „Walkies“, Task C Bildschirmseite 1 wie folgt vor:

((liest die aufgabenstellung)) (3.0) ähm (-) diskuTIERe (3.0) *prospekt* (3.0) *aufgaben*.((liest die aufgabenstellung)) (2.1) ((liest die aufgabenstellung)) (--) *was IST meine aufgabe (.) tabelle mit vor und nachteilen (--) slogan für den prosPEKT (-) meinen text über die erfindung der wa/ der WAlkies schreiben?* ((liest die aufgabenstellung)) (4.5) (Bonnie_Off, Pos. 12-17)

Als einzige machte in den Lautdenkprotokollen Schülerin Cersei eine (wohl an die Forscherin gerichtete) Bemerkung dazu, dass sie die in der Schulsprache stehenden Fragen bzw. die Antwortoptionen immer im Voraus für sich auf Französisch zu übersetzen versuchte:

also ich (--) proBIERE (-) IMMer (-) dIE (-) möglichst auf französisch (-) was das HEISSen könnte (-) und dann kann ich dIE hier im text (-) SUchen? (---) (Cersei_Szenario, Pos. 65-67)

Die Strategie „Paraphrasieren und/oder Übersetzen der Aufgabe“ wurde in den *Stimulated-Recall*-Interviews 16 Mal erwähnt, darunter 2 Mal aufgabenübergreifend. Aus zwei codierten Stellen ging hervor, dass die Lernenden die Aufgabe für sich nochmals zu paraphrasieren versucht hatten, u.a. Bonnie:

*also ich habe wieder oben gelesen (-) [...] dann halt die FRAGE und dann (-) die ANTWORTEN (-) dann habe ich (-) zuerst noch einmal überlegt auf was ich mich quasi befassen muss (Bonnie_Off, Pos. 541-544)

In 14 weiteren Fällen berichteten die Schüler/innen, dass sie sich überlegt hatten, was die Wörter aus der deutschen Aufgabenstellung auf Französisch hiessen. Dies hatte u.a. Cybermen beim Szenario „Stadttrip“, Task B, Bildschirmseite 2 gemacht, um sich auf das Vorkommen dieser Wörter bzw. Themen im Hörtext besser vorzubereiten:

bevor ich (.) play gedrückt habe ich (--) schnell überle:gt so (--) von den wörtern hier welche ich so auf französisch kenne (Cybermen_Szenario, Pos. 756-757)

Strategie „Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren“ (TM03)

Die Strategie „Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren“ wurde codiert, wenn die Lernenden verbalisierten, dass sie 1) in der Aufgabenstellung ein deutsches bzw. 2) bei Off-Szenario in der Aufgabe „Berufswahl“ ein französisches Wort nicht verstanden, oder 3) nicht wussten, wie das Vokabular aus der (deutschen) Aufgabenstellung auf Französisch hiess. Diese Strategie konnte in den Lautdenkprotokollen 2 Mal beobachtet werden und zwar in beiden Fällen beim Schüler Omega. So stellte sich heraus, dass er bei Szenario „Walkies“, Seite 2 nicht wusste, was die Antwortoption „Die Texte der anderen Korrektur lesen“ bedeutet:

```
texte der anderen korrekturlesen- (---) *KEINE ahnung was das ist-*  
(Omega_Off, Pos. 32-33)
```

Im zweiten Fall kannte er in der Frage „Muss man im Beruf *représentant commercial*“ (Szenario Berufswahl, Seite 1) die Bedeutung der französischen Berufsbezeichnung nicht:

```
<<verzieht das gesicht> *eh was IST das; "representant (commercial)">  
(Omega_Off, Pos. 83)
```

Über diese Strategie wurde 11 Mal in den *Stimulated-Recall*-Interviews berichtet, u.a. von Omega, der im Interview beide oben geschilderte Fälle ansprach. Ansonsten erwähnten die Schüler/innen, dass sie die Schlüsselwörter aus den Aufgabenstellungen auf Französisch nicht kannten. Von 6 Lernenden (u.a. von Cersei) wurde erwähnt, dass ihnen das französische Wort für „Regen“, das in Szenario „Stadttrip“, Task B, Seite 2 vorkam, nicht bekannt war:

```
ich wusste nicht was REgen (-) auf französisch heisst und so konnte  
ich es im text auch nicht WENN es vorgekommen (-) ist auch nicht (-)  
hören (Cersei_Szenario, Pos. 921-923)
```

Weitere unbekannte Wörter aus den Aufgabenstellungen, die jeweils 1 Mal erwähnt wurden waren „Aussicht“, „Schloss“, „Produkte“ und „Prospekte“. Omega konnte z.B. im Voraus nicht sagen, was „Produkte“ auf Französisch hiess. Er hatte aber gehofft, dass er das Wort identifizieren würde, wenn es im Text vorkäme:

```
produkte wa/ war ich mir nicht so sicher aber ich wüsste wenn ich es  
sehen würde würde ich es erkennen` (Omega_Off, Pos. 903-904)
```

Strategie „Vorhersage machen“ (TM04)

Die Strategie „Vorhersage machen“ wurde codiert, wenn aus den Transkripten hervorging, dass sich die Schüler/innen vor der Rezeption des Inputtextes zwar überlegten, welche Antwortoption/en evtl. stimmen und welche ausgeschlossen werden könnten, aber noch nicht definitiv entschieden, welche Antwortoption(en) zu wählen bzw. auszuschliessen ist(sind).

In den Lautdenkprotokollen wurde diese Strategie bei Aria und Böbu beobachtet. In beiden Fällen handelte es sich um die gleiche Hörverstehensaufgabe im Szenario „Stadttrip“ (Task B, Bildschirmseite 2), die auch bereits im Zusammenhang mit der Strategie „Prosodie beim Hören nutzen“ erwähnt wurde. Die Frage lautete „Was hat Lou in Avignon besonders beeindruckt?“ und eine der drei Antwortoptionen war „Der starke Regen in der Stadt“. Sowohl Aria wie auch Böbu vermuteten vor dem Hören des Inputtextes, dass diese (übrigens korrekte) Antwortoption wohl kaum in Frage kommen würde. Böbu sagte dazu vor dem Hören:

was hat sie am: (-) nein am (-) besonders beeindruckt; (-) das gute essen`(--) der starke re/ (4.0) °hh oke; (--) denke mal der starke regen in der stadt ist es nicht- (2.5) (Boebu_Szenario, Pos. 159-161)

Nachdem er den Text gehört hatte, entschied er sich trotzdem für eine andere Antwort. Dies schien aber eher damit zu tun zu haben, dass er die entsprechende Stelle im Inputtext nicht verstanden hatte:

un:d (-) d/ starke regen in der stadt (-) °h das habe ich nicht wirklich verstanden was das genau war- (Boebu_Szenario, Pos. 166-167)

Ähnlich sah es bei Aria aus. Sie hielt diese Antwortoption zwar schon vor dem Hören für wenig plausibel:

also der ka ((macht eine bewegung mit der zunge)) der starke regen in der stadt ist sicher nicht weil das ist ja schLECHT (Aria_Szenario, Pos. 122-123)

Sie entschied sich aber nach dem Hören gegen diese Option, weil sie (fälschlicherweise) verstanden hatte, dass der Regen die Bloggerin Lou krank machte:

*starker regen (--) ((blickt auf die seite und dann wieder zu den antwortoptionen)) (das ist ja etwas wo sie) ist krank geworden (---) (Aria_Szenario, Pos. 129-131)

In den *Stimulated-Recall*-Interviews wurde die Strategie „Vorhersage machen“ 5 Mal erwähnt, u.a. von Aria, die im Interview Stellung zur oben besprochenen Stelle aus ihrem Lautdenkprotokoll nahm:

als ich es gelEsen habe` habe ich glaube ich nachher gedacht [...] ob ich etwas schon wie ein bisschen ausschliessen kann weil vorher habe ich ja zum beispiel das mit dem REGEEn schon EHER ausgeschlossen- weil (--) ich nicht/ ich mir nicht vorstellen konnte dass sie das POSitiv beeindruckt hat- und da: habe ich dann aber nicht wirklich etwas gefunden` (Aria_Szenario, Pos. 840-848)

Aria bestätigt hier also, dass sie von Anfang an vermutet hatte, dass die Antwortoption mit dem Regen nicht stimmen würde und sie sie „eher“ ausgeschlossen hatte. Sie hatte sich aber dann gegen diese Option entschieden, weil sie im Inputtext dazu nichts fand bzw. verstand, und nicht (nur), weil sie sie zuvor bereits ausgeschlossen hatte.

Im Zusammenhang mit dieser Strategie machte der Schüler Omega eine für Testentwickler/innen besonders interessante Bemerkung. Omega wandte die Strategie „Vorhersage machen“ an, um zu sehen, ob er die Fragen lediglich aufgrund seines Hintergrundwissens und ohne den Inputtext lesen zu müssen beantworten könnte. Dies schien ihm nicht der Fall gewesen zu sein (was auch eine gute Nachricht für die Testentwickler/innen ist), weshalb er die Texte dann auch durchlas:

```
ich habe zuerst (-) die fragen gelesen glaube ich` (-) [...] dann überlegt  
(--) ob ich eine eigentlich schon ausschliessen könnte mit  
irgendwelchen informationen oder keine ahnung die (-) wenn es  
wissensfragen auch ein bisschen gewesen wären` aber waren es nicht`  
deshalb habe ich nachher (--) mal gelesen (Omega_Off, Pos. 356-362)
```

Einmal wurde diese Strategie auch aufgabenübergreifend erwähnt. Die Schülerin Elena berichtete, dass sie sich nach der Lektüre der Aufgabenstellung und der Fragen immer vorzustellen versuche, was sie an der Stelle der im Inputtext beschriebenen Personen machen oder wählen würde, weil

```
*dann ist es schon einfacher [...] obwohl ich nicht die person bin aber  
(-) was könnte ich machen das dann (-) finde ich es einfacher (--) dann  
zu wählen was (-) es ist;* (Elena_Szenario, Pos. 920-924)
```

Strategie „Frage und Antwortoptionen beim Hören anschauen“ (TM05)

Diese Strategie wurde in den Lautdenkprotokollen insgesamt 39 Mal aufgrund des Videos codiert. Sie wurde bei allen 20 Lernenden bei der einzigen hier vertieft untersuchten Hörverstehensaufgabe zum Szenario „Stadttrip“, Bildschirmseiten 1 und 2 beobachtet, die sowohl von den Szenario- als auch von den Off-Szenario-Schüler/innen gelöst wurde. Dass sie 39 und nicht 40 Mal codiert wurde, ist darauf zurückzuführen, dass der Schüler Sheldon beim Bearbeiten der zweiten Bildschirmseite eine Antwortoption wählte und sofort auf „weiter“ klickte, ohne den Text gehört zu haben. Folglich wurde bei ihm ausser der Strategie „Raten“ (TW01) auch nichts codiert.

In den *Stimulated-Recall*-Interviews wurde diese Strategie nur von Alison und Howard erwähnt. Alison erklärte, dass sie während des Hörens die Antwortoptionen mit dem verglichen hatte, was sie gerade vom Text verstanden hatte:

```
ich habe auch wieder die gelEsen` und dann (-) während dem höRen (--)  
wieder verglICHen ob das jetzt stImmt- (Alison_Szenario, Pos. 848-850)
```

Auch Howard hatte sich die Antwortoptionen aus einem ähnlichen Grund angeschaut. Er hatte die Schlüsselwörter aus den Antwortoptionen mit dem Hörtext verglichen, um herauszufinden, welche Antwortoption richtig wäre:

dann einfach [...] HINGeschaut [...] einfach so die STICHwörter [...] und dann habe ich einfach so geWARTet bis IRgendeines dieser (-) STICHworte dann vorkommen würde (-) und dann habe ich gehört ob das so IST oder NICHT so ist; (--) (Howard_Szenario, Pos. 934-941)

Gruppe B) Strategien zur Auseinandersetzung mit dem Inputtext

Zu dieser Gruppe zählen drei Strategien, die die Lernenden einsetzten, um sich beim Lösen der Testaufgaben mit dem Inputtext zu befassen: „Fokus auf relevante Textteile“, „Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären“ und „Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden“. Obwohl es sich dabei in allen drei Fällen um Strategien handelt, die bei der Herangehensweise der Lernenden an die Inputtexte eingesetzt wurden, werden sie nicht zu den Lernerstrategien, sondern zu den Testmanagementstrategien gezählt. Der Grund dafür ist, dass der Einsatz dieser Strategien klar durch die Testsituation und die Testaufgaben evoziert wurde: Es sind Strategien, die den Lernenden halfen, den Inputtexten die wichtigsten Informationen im Hinblick auf die richtige Beantwortung der Testfragen zu entnehmen.

Strategie „Fokus auf relevante Textteile“ (TM06)

Als Strategie „Fokus auf relevante Textteile“ wurden in den Lautdenkprotokollen diejenigen Stellen codiert, aus denen hervorging, dass sich die Lernenden nach der Kenntnissnahme der Frage und der Antwortoptionen bei der Rezeption der Inputtexte auf jene Teile der Inputtexte konzentriert hatten, die sie als wichtig für die Beantwortung der Frage hielten. Sie ignorierten dabei evtl. Teile des Inputtextes, die ihnen im Hinblick auf die Frage als wenig relevant erschienen. Die Codierung wurde zum einen aufgrund der Bemerkung(en) der Lernenden vorgenommen, wie im Beispiel von Omega, der sich nach der Lektüre der Aufgabenstellung und vor dem Hören des Inputtextes überlegte, worauf er achten sollte:

```
((liest die aufgabenstellung, runzelt die stirn und spricht leise mit))  
(16.1) *(unv.) ich mal etwas von französische sim karte hören (unv.)  
(--) okay entweder NICHT (-) DORF (-) "HOTEL" (-) sim karte und dann  
sollte man etwas von dem hören* (---) (Omega_Off, Pos. 52-55)
```

Zum anderen wurde diese Strategie auch dann codiert, wenn aus dem Transkript hervorging, dass der/die Schüler/in die Lektüre des Inputtextes unterbrach, um die Antwortoptionen nochmals zur Kenntnis zu nehmen (wie bei Eleven), weil es impliziert, dass die Lernenden in dem Moment die Relevanz der gelesenen Stelle im Hinblick auf die Beantwortung der Testfragen überprüfen:

```
((liest den text und murmelt)) (20.9) ((schaut auf den text, schnalzt  
mit der zunge, schaut kurz zur aufgabenstellung)) (5.3)  
((augenbewegungen zwischen dem text und den antwortoptionen)) (2.0)
```

```
*wo kann man das jetzt sehen* ((liest den text und murmelt)) (7.0)
((schaut abwechselnd auf die antwortoptionen und den text und murmelt))
(10.0) (Eleven_Off, Pos. 20-25)
```

Insgesamt wurde diese Strategie in den Lautdenkprotokollen 93 Mal codiert. Die Schülerin Cersei machte dabei eine (eher an die Forscherin gerichtete) Bemerkung, dass sie in den Inputtexten immer nach den Schlüsselwörtern aus den Antwortoptionen suchen würde:

```
also ich (--) proBIERE (-) IMMER (-) [...] DIE hier im text (-) SUCHEN?
(---) und dann schauen ob sie das (.) emPFEHLEN oder NICHT
(Cersei_Szenario, Pos. 65-68)
```

In den *Stimulated-Recall*-Interviews wurde diese Strategie von den Lernenden 108 Mal angesprochen, davon 23 Mal aufgabenübergreifend. So meinte beispielsweise die Schülerin Picard, dass sie zwar den Inputtext sorgfältig und sogar mehrmals gelesen hatte, sich aber im Endeffekt vor allem darauf konzentriert hatte, was sie ihrer Meinung nach für die richtige Beantwortung der Testfrage brauchen würde:

```
also als erstes habe ich (-) Alles mal (2.0) GUT durchgelesen` und wenn
ich halt stellen NICHT so gut verstanden habe` habe ich sie halt (-)
NOCH einmal gelesen` und wenn ich gefunden habe das (-) ist NICHT so
wichtig habe ich es halt (-) weggelassen- und mich auf das konzentriert
was ich (1.0) brAUche oder was wichtig ist; [...] also (-) wenn ich alles
einmal durchlEse` dann habe ich ähm (-) einen ÜBERblick was vorkommt`
u:nd (-) so kann ich halt entscheiden (-) wAs: brAUche ich [...] zum
verSTEHEN` um die lösung zu bekommen- u:nd em (--) wenn ich sehe das
brAUche ich nicht oder das ist nicht wICHTIG dann (-) schliesse ich es
Aus weil (-) es sonst zu viel ist zum (-) die richtige Antwort finden;
(Picard_Off, Pos. 204-220)
```

Strategie „Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären“ (TM07)

Diese Strategie wurde in den Lautdenkprotokollen codiert, wenn aus dem Video (wie bei Alison) bzw. aus den Verbalisierungen der Lernenden (wie bei Loup) hervorging, dass sie die Bedeutung von bestimmten Teilen des französischen Inputtextes nicht (ganz) verstanden und deshalb mithilfe der auf Deutsch vorhandenen Frage und/oder der Antwortoptionen zu klären oder zu bestätigen versuchten.

```
das "cOq rouge" ist ein_e res:to ((blickt zu den antwortmöglichkeiten
und geht zurück zum text)) (.) jA. (Alison_Szenario, Pos. 28)
```

```
dann ((schaut auf die aufgabenstellung)) (-) viel REIsen (-) aHA (-)
voyage heisst REIse (-) gut, (2.0) (Loup_Off, Pos. 101-102)
```

Diese Strategie konnte in den Lautdenkprotokollen 5 Mal beobachtet werden. Im *Stimulated-Recall*-Teil wurde sie von den Schüler/innen 3 Mal erwähnt, davon 1 Mal itemübergreifend von Alison:

S: manchmal sind die tExte ein wenig schwierig zu verstehen manche WÖRter- dann habe ich zuerst ein wenig dIE da hier (-) gelEsen`

MB: also die hier Unten ((meint die antwortoptionen))-

S: ja. [...] also die empfEHlungen oder eben nIcht empfehlungen [...] und dAnn probiert das im text herauszulEsen; (Alison_Szenario, Pos. 259-267)

Strategie „Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden“ (TM09)

Wenn den Videoaufnahmen und/oder den Bemerkungen der Lernenden entnommen wurde, dass sie den Inputtext nicht weiterlasen bzw. weiterhörten, nachdem sie die (zurecht oder vermeintlich) für die Beantwortung der Frage benötigten Informationen im Text gefunden hatten, wurde diese Transkriptpassage als Strategie „Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden“ codiert. An dieser Stelle muss erwähnt werden, warum diese Strategie als konstruktrelevante Testmanagement- und nicht als konstruktirrelevante *test-wiseness*-Strategie eingestuft wurde. Auch wenn man argumentieren könnte, dass es nicht im Sinne des Testkonstrukts ist, dass die Testteilnehmenden aufgrund der Fragestellung nur Ausschnitte der Inputtexte rezipieren, entspricht diese Vorgehensweise genau dem, was unter *search reading* auf lokaler Ebene verstanden und z.B. beim schnellen Durchschauen von verschiedenen Rezensionen praktiziert wird. Weil solche Items auch in die IFB-Aufgaben einbezogen wurden, wurde entschieden, diese Strategie in dieser Arbeit grundsätzlich als konstruktrelevant zu betrachten.⁹⁶

Die Codierung dieser Strategie wurde in drei Lautdenkprotokollen vorgenommen. Beim Lösen von Task C, Bildschirmseite 1 (Szenario „Stadttrip“) realisierte beispielsweise Aria, dass sie einen Teil des Inputtextes mit dem Hotelreview von Miguel nicht versteht. Sie bemerkte aber auch, dass sie die Frage trotzdem (und zwar korrekt) beantworten konnte, und ging zum Lesen des weiteren Hotelreviews über.

ich habe (2.0) dreissig euro beZahlt (---) ((schaut zu den antwortoptionen)) in dem fall gibt es ((wählt bei C2_3 a)) ((schaut auf den inputtext)) ehm (--) ce n'est pas beaucoup (.) es ist nicht viel (unv.) (4.0) *keine ahnung (--) sauber °hh (-) jetzt muss ich nicht WEiterlesen (.) ja hh° (2.0) jetzt kann ich (unv./dort weiterlesen) ((meint text 3))* (Aria_Szenario, Pos. 166-170)

Die Strategie wurde von den Schüler/innen 15 Mal im *Stimulated-Recall*-Interview angesprochen (6 Mal itemübergreifend). So berichtete die Schülerin Arya, dass sie beim

⁹⁶ Im Weiteren werden auch Fälle dargestellt, in denen die Strategie als konstruktirrelevant betrachtet werden könnte.

Bearbeiten der Aufgabe C im Szenario „Stadttrip“ den Inputtext nicht weiterlas, nachdem sie sich sicher gewesen war, darin die richtige Antwort gefunden zu haben:

meistens wenn ich eine antwort (2.0) SEhr genau (-) beantwortet gefunden habe (-) habe ich sie direkt ein/ em: einplatziert; [...] also: zum beispiel hier (-) "le wifi est gratuit à la réception"; und dann wusste ich OKAY es ist gratis; punkt. und dann konnte ich auch direkt beantworten (-) musste nicht noch andere lesen um (-) sicher zu werden; (Arya_Szenario, Pos. 1305-1314)

Diese Strategie ging für Arya in diesem Fall nicht auf: Die Frage, auf die sich die Schülerin hier bezieht, lautete nämlich, ob man im Zimmer gratis WLAN haben könne. In dem von Arya erwähnten Textteil stand zwar, dass das WLAN *an der Rezeption* kostenlos sei, gleich danach kam aber die Information, dass man 15€ mehr für die Woche zahlen müsse, um die Verbindung auch im Zimmer zu haben. Hätte Arya den Text weitergelesen, hätte sie sich vielleicht doch für eine andere Antwort entschieden.

Die obigen Beispiele illustrieren gut, was beim näheren Betrachten der mit dieser Strategie codierten Stellen festgestellt werden konnte: Diese Strategie wurde dann mit Erfolg eingesetzt (d.h. trug zur Auswahl der korrekten Antwortoption bei), wenn die Schüler/innen sowohl die richtigen Textstellen identifizierten als auch ihren Kontext ausreichend verstanden (Fall Aria). Bei denjenigen Lernenden, die diese Strategie als Abkürzungsweg wählten (den man übrigens durchaus als konstruktirrelevant bezeichnen könnte), führte sie hingegen meistens zum Misserfolg (Fall Arya). Es kam aber auch vor, dass trotz des Einsatzes dieser Strategie als Abkürzungsweg die richtige Antwort gewählt wurde, aber aus einem „falschen“ Grund. Dies konnte beim Schüler Howard bei Szenario „Stadttrip“, Task C, Bildschirmseite 2, Item „Stadtzentrum“ beobachtet werden: Er wählte dort nur deswegen eine richtige Antwort, weil er aufgrund des (für die Beantwortung der Frage irrelevanten) Ausdrucks „gare centrale“ („Hauptbahnhof“) aus dem Inputtext den falschen Schluss zog, dass sich das Hotel im Zentrum befinden müsse, wenn es neben dem Hauptbahnhof liegt. Diese Stelle aus dem Inputtext lässt sich auf Deutsch wie folgt übersetzen: „Letztes Jahr waren wir in einem Hotel neben dem Hauptbahnhof... der überhaupt nicht zentral ist!“. Howard las den Text nur bis zur Stelle „neben dem Hauptbahnhof“ und antwortete, dass sich das Hotel im Zentrum befindet. Seine Antwort war trotzdem korrekt, weil aus dem weiteren Teil des Inputtextes hervorging, dass das Hotel, um das es im Inputtext eigentlich geht, im Zentrum liegt. Dies ist ein wichtiger Hinweis darauf, dass diese Stelle im Inputtext überarbeitet werden müsste, weil sie gewisse Testteilnehmende zu falschen Inferenzen verleiten kann, die dann unerwünschterweise zu richtigen Antworten führen.

Gruppe C) Strategien zur Vorgehensweise bei der Wahl der Antwortoption(en)

Als zu dieser Gruppe gehörig wurden diejenigen Testmanagementstrategien betrachtet, auf die die Lernenden zurückgriffen, um die Antwortoption bzw. -optionen zu wählen. Dazu zählen folgende 11 Strategien: „Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“, „Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen“, „Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war“, „Antwortoptionen miteinander vergleichen“, „Auswahl einer Antwortoption, weil andere nicht plausibel sind“, „Kenntnisse über das Itemformat nutzen“, „Richtigkeit der Antwortoption überprüfen“, „Zeitmanagement“, „Selbstkorrektur F-R (falsch-richtig)“, „Selbstkorrektur R-F (richtig-falsch)“, „Selbstkorrektur F-F (falsch-falsch)“. Hierzu zählt auch eine Strategie, die nur für die Codierung im *Stimulated-Recall*-Teil vorgesehen war: „Selbstkorrektur beim SR“.

Strategie „Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“ (TM08)

Diese Strategie wurde codiert, wenn aus den Äusserungen der Schüler/innen hervorging, dass sie eine Antwortoption basierend auf dem vom Inputtext Verstandenem auswählten oder ausschlossen. Bei der genaueren Betrachtung der codierten Transkriptausschnitte stellte sich schnell heraus, dass in vielen Fällen das, was die Schüler/innen vom Text verstanden, nicht mit dem übereinstimmte, was im Inputtext tatsächlich zu lesen war. Um solche Fälle zu kennzeichnen, wurden sie in einem weiteren Schritt zusätzlich mit der Gewichtung „15“ codiert.

Diese Strategie konnte in den Lautdenkprotokollen insgesamt 120 Mal, also recht häufig beobachtet werden, z.B. bei Omega:

also wie ich das verstehe (.) spricht sie die sprachen (-) welche die klienten sprechen (-) das bedeutet eigentlich dass sie mehrere sprachen können muss; (Omega_Off, Pos. 133-135)

Beim Vergleich der Lautdenkprotokolle mit den Inputtexten zeigte sich, dass die Lernenden die Inputtexte in 33 von 120 codierten Fällen falsch oder ungenügend verstanden. Zu solchen mit der Gewichtung „15“ markierten Stellen gehört z.B. die untenstehende Aussage von Mogli, die er zum Text von Mélanie (Szenario „Berufswahl“, Seite 1) machte, kurz bevor er die (falsche) Antwortoption wählte, dass Mélanie in ihrem Beruf mehrere Sprachen sprechen muss:

(in) einer grÜppe- (---) ((lehnt sich zurück und schaut auf den text))
(3.2) ich spreche Italienisch (-) dEutsch italienisch (-) und englisch;
((cursor bewegt sich über das ende des ersten textes, S liest)) (7.7)

```
jA (-) *(da muss man) mEHrere sprachen.* ((S klickt Cl_2a an)) (2.0)
(Mogli_Off, Pos. 89-93)
```

Im Text von Mélanie war hingegen zu lesen, dass es in ihrem Team *andere* Verkaufsvertreter/innen gäbe, die Deutsch, Italienisch oder Englisch sprechen und dass jede/r von ihnen sich Kunden aussuche, die die von ihnen gesprochene Sprache sprechen.

In den *Stimulated-Recall*-Interviews wurde diese Strategie ebenfalls 120 Mal codiert. Bei 13 von 120 Codierungen handelte es sich um Äusserungen, bei denen aus dem Kontext hervorging, dass sie sich nicht nur auf eine spezifische Aufgabe bezogen. Unter den 120 Codierungen wurden darüber hinaus 26 Stellen identifiziert, an denen die Lernenden den Inputtext falsch verstanden hatten.

Strategie „Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen“ (TM12)

Die Strategie „Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen“ wurde codiert, wenn in den Transkripten beobachtet werden konnte, dass die Lernenden eine Antwort aufgrund von Schlussfolgerungen auswählten oder auch ausschlossen, die sie aus dem Inputtext zogen. Diese Strategie ist konstruktrelevant, weil gewisse Items aus dem Test, in denen die Antwort nicht wortwörtlich im Text stand, nach solchen Inferenzen verlangten (z.B. Szenario „Stadtrip“, Task B, Seite 2). Neben diesen Items gab es im Test aber auch solche, die keine Inferenzen verlangten, weil die Antwort auf die Frage eigentlich klar im Text vorkam. Manche Lernende zogen auch bei diesen Items Schlussfolgerungen, die es dort gar nicht brauchte. Mit dem hier beschriebenen Code wurden also auch Stellen markiert (ähnlich wie bei der Strategie „Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“), bei denen beobachtet werden konnte, dass die Lernenden beim Lösen einer Aufgabe falsche Schlussfolgerungen zogen. Dies passierte oft dann, wenn der Text nicht genügend verstanden wurde. Aria verstand z.B. bei Szenario „Stadtrip“, Task A Seite 2 fälschlicherweise, dass den Schüler/innen im Chat eine Stadtrundfahrt mit dem Sightseeingbus empfohlen würde und schlussfolgerte daraus, dass in diesem Fall der Stadtrundgang zu Fuss nicht empfohlen wäre:

```
und *stadtrundgang machen (--)) ((schaut auf den inputtext)) machen sie
ja mit dem bu_s (Aria_Szenario, Pos. 62)
```

Bei gewissen so codierten Segmenten können auch die von den Schüler/innen gezogenen Schlüsse als zu weit hergeholt betrachtet werden, d.h. es gibt im Text nicht genügend Anhaltspunkte, die die Interpretation der Schüler/innen stützen. Die Schülerin Dustin schloss beispielsweise bei Szenario „Stadtrip“, Task B Seite 2 die Antwortoption „Ihre neue

französische SIM-Karte war noch nicht aktiviert“ aus, weil sie sich aufgrund ihres (falschen) Weltwissens Folgendes überlegte:

ich nehme an wenn (-) die sim karte (-) nicht aktiviert ist kann sie
ja trotzdem ins INternet (Dustin_Off, Pos. 46-47)

Nach dem Vergleich der codierten Passagen mit dem Inputtext wurden solche Codierungen, bei denen die Schüler/innen (eine) Antwort(en) wegen unplausiblen Schlussfolgerungen wählten bzw. ausschlossen, mit der Gewichtung „20“ markiert. Im Lauten Denken wurden 11 von 26 und beim *Stimulated Recall* 16 von 33 Codierungen mit dieser Gewichtung gekennzeichnet.

Strategie „Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war“ (TM13)

Diese Strategie wurde codiert, wenn die Schüler/innen eine Antwortoption auswählten, weil sie im Inputtext korrekterweise ein zu den Schlüsselwörtern aus dieser Antwortoption passendes Wort oder passende Wörter identifizierten. In den Lautdenkprotokollen wurde diese Strategie insgesamt 26 Mal codiert. Sie kam oft bei Items vor, die auf selektives Lesen (*search reading*) abzielten, wie z.B. gewisse Items bei Szenario „Stadttrip“, Task C. Eine der Fragen in dieser Aufgabe lautete: „Gibt es im Hotel Zimmer unter 35 € pro Nacht?“ und in der entsprechenden Passage im Inputtext (Hotelreview von Miguel, Bildschirmseite 1) war zu lesen: „Nous avons payé environ 30€ pour la nuit“ („Wir haben etwa 30€ für die Nacht bezahlt“). Cybermen machte beim Lösen dieser Aufgabe folgende Bemerkung:

dann hier (.) ((schaut auf den text)) dreissig euro "pour la nuit-"
(2.0) ((schaut auf die aufgabenstellung)) <<sehr leise> also ist das
auch ja-> ((wählt C1_3a an und zieht 2 nach C1_3rev))
(Cybermen_Szenario, Pos. 125-126)

Im *Stimulated-Recall*-Teil wurde diese Strategie 22 Mal angesprochen, darunter von Cybermen, der sich so an seine Vorgehensweise beim Bearbeiten des oben dargestellten Items erinnerte:

dann habe ich halt gerade gesehen (--) dreissig euro: (.) für eine
nacht- (-) dann habe ich das glaube ich dann schon direkt (--) wieder
ausgefüllt- (Cybermen_Szenario, Pos. 973-975)

Allerdings muss an der Stelle eingeräumt werden, dass sich bei einigen codierten Verbalisierungen der Lernenden nicht mit Sicherheit sagen liess, ob sie die entsprechenden Teile der Inputtexte auch tatsächlich verstanden und deswegen die richtige Antwort gewählt hatten, oder ob es doch eher ein Zufall war, dass sie sich für diese Antwort und nicht für eine falsche Antwortoption entschieden hatten. Diese Zweifel kommen u.a. bei Bonnie auf. Eine

der Fragen in der von ihr bearbeiteten Aufgabe „Berufswahl“ Seite 2 lautete, ob man im Beruf „représentant commercial“ viel reisen müsse. Bonnie entdeckte im Inputtext von Camille die richtige Stelle „je dois beaucoup voyager en Europe pour mon métier“ („ich muss für meinen Job viel in Europa reisen“) und sagte dazu im Lautdenken-Teil:

der hat etwas geschrieben von viel (.) reisen (--) oder sie? ((zieht 2 nach C2_3rev und wählt C2_3a)) (---) (Bonnie_Off, Pos. 181-182)

Sie wählte die richtige Antwort und machte das auch aufgrund von richtigen Schlüsselwörtern. Im *Stimulated-Recall*-Interview sagte sie dazu:

sie hat (-) auch noch geschrieben dass sie (-) also viel REIST (-) und also (-) in europa vor allem (-) und dann habe ich halt (-) überlegt dass das ja eigentlich zutrifft wenn sie ja viel reist. (Bonnie_Off, Pos. 1008-1011)

Bei der mündlichen Zusammenfassung des Inputtextes im *Stimulated-Recall*-Teil sagte sie zu dieser Textstelle lediglich:

und (-) sie REIST eben viel (-) vor allem in europa (--) ja. (---) (Bonnie_Off, Pos. 1037)

Es lässt sich also nicht zweifelsfrei sagen, ob Bonnie auch verstand, *aus welchem Grund* Camille so viel reist, d.h. ob sie die vielen Reisen tatsächlich für berufliche Zwecke unternehmen muss: Weil die IFB-Aufgaben so konstruiert waren, dass auch die Schlüsselwörter aus den falschen Antwortoptionen in den Inputtexten zu finden waren, hätte es auch sein können, dass die Reisen von Camille nicht mit ihrem beruflichen, sondern mit ihrem privaten Leben zu tun hatten, was in anderen Inputtexten vorkam. Weil in den mit dieser Strategie codierten Stellen die Schlüsselwörter korrekt identifiziert und die gewählte Antwort auch korrekt war, wurde in solchen Fällen trotz der erwähnten Zweifel entschieden, diese Transkriptpassagen als konstruktrelevante Strategie „Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war“ zu codieren.

Strategie „Antwortoptionen miteinander vergleichen“ (TM11)

Mit dieser Strategie wurden Transkriptpassagen codiert, die darauf hindeuteten, dass die Lernenden unter Berücksichtigung des Inputtextes bzw. seiner Teile mehrere Antwortoptionen miteinander verglichen, bevor sie sich für eine (richtige oder auch falsche) Antwortoption entschieden. In den Lautdenkprotokollen wurde diese Strategie aufgrund der Verbalisierungen der Lernenden und/oder des Videos insgesamt 39 Mal codiert, u.a. bei Alison, die bei der Auswahl einer Antwortoption beim Szenario „Stadttrip“, Task A, Seite 2 so vorging, bevor sie sich letztlich für eine (in diesem Fall richtige) Antwortoption entschied:

((S kommt ein bisschen näher an den bildschirm heran, schaut zum text und dann zu den antwortmöglichkeiten)) (5.1) eine boo:tOUr (--) könnte auch sein` (---) einen stAdtrundgang (---) mit einem: smArtphone ((S bewegt die augen zwischen dem text und den antwortmöglichkeiten hin und her)) (15.7) (Alison_Szenario, Pos. 46-50)

In den *Stimulated-Recall*-Interviews wurden 34 Stellen codiert, in denen die Lernenden berichteten, dass sie die Antwortoptionen vor der definitiven Auswahl miteinander verglichen hatten, u.a. Eleven:

beim LETZTen bin ich mir [...] gar nicht mehr so sicher geWESen (-) [...] und dann habe ich [...] nochmal alle durchgelesen und dann habe ich es so auch grad noch mit den ANderen verglichen. (Eleven_Off, Pos. 833-836)

Oft erwähnten die Schüler/innen dabei auch, dass sie sich nicht sicher gewesen waren, welche Antwort die richtige war und dass sie bei der Wahl der Antwortoption neben dem Vergleich der Antwortoptionen auch auf andere (sowohl konstruktrelevante wie auch konstruktirrelevante) Strategien zurückgegriffen hatten. Arya erzählte beispielsweise, dass sie es verwirrend fand, dass im Inputtext zu Szenario „Stadttrip“, Task B, Seite 1 alle Schlüsselwörter aus den Antwortoptionen vorkamen. Dies hatte sie dazu verleitet, die erste Antwortoption zu wählen, zu der sie ein passendes Wort im Inputtext fand (was auch als *test-wiseness*-Strategie „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“, codiert wurde):

irgendwie (-) sie hat ALLES erwähnt was hier in diesen antworten stand; wirklich ALLES; und [...] da habe ich BEInahe ein wenig geschätzt oder geRATEN weil- am anfang kam sie hatte keine internetverbindung und dann sie war in einem kleinen DORF dann war (es) aber wieder im HOTEL und (-) [...] INternet das habe ich NICHT verstanden; und dann (-) wieder (.) ziemlich kla:r und lange von der SIMkarte erzählt- und da wusste ich nicht (2.0) W:AS ist jetzt die richtige antwort` [...] und ja (-) ich habe einfach (-) °h wie die ERste antwort im text genommen; die ich gehört habe- und das war eben das mit dem kleinen dorf und ohne internetempfang; (Arya_Szenario, Pos. 922-934)

Strategie „Auswahl einer Antwortoption, weil andere nicht plausibel sind“ (TM14)

Diese Strategie wurde codiert, wenn den Bemerkungen der Schülerin oder des Schülers zu entnehmen war, dass sie oder er eine Antwortoption oder mehrere Antwortoptionen auswählte, nachdem sie oder er andere Antwortoptionen ausgeschlossen hatte, die ihr oder ihm basierend auf ihrem/seinem Verständnis des Inputtextes als nicht plausibel erschienen waren. Wie bereits bei der Darstellung des Kategorienleitfadens erwähnt (vgl. Kap. 3.7.3.2), könnte man gegen die Klassifizierung dieser Strategie als Testmanagementstrategie einwenden, dass es sich hier eigentlich um eine konstruktirrelevante *test-wiseness*-Strategie handelt, weil sich der/die Schüler/in nicht *für* eine Antwortoption, sondern *gegen* die restlichen Antwortoptionen entschied. In dieser Arbeit wurde diese Strategie dennoch als

konstruktrelevant eingestuft, weil sie impliziert, dass der/die Lernende nicht einfach den Test umgehen wollte, sondern die Frage aufgrund dessen zu beantworten versuchte, was er/sie vom Text verstand (unabhängig davon, ob dieses Verständnis richtig oder falsch war).

In den Lautdenkprotokollen konnte diese Strategie 14 Mal codiert werden. Bei der Durchsicht dieser 14 codierten Segmente fällt auf, dass die Lernenden oft nicht sicher waren, ob die von ihnen letztendlich gewählte Antwortoption auch korrekt war. Auch griffen sie in solchen Fällen bei der Auswahl ihrer Antwort auf weitere konstruktrelevante sowie konstruktirrelevante Strategien zurück. Dies kann z.B. bei Alison beobachtet werden, die beim Lösen von Task B, Seite 2 in Szenario „Stadttrip“ die (falsche) Antwortoption „Die schöne Aussicht vom Schloss“ wählte, u.a. weil sie die zwei anderen Optionen ausschloss. Sie verbalisierte dazu:

```
am ersten tag war sie (-) beim sCHLOss- beim zweiten (1.0) war sie
glaube ich KRAnk (und) vom rEgen (3.0) (hat) sie glaube ich nichts
gesagt- (--)) ((S klickt B2c an)) (Alison_Szenario, Pos. 88-90)
```

Beim Ausschluss der (in ihren Augen) unplausiblen Antwortoptionen scheint hier Alison auch weitere Strategien einzusetzen, die als Testmanagementstrategien „Antwortoptionen miteinander vergleichen“ und „Einsatz von Verstandem bei Auswahl/Ausschluss der Antwortoption(en)“ sowie *test-wiseness*-Strategie „Ausschluss/Auswahl einer Antwortoption bzw. mehrerer Antwortoptionen mit nicht rezipierten oder nicht verstandenen Wörtern“ codiert wurden.

Die Strategie „Auswahl einer Antwortoption, weil andere nicht plausibel sind“ wurde im *Stimulated-Recall*-Teil 21 Mal codiert. So berichtete beispielsweise die Off-Szenario-Schülerin Loup im Interview, dass sie beim Lösen von Task B, Seite 2 in Szenario „Stadttrip“ ähnlich wie Alison vorgegangen war:

```
S: dann habe ich halt verstanden dass [...] das GUTe essen am zweiten
TAG (-) [...] das war es SICHER nicht? (--)) weil sie hat gesagt es
SCHMECKTe nicht gut, (--)) und ähm (-) REgen (-) etwas vom regen habe
ich eben NICHT gehört [...] und dann da/ dachte ich einfach ah (-) das
LETZTe vielleicht; (--)) dann hab ich halt DAS genommen. ((lacht))
MB: hast du vom letzten auch (.) was geHÖRT oder hast du (-) das NUR
genommen weil du die zwei ersten ausgeschlossen hast;
S: ich habe (--)) eigentlich (-) habe ich die zwei er/ äh die zwei
ersten ausgeschlossen und dann dachte ich ja (--)) DAS,
MB: muss DAS sein;
S: ja (-) (Loup_Off, Pos. 734-750)
```

Strategie „Richtigkeit der Antwortoption überprüfen“ (TM10)

Wenn den Bemerkungen der Lernenden oder im Lautdenken-Teil auch dem Video entnommen wurde, dass sie vor oder nach der definitiven Auswahl einer Antwortoption zur Frage und/oder zum Inputtext zurückgingen, um zu prüfen, ob die in Erwägung gezogene

Antwortoption korrekt ist oder nicht, wurde die Strategie „Richtigkeit der Antwortoption überprüfen“ codiert. Ihr Einsatz konnte in 14 Lautdenkprotokollen beobachtet werden, darunter u.a. bei Sheldon und Toby.

```
((S zieht 3 zu C1_2rev zögert lange, schaut währenddessen kurz auf die  
antwortmöglichkeiten und dann auf den text, und lässt das feld dann in  
C1_2rev springen)) (11.2) (Sheldon_Szenario, Pos. 135)
```

```
ALso zuerst hat sie ja gesagt dass das internet nicht funktioniert habe  
(-) aber DANN hat sie noch etwas gesagt dass es zu TEUer wäre mit ihrer  
(-) kaNAdischen simkarte (-) deswegen bin ich jetzt ein bisschen  
verunsichert, (--) hm; ((liest nochmals die antwortoptionen))(13.5)  
(Toby_Szenario, Pos. 89-92)
```

Diese Strategie wurde von 4 Lernenden im *Stimulated-Recall*-Teil erwähnt, davon 3 Mal itemübergreifend. Eleven antwortete z.B. auf die Frage der Forscherin dazu, was er gemacht hatte, wenn er eine passende Stelle im Text gefunden hatte:

```
[*ich habe es zuerst schonmal] REINgetan und dann (-) habe ich es am  
SCHLUSS nochmal angeschaut weil wenn ich vielleicht noch etwas gefunden  
hätte was vielleicht falsch sein könnte hätte ich es noch so ändern  
können.* (Eleven_Off, Pos. 828-831)
```

Strategie „Kenntnisse über das Itemformat nutzen“ (TM15)

Diese Strategie wurde nur bei Szenario „Stadttrip“, Task A codiert. In dieser Aufgabe sollten den vier Kästchen vier zur Wahl stehende Aktivitäten zugeordnet werden, die im Inputtext entweder empfohlen (zwei Aktivitäten) oder nicht empfohlen wurden (ebenfalls zwei Aktivitäten). Es konnte beobachtet werden, dass die Lernenden beim Lösen dieser Aufgabe zuerst zwei empfohlene oder nicht empfohlene Aktivitäten fanden und dann die restlichen zwei quasi automatisch als nicht empfohlen bzw. empfohlen einstufen. Diese Strategie wurde aufgrund der entsprechenden Bemerkungen der Schüler/innen oder auch des Videos codiert, wenn z.B. die nicht empfohlenen Aktivitäten ohne weitere Überlegungen gewählt wurden, nachdem zwei empfohlene Aktivitäten im Inputtext identifiziert worden waren. Diese Strategie wurde hier als konstruktrelevante Testmanagementstrategie (und nicht als konstruktirrelevante *test-wiseness*-Strategie) angesehen, weil es wahrscheinlich dem Aufgabenformat geschuldet war, dass manche Lernende zu lesen aufhörten, nachdem sie zwei Empfehlungen oder zwei Nicht-Empfehlungen gefunden hatten: Auch im realen Leben würde man u.U. so vorgehen, wenn man eine ähnliche Aufgabe zu lösen hätte. Es ist aber ein Grenzfall und man könnte diese Strategie auch als konstruktirrelevant einstufen. Es kann vermutet werden, dass das Hinzufügen von zusätzlichen leeren Kästchen (z.B. 6 oder 8 statt nur 4 leere Kästchen, denen 4 Antwortoptionen zugeordnet werden müssen) den Einsatz dieser Strategie reduzieren würde.

Der Einsatz der Strategie „Kenntnisse über das Itemformat nutzen“ wurde in 6 Lautdenkprotokollen beobachtet. So ermittelte z.B. Cersei richtigerweise, welche zwei Aktivitäten im Inputtext empfohlen wurden, und ordnete dann sofort die restlichen zwei Antwortoptionen der Spalte „nicht empfohlen“ zu:

dann DENke ich (-) ((schaut auf die Antwortmöglichkeiten)) dass sie das hier AUCh (---) empfehlen (.) und die nicht; ((zieht 4 nach Alx, zieht 1 nach Aly, zieht 2 nach Alz)) (3.0) (Cersei_Szenario, Pos. 34-35)

Über den Einsatz dieser Strategie wurde in den *Stimulated-Recall*-Interviews 8 Mal berichtet, davon 7 Mal itemübergreifend. Aria sagte dazu beispielsweise:

und habe es dann so ein bisschen (aus) (--)sortiert also [...] welches sie gUt fanden- oder was sie empFEHlen` und dann ist ja LOGisch dass sie die Anderen zwei nicht empfehlen- (Aria_Szenario, Pos. 386-391)

Strategie „Zeitmanagement“ (TM16)

Eine Transkriptstelle wurde mit dieser Strategie codiert, wenn die Schüler/innen eine Aussage über ihr Zeitmanagement beim Lösen der Testaufgaben machten. Diese Strategie wurde nur in den *Stimulated-Recall*-Interviews codiert, wo sie 3 Mal erwähnt wurde, davon 2 Mal itemübergreifend. Alle mit dieser Strategie codierten Stellen werden im Folgenden dargestellt, weil sie interessante Hinweise für die Testentwickler/innen enthalten.

Eine wichtige Erkenntnis zum möglichen Einfluss des Aufgabendesigns auf die Aufgabenbewältigung stammt aus dem Interview mit Aria. Ihrer Aussage ist zu entnehmen, dass sie sich beim Bearbeiten von Task B in Szenario „Stadtrip“ wegen dem in Abbildung 18 rot markierten Balken unter zeitlichen Druck gesetzt fühlte.

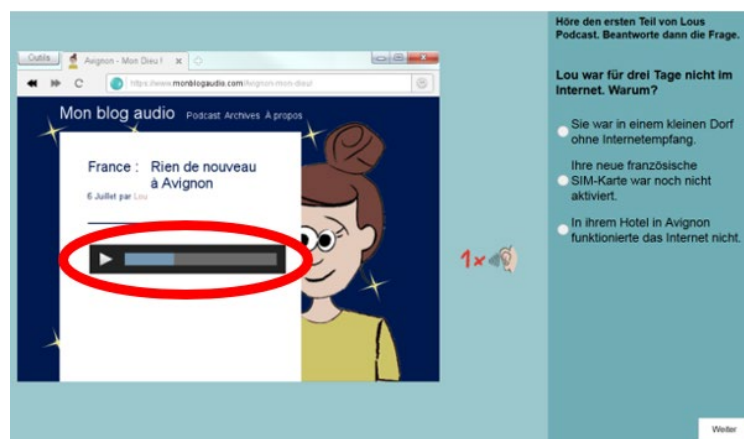


Abbildung 18: Die visuelle Aufmachung der Hörverstehensaufgabe (Szenario „Stadtrip“, Task B).

Der Balken in Abbildung 18, die die Bildschirmseite 1 von Task B zeigt, war statisch und sollte vor allem zur Authentizität der Hörverstehensaufgabe beitragen. Die Länge des Balkens sollte

auch den jeweils abgespielten Teil des Audioblogs abbilden, d.h. der Balken sollte illustrieren, dass man auf der ersten Bildschirmseite den ersten von drei Teilen des Audioblogs hört. Aria ging jedoch davon aus, dass der Balken wie in den Mediaplayer-Programmen dynamisch sei und die Progression des gehörten Textes aufzeige. Dies hatte sie besonders beim Bearbeiten der dritten Bildschirmseite verwirrt, wo der ganze Balken blau markiert war: Sie hatte nicht einschätzen können, wie lange der Hörtext noch dauern würde und ob sie noch Zeit hätte, die richtige Antwort herauszuhören:

```
*[...] Irgendwo war wirklich der ganze balken schon bLAu` und dann (-)
fand ich das schwieriger weil ich da nicht wusste wann es aufhört- weil
sonst wenn ich SEHE wie es so ABLäuft so eine art` habe ich wie das
gefühl ja jetzt (--) also habe ich zum beispiel gedacht jA jetzt habe
ich noch ZEIT zum vielleicht etwas rauszuhören; und wenn man nicht
wEiss ob es jetzt dann gerade fErtig ist` finde ich das schwieriger;*
(Aria_Szenario, Pos. 812-819)
```

Die Aussage von Aria zeigt ausserdem, dass sie es wichtig fände, die Dauer der Inputtexte zu kennen. Bei einem computerbasierten Test könnte es daher überlegenswert sein, die Progression des gehörten Inputtexts tatsächlich dynamisch abzubilden, wie es in Programmen zum Abspielen von Musik oder Filmen üblich ist (was im IFB-Test aus technischen Gründen nicht möglich war).

Dass die Testteilnehmenden es schätzen würden, einen Überblick über die zur Verfügung stehende Zeit zu haben, zeigt auch das Gespräch mit Alison. Beim Lauten Denken konnte beobachtet werden, dass die Schülerin beim Lösen der Testaufgaben die schriftlichen Inputtexte eher schnell und oberflächlich zur Kenntnis nahm, was an einigen Stellen zur Folge hatte, dass sie trotz ihrer schätzungsweise allgemein guten Französischkenntnissen eine falsche Antwortoption wählte. Während des Interviews nahm sie sich aber Zeit für die genauere Lektüre und konnte ihre Fehler meistens auch selber entdecken und korrigieren. Alison wies im Gespräch darauf hin, dass im IFB-Test weder die Information zur abgelaufenen und restlichen Testzeit noch die Testprogression zu sehen waren. Da sie nicht wusste, wie viel Zeit sie sich pro Aufgabe ungefähr einplanen soll und wie viele Aufgaben sie noch zu lösen hatte, versuchte sie, die Fragen möglichst schnell zu beantworten (und das obwohl ihr am Anfang mitgeteilt wurde, dass sie sich für die Testaufgaben so viel Zeit nehmen könne, wie sie dafür bräuchte). Sie hätte es daher besser gefunden, zu wissen, wie viel Zeit sie noch habe und wie viele Aufgaben der Test beinhalte:

```
also vielleicht wenn man (---) jetzt eine zEITvorgabe hat und weiss
wie VIEL (--) ( ) Aufgaben DASS es hAt (---) könnte man auch denken
wie viel zEIT dass man pro aufgabe hAt- und dann hätte ich mir auch
(---) die vOrgegebene zEit genommen; (Alison_Szenario, Pos. 404-407)
```

Alison behauptete ausserdem, dass sie ihre Fehler wahrscheinlich entdeckt hätte, hätte sie im Test zurückgehen und ihre Antworten nochmals überprüfen können. Dies war jedoch im IFB-Test aus technischen Gründen nicht möglich (und wäre im Übrigen beim Lauten Denken ohnehin kontraproduktiv).

jetzt- hIEr kann man es ja eigentlich nicht noch einmal überprüfen wenn man die Anderen aufgaben schon gemacht hat- ich denke bei einer prüfung wenn man es nOch einmal (--) anschauen kann hätte ich es vielleicht gemErkt- die fehler- (Alison_Szenario, Pos. 392-395)

Dem Schüler Toby war hingegen klar, dass er sich beim Lösen der Testaufgaben nicht beeilen musste. Die Frage der Forscherin, ob er auch bei einem Test mit einer wichtigen Note gleich wie beim Bearbeiten der IFB-Aufgaben vorgegangen wäre, bejahte er. Er meinte aber, dass seine Vorgehensweise bei einem wichtigen Test wohl auch von der zur Verfügung stehenden Zeit abhängig wäre. Er fühlte sich beim Lösen der IFB-Testaufgaben zwar wegen der Zeitvorgaben nicht gestresst, gab aber auch zu, dass ihn das Laute Denken etwas ablenkte:

MB: okay (--) ähm (-) hast du alles GLEICH gemacht wie wenn das jetzt (-) ein TEST wäre in der KLASSE wo du eine wichtige NOTE bekommen würdest (-) [...]
S: alles gleich ähm- ((lacht)) äh ja:: (2) es kommt auf die ZEIT (-) drauf an also hier hatte ich jetzt VIEL zeit (-) ähm (-) war aber ein bisschen ABgelenkt von (-) auch noch SPRECHen? (--) äh von DEMher hätte ich vielleicht bei einem (-) TEST es ein bisschen ANDers gemacht (-) auch von der ZEIT her also da war jetzt KEINE zeitlimite? (-)
(Toby_Szenario, Pos. 391-399)

Strategien „Selbstkorrektur F-R“ (falsch-richtig) (TM17), „Selbstkorrektur R-F“ (richtig-falsch) (TM18) und „Selbstkorrektur F-F“ (falsch-falsch) (TM19)

Diese Strategien wurden dann codiert, wenn aufgrund des Videos und/oder der Bemerkungen der Schüler/innen beobachtet wurde, dass sie ihre bereits gewählte Antwort nochmals wechselten. Was beim Einsatz dieser drei Strategien jeweils anders war, war das Endergebnis: Die Strategie „Selbstkorrektur F-R“ (falsch-richtig) wurde dann codiert, wenn die zuerst gewählte Antwortoption falsch und die korrigierte richtig war; im Fall der Strategie „Selbstkorrektur R-F“ (richtig-falsch) war es genau umgekehrt; und die Strategie „Selbstkorrektur F-F“ (falsch-falsch) wurde dann codiert, wenn die Lernenden zuerst eine falsche Antwortoption wählten, um danach eine andere, ebenfalls falsche Antwortoption anzuklicken. Alle drei Strategien wurden als konstruktrelevant eingestuft, weil die Lernenden bei ihrem Einsatz nicht den Test umzugehen versuchten, sondern sich darum bemühten, die Testaufgaben aufgrund ihres (im Fall der Strategien „Selbstkorrektur R-F“ und „Selbstkorrektur F-F“ jedoch evtl. falschen oder unzureichenden) Textverständnisses zu lösen.

Die Strategie „Selbstkorrektur F-R“ wurde in den Lautdenkprotokollen an 11 Stellen codiert. Toby wählte z.B. bei Szenario „Stadttrip“, Task A, Seite 1 zuerst eine falsche Antwort: Er meinte, die geführte Stadiontour würde im Inputtext empfohlen. Dann las er aber die Textstelle genauer durch und korrigierte seine Antwort, nachdem er bemerkte hatte, dass die Stadiontour wegen ihres hohen Preises doch nicht zu den empfohlenen Aktivitäten gehörte:

```

ähm (-) ((schaut auf den text)) ANNE sagt aber dass das (-) NEUe stadion
(-) also sportstadion aber sehr (-) äh GUT sei und man müsse unbedingt
eine (-) eine TOUR machen (-) eine geFÜHrte tour? ((zieht 2 nach
Alw))(4.0) ((liest den text weiter)) (12.4) ((schaut auf die
antwortoptionen)) (1.5) ((liest den text)) (2.1) ah nein (-) FALSCH
(-) ähm es sei zu TEUer? ((zieht 2 von Alw nach Alz)) (--)
(Toby_Szenario, Pos. 22-30)

```

In den *Stimulated-Recall*-Interviews wurde diese Strategie 3 Mal erwähnt, davon 2 Mal von Toby. Er erinnerte sich im Interview daran, dass er sich bei Szenario „Stadttrip“, Task A, Seite 1 für eine Antwort zu schnell entschieden hatte, d.h. ohne den Rest des Inputtextes zur Kenntnis zu nehmen. Das hatte er korrigiert, als er es bemerkt hatte. Er meinte gleichzeitig, dass er vielleicht auch an anderen Stellen solche Fehler gemacht hatte, die ihm dann aber entgangen seien.

```

S: äh das war der FEHler ich <<lachend>habe> ich habe es äh soFORT
eingefügt und NICHT fertig gelesen eben und dann eins zwei mal war es
wieder dann ANders im zweiten satz?
MB: dann hast du dich sofort korriGIERT;
S: und dann ja: (-) DAS das ich gemerkt habe aber ich denke es sind
VIEle die (-) ich nicht bemerkt habe. ((lacht)) (Toby_Szenario,
Pos. 459-464)

```

Die beiden Strategien „Selbstkorrektur R-F“ (richtig-falsch) und „Selbstkorrektur F-F“ (falsch-falsch) wurden nur in den Lautdenkprotokollen codiert. In 5 Fällen war die ursprünglich gewählte Antwort richtig und die korrigierte falsch. In 3 Fällen kam es hingegen vor, dass sich die Lernenden sowohl bei der Auswahl der ursprünglichen Antwortoption wie auch bei der Korrektur für einen Distraktor entschieden. Den mit diesen beiden Strategien codierten Transkriptausschnitten war zu entnehmen, dass sich die Lernenden bei der Wahl ihrer Antworten nicht sicher waren, ob sie den Inputtext richtig verstanden (was auch stimmte), wie z.B. Cersei (Strategie „Selbstkorrektur R-F“) beim Bearbeiten von Szenario „Stadttrip“, Task A, Seite 2. Sie wählte zuerst die richtige Antwortmöglichkeit (d.h. die Option, dass die Anschaffung der „Carte Avignon“ empfohlen würde). Weiter im Text stand: „La carte n'est pas valable sur les bateaux, mais c'est pas grave.“ („Die Karte ist auf Booten nicht gültig, aber das ist nicht schlimm.“). Cersei bemerkte, dass über die Karte etwas Negatives gesagt wurde und wechselte daraufhin ihre Antwort. Sie verstand aber nicht bzw. überlas den weiteren Textteil, in dem stand, dass die Bootstouren langweilig wären und es deswegen nicht schlimm wäre, dass die Karte auf Booten ungültig ist.

da (.) schreiben sie wieder von (-) der karte (.) u::nd (-) ((schaut auf die antwortmöglichkeiten)) ich DENke dass die DIE dem empfehlen; ((zieht 2 nach A2w)) ((liest den text)) (5.3) aber da UNTen (2.0) ((schaut kurz auf die antwortmöglichkeiten und dann wieder auf den text, zeigt mit der maus auf "la carte n'est pas valable")) dass sie NICHT gut ist; ((liest den text und zeigt mit der maus auf die zweite nachricht)) (7.3) also die ist dann ((schaut auf die aufgabenstellung)) eher KEine empfehlung. [...] ((S zieht 2 nach A2y)) (Cersei_Szenario, Pos. 42-53)

Strategie „Selbstkorrektur beim SR“ (Stimulated Recall) (TM20)

Die Codierung der Strategie „Selbstkorrektur beim SR“ wurde, wie schon mit der Bezeichnung angezeigt, nur für den *Stimulated-Recall*-Teil vorgesehen. Bereits bei der Datenerhebung konnte beobachtet werden, dass die erneute und für einige Schüler/innen auch intensivere Auseinandersetzung mit den Inputtexten und Aufgaben dazu führte, dass sie nun mehr oder etwas anderes als bei der Bearbeitung der Testaufgaben während des Lautdenken-Teils verstanden. Auch sagten sie im Interview, dass sie ihre Antworten jetzt evtl. korrigieren würden bzw. dass sie sich nicht mehr sicher seien, ob die früher gewählte Antwort tatsächlich stimmen würde. Mit diesem Code wurden in den Interviews 29 Stellen gekennzeichnet.

Die Gründe dafür, dass die Schüler/innen bei der nochmaligen Kenntnisnahme der Inputtexte ihre Antworten anpassten, waren vielfältig. Es kam vor, dass die Lernenden bei der Bearbeitung der Testaufgaben die Antworten sehr schnell wählten, d.h. ohne den vollständigen Inputtext gelesen zu haben. Dies geschah u.a. Alison bei Szenario „Stadttrip“, Task C, Seite 1: Eine der Fragen zu diesem Inputtext lautete, ob es im Zimmer gratis WLAN gäbe. In der entsprechenden Textpassage stand: „Le Wi-Fi est gratuit à la réception, mais comme nous voulions une connexion dans notre chambre, nous avons payé 15€ de plus pour la semaine.“ (Wi-Fi ist an der Rezeption kostenlos, aber da wir eine Verbindung in unserem Zimmer haben wollten, haben wir 15€ mehr für die Woche bezahlt.“). Erst während des Interviews bemerkte Alison, dass sie im Test die falsche Antwort gewählt hatte, weil sie damals nur den Anfang des Satzes gelesen hatte:

dann habe ich hier Wifi (-) ist gratis (-) an der rezeptION- ((S schaut auf den bildschirm)) (9.8) ja; jetzt hätte ich noch WEiterlesen sollen wEil ja unten steht dass es im zIMmer fünfzehn euro mEHr kostet (-) in der woche` [...] ((S blickt auf den bildschirm)) (6.2) ja. also wenn man: bei der rezeptION ist Ist es grAtis (-) aber wenn man im zimmer auch wlan will muss man bezahlen. (Alison_Szenario, Pos. 1069-1079)

Bei der Besprechung der Testaufgaben wurden die Lernenden gebeten, das Verstandene in ihrer Schulsprache bzw. in ihrem Dialekt zusammenzufassen. Dafür konnten sie die Inputtexte nochmals lesen oder auch nochmals hören (wobei sie diesmal die Audioaufnahme an einer beliebigen Stelle stoppen konnten). Durch diese wiederholte und auch intensivere Auseinandersetzung mit den Inputtexten verstanden viele Lernende mehr als während des

eigentlichen Tests. Dies war z.B. bei Howard zu beobachten, der beim nochmaligen Hören des Inputtextes zu Szenario „Stadttrip“, Task B, Seite 2 richtigerweise bemerkte, dass die von ihm gewählte Antwortoption „Die schöne Aussicht vom Schloss (hat die Bloggerin Lou besonders beeindruckt)“ gar nicht stimmen konnte, weil Lou das Schloss nur vom Bus aus sah:

```
((S hört weiter)) äh AH ((fast sich mit der hand an die stirn)) ich
habe das CHAteau durch das fenster vom BUS gesehen (-) hat sie gesagt
(-) ah nein shit ((lacht)) das habe ich gar nicht richtig (-) ja. ((S
hört weiter)) je pense (-) ich GLAUbe es es äh (-) es WÄRe eine
megaschöne aussicht von dort (-) nein. ((grinst)) (Howard_Szenario,
Pos. 1089-1095)
```

Es gab aber auch umgekehrte Fälle, wo die Schüler/innen bei der Bearbeitung der Testaufgaben eine richtige Antwort wählten, im *Stimulated-Recall*-Interview aber behaupteten, ihre ursprüngliche Antwort wäre falsch gewesen. Elena wählte z.B. bei Szenario „Stadttrip“, Task B, Seite 1 die richtige Antwort „In ihrem Hotel in Avignon funktionierte das Internet nicht“. Beim nochmaligen Hören des Textes meinte sie dann, verstanden zu haben, dass das Internet im Hotel doch funktionieren würde und dass sie hier einen Fehler gemacht hätte.

```
((S hört weiter)) (4.0) das interne äh (-) funktioNIERT im hotel (-)
*also ist das mal falsch;* (Elena_Szenario, Pos. 798-800)
```

Im Interview mit Elena stellte sich weiter auch heraus, dass sie im Szenario „Stadttrip“, Task C, Seite 1 nun ihre Antwort ändern würde, weil sie beim Lösen dieser Aufgabe von falschen Prämissen ausgegangen wäre, da sie die Aufgabeninstruktionen nicht oder nicht genügend zur Kenntnis genommen hätte. In dieser Leseverstehensaufgabe mussten 1) pro Bildschirmseite drei Fragen zu drei Inputtexten mit Hotelreviews mit „ja“ oder „nein“ beantwortet werden und 2) es musste auch angegeben werden, in welchem der drei Reviews die Antwort gefunden wurde. In der Instruktion stand dazu: „Manchmal findest du in der gleichen Bewertung mehrere Antworten.“ Diesen Satz hatte Elena nicht gelesen und gedacht, dass sie in dieser Aufgabe alle drei Reviews den drei Fragen zuordnen müsse. Nachdem sie beim Testlösen bereits zwei Reviews den Fragen zugeordnet hatte, ordnete sie automatisch der dritten Frage das restliche Review zu und kreuzte dort „nein“ an, weil sie in diesem Text keine Antwort auf die Frage gefunden hatte. Nachdem sie im Interview darauf hingewiesen wurde, dass sie das gleiche Review auch mehrmals den Fragen hätte zuordnen können, sagte sie (richtigerweise), dass sich die Antwort auf die gerade besprochene Frage im Review von Thérèse befände, das sie bereits (fälschlicherweise) einer anderen Frage zugeordnet hatte:

```
S: *was sie jetzt gesagt haben dass eine person zweimal sagen kann das
habe ich nicht gewusst; (--) vielleicht ist es ja dann (-) sollte man
das irgendwo hinschreiben oder (--) nicht;*
```

MB: hast du das dann AUCH gelesen? ((zeigt auf aufgabenstellung))
S: *nein (-) nicht unbedingt. ((S liest aufgabenstellung)) (2.8) ja
(-) das habe ich NICHT gelesen; ((lacht)) [...] ich habe gedacht [...] ein
TEXT gehört zur person (-) aber jetzt (-) zum beispiel THereSe so es
können ja zwei texte gehören (-) jetzt bin ich mir sogar sicher (-)
dass sie zu DEM gehört;* (Elena_Szenario, Pos. 1060-1075)

Resümee

Insgesamt konnten in den Daten 19 bzw. 20 Testmanagementstrategien festgelegt und codiert werden. Sie machten ca. 57% der in den Lautdenkprotokollen und etwa 60% der in *Stimulated-Recall*-Interviews codierten Strategien aus und liessen sich anschliessend in A) Strategien zur Auseinandersetzung mit den Instruktionen und/oder Itemelementen, B) Strategien zur Auseinandersetzung mit dem Inputtext, sowie C) Strategien zur Vorgehensweise bei der Wahl der Antwortoption(en) einteilen.

Der hohe Anteil der codierten Testmanagementstrategien spricht dafür, dass sich die Schüler/innen mit den Inputtexten jeweils bewusst im Hinblick auf die zu lösenden Aufgaben auseinandersetzten. Wie schon im Fall der Lernerstrategien konnte auch bei den Testmanagementstrategien beobachtet werden, dass das erfolgreiche Lösen der Testaufgaben unter dem Einsatz dieser Strategien in erster Linie davon abhängig war, wie viel die Schüler/innen vom französischen Inputtext verstanden. Auch war hier von Bedeutung, wie aufmerksam die Lernenden die Inputtexte zur Kenntnis nahmen: Es zeigte sich, dass oberflächliches und schnelles Vorgehen beim Lösen der Testaufgaben auch unter dem Einsatz von konstruktrelevanten Strategien nicht unbedingt zum Erfolg führte.

Im nächsten Kapitel werden die konstruktirrelevanten *test-wiseness*-Strategien vorgestellt, die in den Daten beobachtet und codiert wurden.

4.2.2.2 TEST-WISENESS-STRATEGIEN

In den Lautdenkprotokollen wurden (nach Abzug der Codierungen mit Gewichtung „2“, also der Codierungen mit der wiederholten Aussage zu einer bereits codierten Strategie) im Ganzen 109 und in den *Stimulated-Recall*-Interviews 147 *test-wiseness*-Strategien codiert. Dies entspricht ca. 8% der in den Lautdenkprotokollen und ca. 16% der in den *Stimulated-Recall*-Interviews beobachteten Strategien. Die Häufigkeiten der jeweiligen Strategien samt Gewichtungen sind Tabelle 16 zu entnehmen. Im weiteren Teil dieses Kapitels werden die codierten Strategien beschrieben und dort, wo es sich anbietet, auch daraus folgende Konsequenzen für die Testkonstruktion angesprochen.

Tabelle 16: Häufigkeit der test-wiseness-Strategien inkl. Gewichtungen.

Kurzbezeichnung Code	Code LD	Anzahl Codierungen in LD-Protokollen (n=20)	Gewichtung „2“ ⁹⁷	Gewichtung „5“ ⁹⁸	Anzahl Codierungen in LD-Protokollen nach Abzug Codierungen mit Gew. „2“	Code SR	Anzahl Codierungen in SR-Interviews (n=20)	Gewichtung „2“	Gewichtung „5“	Anzahl Codierungen in SR-Interviews nach Abzug Codierungen mit Gew. „2“
Raten	LD-TW01	8	0	2	8	SR-TW01	26	1	6	25
Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war	LD-TW02	72	1	0	71	SR-TW02	78	6	2	72
Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern	LD-TW03	17	0	0	17	SR-TW03	27	0	2	27
Auswahl einer Antwortoption mit bekannten Wörtern	LD-TW04	0	0	0	0	SR-TW04	4	0	0	4
Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen	LD-TW05	6	0	1	6	SR-TW05	6	0	1	6
Auswahl einer anders aussehenden Antwortoption	LD-TW06	0	0	0	0	SR-TW06	2	0	0	2
Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats	LD-TW07	7	0	0	7	SR-TW07	12	1	0	11
Total		110	1	3	109		155	8	11	147

⁹⁷ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

⁹⁸ Eine allgemeine(re) Aussage zum eigenen Verhalten beim Lösen der Testaufgaben, die aufgaben-, item- oder textstellenübergreifend ist und nicht nur das gerade besprochene Item oder die betroffene Aufgabe oder Textstelle betrifft.

Strategie „Raten“ (TW01)

Mit „Raten“ wurden in den Transkripten solche Stellen codiert, die darauf hindeuten, dass die Lernenden bei der Auswahl der Antwortoptionen intelligent oder auch zufällig geraten hatten. Durch die Testaufmachung wurden die Testteilnehmenden zur Wahl einer Antwort gezwungen, da sie im Test nicht hätten weiter machen können. Sie waren sich aber nicht sicher, ob die gewählte Antwort auch tatsächlich stimme, weil sie den Inputtext entweder nicht oder nicht aufmerksam genug zur Kenntnis genommen hatten; oder aber nichts davon bzw. nicht genug davon verstanden hatten, um eine Antwort aufgrund des Verstandenen zu wählen.

Diese Strategie wurde in den Lautdenkprotokollen 8 Mal codiert, davon 2 Mal mit der Gewichtung „5“ (itemübergreifend). Sie wurde u.a. beim Schüler Steve beobachtet, der beim Bearbeiten der Aufgabe „Berufswahl“, Seite 2 die (falschen) Antwortoptionen zumindest teilweise zufällig gewählt zu haben schien, weil er den Inputtext von Camille nicht verstand:

```
(von) diesem tExt weiss ich (.) Nichts; ((cursor auf text von camille))  
((zieht 1 zuerst zum antwortfeld C2_2rev)) (4.0) (unv.) (vielleicht)  
mehrere (spr/) ((platziert 1 dann in C2_1rev)) (---) viele produkte  
testen ((klickt auf C2_1a)) (1.1) (Steve_Off, Pos. 85-89)
```

„Raten“ wurde auch bei Schülerin Loup codiert, die beim Lösen von Szenario „Stadttrip“, Task B, Seite 2 Folgendes verbalisierte:

```
ALso sie war KRANK (-) malade (-) das haben wir gerade in der SCHULE  
gelernt? (--) malade also (-) ja (-) und (-) ähm sie hat (-) ähm das  
ESSen (-) hat sie NICHT so (-) LECKER gefunden? (--) und (-) das mit  
dem (.) REgen (-) pleut (-) habe ich gar nicht gehört? (--) hm ich  
weiss jetzt NICHT ähm (-) die schöne AUSSicht vielleicht (-) schloss?  
((wählt B2c und klickt "weiter" 2x)) (Loup_Off, Pos. 68-74)
```

Dem Transkriptausschnitt ist zu entnehmen, dass Loup neben (intelligentem) Raten auch weitere Strategien einsetzte, die hier als die *test-wiseness*-Strategie „Ausschluss/Auswahl einer Antwortoption bzw. mehrerer Antwortoptionen mit nicht rezipierten oder nicht verstandenen Wörtern“, die Testmanagementstrategien „Einsatz von Verstandenen bei Auswahl/Ausschluss der Antwortoption(en)“ und „Antwortoptionen miteinander vergleichen“ sowie die Lernerstrategie „Intensive Auseinandersetzung mit Teilen des Inputtextes“ codiert wurden.

Die beiden mit der Gewichtung „5“ gekennzeichneten Codierungen der Strategie „Raten“ stammen aus dem Lautdenkprotokoll von Elena. Sie schien sich für diese Strategie entschieden zu haben, weil sie nicht besonders motiviert war, sich mit den Inputtexten vertieft auseinanderzusetzen:

manchmal rate ich einfach weil ich keine lust habe das alles zu lesen.
 (5.0) ((zieht 1 nach C2_1rev, bewegt den cursor zwischen a und b hin und her und wählt dann C2_1a, klickt "weiter")) (Elena_Szenario, Pos. 88-89)

Beim Lösen der dritten Bildschirmseite von Task C, Szenario „Stadttrip“ gab Elena auch zu, dass sie bei dieser Aufgabe manchmal nur den Inputtext den Fragen zuordnete (z.B. dass ihrer Meinung nach die Antwort auf die Frage „Gibt es im Hotel gratis WLAN“ im Inputtext von Thérèse vorkommt) und sich dann beim Anklicken der Antwortmöglichkeiten „ja“ oder „nein“ zufällig für etwas entschied:

manchmal lese ich auch nur die texte (-) also textteile (-) und dann (-) wenn es (-) wenn ich weiss von wem es ist dann rate ich die andern zwei ((zeigt mit der maus auf "ja" und "nein" in der aufgabenstellung)) einfach. (Elena_Szenario, Pos. 96-98)

Elena scheint übrigens mit dieser Strategie nicht viel Erfolg gehabt zu haben: Nur eine ihrer sechs Antworten auf die Fragen zu den Bildschirmseiten 1 und 2 in Task C war korrekt.

Im *Stimulated-Recall*-Teil wurde die Strategie „Raten“ 25 Mal (davon 6 Mal itemübergreifend) erwähnt. Gemäss den Interviewäusserungen der Lernenden griffen sie dann auf diese Strategie zurück, wenn sie die Inputtexte nicht richtig verstanden. Sie waren sich dann auch nicht sicher, ob die von ihnen gewählte Antwort tatsächlich stimmen würde. So berichtete beispielsweise Omega über den Einsatz dieser Strategie beim Wählen einer der Antwortmöglichkeiten bei Szenario „Berufswahl“, Seite 1:

un:d produkte habe (-) ich vom "PRODUITS" (-) irgendwo- da war mich (-) war ich mir halt das NICHT sicher ob sie es (-) selber testet oder (-- es getestet wird` [und dort habe ich einfach] geraten; (Omega_Off, Pos. 914-917)

Steve gab im Interview an, dass er von den Inputtexten nicht viel verstanden hatte und deswegen fast überall im Test (in 9 von 10 Items übrigens erfolglos) hatte zufällig raten müssen:

S: hier habe ich noch weniger verstanden; [...]
 MB: hast du ALLES geraten` oder [...] hast du IRGENDwelche (--) indizien ((MB lacht kurz)) gehabt-
 S: ich habe beinahe alles geraten; ich habe recht viel geraten; [...]
 MB: mh_mh °h also bei diANA hast du gesagt dass sie (--) selbst neue projekte (-) eh produkte testen muss- [...] hast du IRGENDetwas (-) drin gehabt` ein wort (-) das-
 S: nein.
 MB: dir das verraten hat- nichts`
 S: nein; ich habe recht (3.0) gerAten; (Steve_Off, Pos. 893-917)

Andere Schüler/innen setzten hingegen neben dem Raten auch andere Strategien ein. Der Schüler Mogli hatte z.B. beim Lösen von Szenario „Stadttrip“, Task B, Seite 1 zuerst eine

Antwortmöglichkeit ausgeschlossen und sich dann bei der Wahl einer der zwei restlichen Antwortmöglichkeiten auf sein Bauchgefühl verlassen:

```
S: *Und nachher (-) ja: habe ich mehr oder weniger eigentlich geraten;*  
((S lächelt))  
MB: okAy [...] also hast du wIRKlich geraten gerATen` oder hast du auch  
(-) DOCH einiges verstanden`  
S: *ja so mit dem HOTel (.) [...] die Erste frage habe ich Ausgeschlossen  
also da habe ich gedacht die ist es sICher nicht; aber sie hat von der  
SIMkarte gesprochen und dass es TEUER ist (-) und aber dass sie  
irgendwie "l'Internet c'est casse" also irgendwie so-[...] und dann habe  
ich gedacht (ich) rAte und dann habe ich gedacht JA es ist das;*  
(Mogli_Off, Pos. 660-674)
```

Auch für die Schülerin Picard war Raten dann eine (letzte) Option, wenn sie die richtige Antwort nicht ermitteln konnte, weil sie den Text nicht ganz verstand:

```
wenn es ein wort gibt das ich wIRKlich nicht verstehe dann schAue ich  
(-) em (-) wAs es heissen könnte` weil es Irgendeinen zusammenhang  
geben MUSS- [...] und wenn ich nicht darauf komme dann (-) rAte ich  
(Picard_Off, Pos. 357-361)
```

Strategie „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“ (TW02)

Diese Strategie könnte auch als eine Variante des intelligenten Ratens bezeichnet werden. Sie wurde dann codiert, wenn den Verbalisierungen der Lernenden zu entnehmen war, dass sie sich deswegen für eine Antwortoption entschieden, weil ein Wort oder mehrere Wörter aus den Antwortoptionen auch im Inputtext in der gleichen oder einer ähnlichen Form vorkamen. Dabei wurde aber von ihnen die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigt bzw. nicht verstanden. Die Strategie wurde (im Gegensatz zur ähnlichen Testmanagementstrategie „Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war“) deswegen als konstruktirrelevant eingestuft, weil die Schüler/innen bei ihrem Einsatz nicht die kognitiven Prozesse durchliefen, die durch das Item vorgesehenen waren. Diese Strategie wurde daher sowohl bei den falschen wie auch bei den richtigen Antworten codiert, wenn 1) die Antwort aufgrund eines unpassenden Wortes gewählt wurde, welches sowohl in der Antwortoption als auch dem Inputtext vorkam; oder 2) es klare Indizien dafür gab, dass die Schüler/innen den Text um das gefundene Wort herum nicht verstanden oder gar nicht gelesen hatten.

Diese Strategie gehört zu den mit Abstand am häufigsten beobachteten *test-wiseness*-Strategien. In den Lautdenkprotokollen wurde sie 71 Mal codiert. Sie wurde u.a. von Howard angewandt: Nachdem der Schüler im Inputtext von Miguel (Szenario „Stadttrip“, Task C, Seite 1) die Stelle „au centre-ville d'Avignon“ („im Zentrum von Avignon“) entdeckte, die in seinen Augen die positive Antwort auf die Frage „Ist das Hotel im Stadtzentrum?“ beinhaltete,

wählte er sofort diese Antwortoption und nahm den Rest vom Inputtext nicht genau zur Kenntnis:

```
((zieht 2 nach C1_2rev)) (---) ähm also (-) miguel hat ja gesagt  
(---) äh dass es im (-) centre (.) ville de avignon ist also ist es  
(-) im STADTzentrum; ((wählt C1_2a)) (Howard_Szenario, Pos. 95-98)
```

Der ganze Satz lautete aber „J'ai passé une nuit ici avec un ami, après un concert au centre-ville d'Avignon, en novembre.“ („Ich habe hier mit einem Freund übernachtet, nach einem Konzert im Zentrum von Avignon, im November.“). Die Strategie von Howard ging folglich (zumindest an dieser Stelle) nicht auf: Nicht das Hotel, sondern das letztjährige Konzert war im Zentrum. Einige Schüler/innen, bei denen diese Strategie beobachtet und codiert wurde, lasen dann aber später den Text nochmals durch und passten ihre Antworten anschliessend an. Unter anderem ging Arya an der Stelle aus dem obigen Beispiel zuerst gleich wie Howard vor, um dann nach der näheren Lektüre ihre Antwort (richtigerweise) anzupassen:

```
*dann würde ich jetzt mal DAs nehmen (.) ((blickt auf den text))  
stadtzentrum ((zieht 2 in C1_2rev)) ((S schaut auf die  
aufgabenstellung)) ((klickt auf C1_2a)) (-) und das ist JA. [...]  
((schaut auf den text)) jetzt bin ich bei nummer zwei nicht sicher (.)  
(noch einmal lesen)- ((S liest den text, cursor auf text 3 danach auf  
text 2)) (15.0) °h ich glaube (.) ((blickt auf die aufgabenstellung))  
das hier ist (.) die (unv.) für das nicht das andere ((zieht 3 in  
C1_2rev)) (-) weil HIER (.) ((S blickt auf den text, cursor zuerst auf  
text 2, dann auf text 3)) sieht man °hh dass es (im herzen der stadt)  
ist und nicht (-) dass sie nicht im (-) ((klickt auf "weiter")) konzert  
waren.* (Arya_Szenario, Pos. 154-165)
```

Der (meist erfolglose) Einsatz dieser Strategie konnte sehr oft bei Schüler/innen beobachtet werden, die die französischen Inputtexte nicht sehr gut verstanden. Cersei verbalisierte z.B. beim Lösen der Aufgabe Szenario „Stadttrip“, Task C, Seite 2 Folgendes:

```
((liest nochmal abschnitt von roberta und zeigt mit der maus, wo sie  
ist)) (12.0) ((schaut auf den text von roberta)) also jetzt INternet  
sans probleme (2.0) dass es ähm (-) ((schaut auf die antwortoptionen))  
ein proBLEM gibt ((zieht 2 nach C2_1rec)) dann denke ich dass das  
internet NICHT ((wählt C2_1b)) (-) GRAtis ist. (Cersei_Szenario,  
Pos. 173-176)
```

Sie wählte also die (falsche) Antwort aufgrund des Fragments „et il y avait une connexion internet sans fil“ („es gab eine drahtlose Internetverbindung“) aus dem Inputtext von Roberta, den sie offensichtlich falsch verstanden hatte.

Dank dem Einsatz dieser Strategie konnten aber auch richtige Antworten gefunden werden. In einigen Fällen war Zufall im Spiel, wie z.B. bei Bonnie, die bei Szenario „Berufswahl“, Seite 1 den richtigen Text mit der Antwort auf die Frage „Muss man im Beruf „représentant commercial“ neue Produkte testen?“ wegen dem Wort „tester“ („testen“) auswählte, aber beim Anklicken von „ja“ oder „nein“ raten musste, womit sie Erfolg hatte:

*dann das (-) neue proDUKTe (.) testen ((zieht 3 nach Cl_1rev)) (--)
 aber (-) da würde ich jetzt (.) NEIN tippen.* (2.0)((klickt weiter))
 (Bonnie_Off, Pos. 139-140)

Im Interview sagte Bonnie zu dieser Stelle:

*also (-) die ERSTe (-) neue produkte herstellen (-) ähm (---) also
 (-) also da oben (-) zuoberst in der linie steht etwas von testen?
 (--) dann habe ich mir halt überlegt ja vielleicht könnte es ja DAS
 sein (-) und dann hat er aber auch wieder etwas von rien also von
 NICHTS erzählt und da war ich mir nicht mehr (-) ganz und dann habe
 ich aber gedacht ja trotzdem es könnte ja vielleicht DAS sein. (--)
 also es war eher geraten.* (Bonnie_Off, Pos. 841-848)

Solch eine zufällige Wahl der richtigen Antwortoption(en) kann bei geschlossenen Antwortformaten wohl nie vermieden werden. Bei zwei Items lag der Erfolg dieser Strategie aber an den Inputtexten, die aufgrund der Erkenntnisse aus den Lautdenkprotokollen von Sheldon, Alison und Howard, über die im Folgenden berichtet wird, angepasst werden müssten. Eine der Fragen im Szenario „Stadttrip“, Task C, Seite 1 lautete, ob das Hotel im Zentrum lag. Beim Lösen dieser Aufgabe stiess Sheldon an folgende Stelle im Inputtext von Curt: „En général, l'hôtel n'était pas mal...“ („Im Allgemeinen war das Hotel nicht schlecht...“) und überlegte dazu Folgendes:

ist das hotel I:m (1.5) zEntrum ((schaut auf den text)) (3.5) das WEIss
 ich n:ICht (.) aber ich denke "en general" hOtEl heisst vielleicht
 (--) das generAlhotel und das generalhotel ist vielleicht ((blickt nach
 oben)) in der mitte der sTADT also denke ich ((schaut auf die
 aufgabenstellung)) könnte es sEin; (--) (Sheldon_Szenario,
 Pos. 140-143)

Sheldon beantwortete die Frage mit „ja“ und gab auch an, dass er die Antwort im Text von Curt fand. Seine Antwort war korrekt, aber aus einem falschen Grund: Die Stelle im Inputtext, die die Antwort beinhaltete, lautete nämlich: „L'hôtel offre une belle vue sur le Rhône et il se trouve en plein cœur d'Avignon.“ („Das Hotel bietet einen schönen Blick auf die Rhone und liegt im Herzen von Avignon.“). Eine ähnliche Situation konnte bei Alison beobachtet werden, die beim Lösen von Szenario „Stadttrip“, Task C, Seite 2 zur gleichen Frage und wie Sheldon die richtige Antwort aus dem falschen Grund wählte. Sie identifizierte richtig, dass sich die positive Antwort auf die Frage, ob das Hotel im Zentrum liege, im Text von Kelly befinde. Der Text lautete: „L'année dernière, on était dans un hôtel à côté de la Gare Centrale ... qui n'est pas du tout centrale! L'hôtel du Sud est situé juste à côté de la mairie, on a fait tout le centre à pied.“ („Letztes Jahr waren wir in einem Hotel neben dem Hauptbahnhof... der überhaupt nicht zentral ist! Das Hôtel du Sud befindet sich direkt neben dem Rathaus, wir sind durch das ganze Zentrum zu Fuss gegangen.“). Die Antwort war hier folglich dem zweiten Satz zu entnehmen. Alison meinte jedoch fälschlicherweise, dass das Hotel im Zentrum sei, weil es am Bahnhof liege:

das hotEl ist neben dem: (--) bAhnHOf- das im zEntrum- (3.0) ((S zieht 3 in C2_2rev)) (2.5) ((S klickt C2_2a an)) (1.8) (Alison_Szenario, Pos. 122-124)

Diese Strategie wurde auch von Howard an der gleichen Stelle angewandt.

DA sagt er es ist ähm (-) beim GARE centrale also (--) und der ist meistens im zentrum (--) ((zieht 3 nach C2_2rev und wählt C2_2a)) also (-) würde ich mal sagen das ist (-) der; (---) (Howard_Szenario, Pos. 117-119)

Es gab aber auch Schüler/innen, darunter Cybermen, die den Satz zu Ende lasen und sich gerade deswegen für die negative (und somit falsche) Antwort entschieden:

dann hier "n est pas (.) tout (.) centrale" (--) ((schaut auf die antwortmöglichkeiten)) zentral (-) das ist halt mitte- (--) also nein ((wählt C2_2b an und zieht 3 nach C2_2rev)) (Cybermen_Szenario, Pos. 144-146)

Aus Sicht der Testkonstruktion sprechen die obigen Auszüge aus den Lautdenkprotokollen dafür, die Sätze, die Sheldon, Alison und Howard zur Wahl der richtigen Antwortoptionen aus falschen Gründen verleiteten, zu überarbeiten. Wenn man diese Sätze z.B. in einen der zwei weiteren Inputtexte auf dieser Bildschirmseite verschieben würde, würden sie noch als gute Distraktoren dienen.

In den *Stimulated-Recall*-Interviews wurde diese Strategie 72 Mal codiert (davon 2 Mal aufgabenübergreifend), also fast ebenso häufig wie in den Lautdenkprotokollen. Bei einigen Schüler/innen, die eigentlich zu den stärkeren Lernenden aus der Stichprobe gehörten (wie Alison und Howard), zeigte sich dabei, dass sie auf diese Strategie zurückgriffen, weil sie die Inputtexte beim Lösen der Testaufgaben nur schnell und oberflächlich zur Kenntnis genommen hatten. Bei der Besprechung dieser Aufgaben im Interview konnten sie ihre Fehler selber entdecken und korrigieren. Andere Schüler/innen setzten diese Strategie wiederum deswegen ein, weil ihre Französischkenntnisse für ein ausreichendes Textverständnis nicht genügten, wie z.B. Omega oder Picard:

dann habe ich auch die viele <<runzelt die stirn und starrt auf den bildschirm> sprachen sprechen> oder irgendetwas habe ich von englisch glaube ich- da "beaucoup d'anglais" (--) ((zeigt mit dem finger auf den bildschirm)) habe ich gesehen und darum habe ich das genommen` (Omega_Off, Pos. 910-913)

MB: warst du dir eher SICHer bei deinen antworten`
S: nEin ((S schüttelt den kopf)) [...] ich habe auch nicht viEl verstanden- dann habe ich halt das zusammengetan was ich em gedacht habe was es sein kÖNnte- u:nd so irgendwie (-) eine antwort zu krIegen-
MB: mh_mh weisst du noch wo du das gefunden hast` bei der (-) cAMille [...]
S: hIEr; (wir) ("tEs(--)ter")
MB: (hier?) ((cursor auf "tester" im text von camille))
S: ja; (Picard_Off, Pos. 804-816)

Strategie „Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern“ (TW03)

Diese Strategie wurde aufgrund der Bemerkungen der Lernenden codiert, wenn 1) der/die Schüler/in eine oder mehrere Antwortoption(en) ausschloss, weil er/sie die entsprechenden Informationen im Inputtext nicht fand oder nicht verstand; 2) beim Itemformat, das das Anklicken von „ja“ oder „nein“ verlangte (Szenario „Szenario Stadtrip“, Task B und Off-Szenario-Aufgabe „Berufswahl“), der/die Schüler/in die Option „nein“ wählte, weil er/sie die passende Information im Text nicht fand oder nicht verstand. Die Strategie wurde u.a. deswegen als konstruktirrelevant eingestuft, weil in allen IFB-Inputtexten neben den Informationen zur richtigen Antwortoption immer auch Elemente aller Distraktoren vorkamen: Wenn die Schüler/innen also behaupteten, zu einer Antwortoption nichts im Text gefunden zu haben, lässt sich daraus schliessen, dass sie den Text nicht richtig zur Kenntnis genommen hatten (oder auch nicht verstanden hatten).

Die Strategie konnte in den Lautdenkprotokollen 17 Mal beobachtet werden. Cersei schloss beispielsweise bei Szenario „Stadtrip“, Task B, Seite 2 die Antwortoptionen „An das gute Essen...“ und „Der starke Regen...“ aus, weil zu diesen Themen ihrer Meinung nach im Inputtext nichts vorkam. Sie entschloss sich dann für die dritte, übrig gebliebene (und falsche) Option „An die schöne Aussicht“.

```
also ich hab gehört dass sie KRANK war? (--) und dass sie (--) den
starken regen (.) das NICHT? (--) von ESSen hat sie glaub ich AUCH
nichts gesagt (.) oder hab ich nicht gehört, (--) da DENke ich dass es
die schöne aussicht ((wählt B2c)) (---) ist. (Cersei_Szenario,
Pos. 106-110)
```

Beim Lösen der Testaufgabe „Berufswahl“, Seite 1 gab die Schülerin Picard an, dass sich die Antwort auf die Frage „Muss man im Beruf „représentant commercial“ selbst neue Produkte testen?“ im Text von Camille befindet. Sie klickte dann auf „nein“, weil sie im Inputtext nichts Passendes dazu fand:

```
((schaut auf die aufgabenstellung und dann kurz auf den text)) m:
"sElber produkte testen"- (---) da habe ich nichts im text gefunden-
((zieht 1 in C1_1rev)) (--) also em (--) (klicke) ich nEin an; (--)
((S klickt C1_1b an)) (Picard_Off, Pos. 82-84)
```

In den Interviews berichteten die Schüler/innen insgesamt 27 Mal über diese Strategie (davon 2 Mal itemübergreifend). In Szenario „Stadtrip“, Task A, Seite 2 ging z.B. Cersei davon aus, dass der Stadtrundgang nicht empfohlen würde, weil sie dazu im Text keine passende Stelle gefunden hatten:

MB: [...] und das mit der STADTrundgang (-) [...]
S: ähm ich habe es eben (-) NICHT gefunden und da dachte ich (-) dass sie es auch nicht emPFEHlen; (Cersei_Szenario, Pos. 553-556)

Auch Picard erinnerte sich im Interview an den Einsatz dieser Strategie beim Task „Berufswahl“, Seite 1:

dann habe ich halt (1.5) mir gedacht sie (-) testet kEIne produkte`
weil darüber irgendwie nichts gestanden ist; ja- (Picard_Off, Pos. 694-696)

Elena berichtete darüber, dass sie im Szenario „Stadttrip“, Task C immer dann „nein“ gewählt hatte, wenn sie im Text keine zur Frage passende Stelle identifizieren konnte:

S: *stadtzentrum (-) steht ja gar nicht deswegen habe ich einfach nein,*
MB: ach so wenn du (-) NICHTS gefunden hast (-) dann hast du NEIN angekreuzt;
S: ja;
MB: okay (-) hast du es IMMer so gemacht?
S: ja; (Elena_Szenario, Pos. 1254-1259)

Strategie „Auswahl einer Antwortoption mit bekannten Wörtern“ (TW04)

Die Strategie „Auswahl einer Antwortoption mit bekannten Wörtern“ wurde codiert, wenn die Lernenden eine Antwortoption gerade deswegen auswählten, weil sie Wörter enthielt, die sie im Inputtext verstehen konnten (also genau das umgekehrte Vorgehen wie bei der Strategie „Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern“). Diese Strategie wurde nur in den *Stimulated-Recall*-Interviews codiert, in denen vier Lernende über ihren Einsatz berichteten. So wählte z.B. Steve bei Szenario „Stadttrip“, Task B, Seite 1 die (falsche) Antwortoption „Ihre neue französische SIM-Karte war noch nicht aktiviert“, weil er verstanden hatte, dass im Inputtext von der SIM-Karte die Rede war. Er gab auch auf die Frage der Forscherin an, dass er zusätzlich auch die zwei anderen Antwortoptionen ausgeschlossen hatte, weil er dazu nichts verstanden hatte (deswegen wurde die Stelle auch als Strategie „Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern“ codiert):

MB: [...] warst du dir SICHer dass das auch die richtige antwort ist`
S: RECHt- ja;
MB: warum` weil du etwas gehört hast oder-
S: weil ich etwas von einer SIMkarte gehört habe;
MB: mh_mh (---) und zu den ANderen zwei nicht-
S: nein nichts;
MB: da hast du die sOfort ausgeschlossen;
S: ja; (Steve_Off, Pos. 538-546)

Strategie „Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen“ (TW05)

Mit dieser Strategie wurden Äusserungen der Lernenden codiert, aus denen hervorging, dass sie *nach* der Rezeption des Inputtextes eine Antwortoption oder mehrere Antwortoptionen ausschlossen oder eine Antwortoption auswählten, weil sie diese aufgrund von eigenem Wissen, eigenen Erfahrungen oder eigenen Vorlieben (und nicht aufgrund des Inputtextes) für falsch oder für richtig hielten.

Diese Strategie wurde in den Lautdenkprotokollen und in den *Stimulated-Recall*-Interviews je 6 Mal codiert, davon je 1 Mal itemübergreifend. Die Analyse der codierten Segmente zeigte, dass die Lernenden diese Strategie in zwei Fällen einsetzten: 1) wenn ihr Textverständnis nicht ausreichend war, und 2) beim Lösen von Szenario „Stadttrip“, Task B, Seite 2; und zwar aus Gründen, die im Folgenden näher beleuchtet werden.

Aus dem Lautdenkprotokoll von Elena ging hervor, dass sie von den französischen Inputtexten allgemein nicht viel verstand (und auch, dass sie nicht besonders motiviert war, sich mit den Testaufgaben auseinanderzusetzen). Beim Lösen von Szenario „Stadttrip“, Task A, wo sie aufgrund des Inputtextes die empfohlenen und nicht empfohlenen Aktivitäten sortieren musste, sagte sie gleich zweimal, dass sie diejenigen Aktivitäten als empfohlen einstufte, die sie selber jemandem empfehlen würde:

```
*also ich denke mir (-) wenn ich SELber tipps geben würde was ich für  
tipps geben würde (--) oder was ich (-) empfehlen würde. ((zieht 2 nach  
A1y und 3 nach A1x)) so* (Elena_Szenario, Pos. 21-23)
```

```
((zieht 2 nach A2y, zieht 4 nach A1z, klickt "weiter")) (4.0) *es ist  
immer soviel TEXT dann muss ich schauen was ich (-) machen möchte.*  
(Elena_Szenario, Pos. 30-31)
```

Im *Stimulated-Recall*-Interview wurde von der Schülerin Picard erwähnt, dass sie sich beim Lösen der Testaufgaben (ähnlich wie Elena) auf ihr Hintergrundwissen gestützt hatte, und zwar dann, wenn sie gemerkt hatte, dass sie die zutreffende Antwortmöglichkeit aufgrund des Inputtextes nicht wählen konnte:

```
wenn ich finde dass em (-) dass ich etwas nicht wEIss` dann habe ich  
auch verschIEDene möglichkeiten und dann (-) schAue ich welche am  
meisten sinn gibt- oder welche am: (2.0) also am (-) idealsten für die  
situation ist; (Picard_Off, Pos. 227-232)
```

Im zweiten Fall wurde der Einsatz dieser Strategie, wie erwähnt, im Zusammenhang mit Task B, Seite 2 im Szenario „Stadttrip“ codiert, und zwar in 4 Lautdenkprotokollen und in 5 *Stimulated-Recall*-Interviews. Es zeigte sich, dass der Einsatz dieser Strategie durch eine ungünstige Wortwahl in der Aufgabenstellung hervorgerufen wurde. Die Frage zu dieser

Hörverstehensaufgabe lautete „Was hat Lou in Avignon besonders beeindruckt?“ und den Lernenden standen drei Antwortmöglichkeiten zur Verfügung: a) Das gute Essen am zweiten Tag; b) Der starke Regen in der Stadt (korrekte Antwort) und c) Die schöne Aussicht vom Schloss. Es stellte sich heraus, dass für viele Schüler/innen das Wort „beeindrucken“ eindeutig positiv konnotiert war, während sie mit dem Regen (vor allem im Urlaub, in dem sich Bloggerin Lou gerade befand) eher etwas Negatives verbanden. Dies führte dazu, dass sie sich schlicht nicht vorstellen konnten, dass Lou vom Regen beeindruckt war, und sie daher diese (korrekte) Antwortoption ausschlossen. Das machte u.a. Dustin, die im Lautdenken-Teil dazu sagte:

sie hat sicher irgend etwas von (--) REgen oder so irgendetwas erzählt (-) [...] starker regen (-) ob das so etwas beEindruckendes ist (-) ((verzieht das gesicht)) glaub ich NICHT und (.) in dem fall nehme ich jetzt einfach das ESSen; (Dustin_Off, Pos. 61-65)

Auch Aria meinte im *Stimulated-Recall*-Interview, dass der Regen sie persönlich nicht beeindrucken würde: Sie würde es eher als nervig empfinden:

der stArke regen in der stAdt; also ich konnte mir einfach nicht vorstellen [...] dass (sie) das speziEll beeindruckt der regen- weil das ist ja irgendwie nICHT so gUt und für mich klingt beeindrucken eher pOsitiv- also [...] dass es sEhr stark rEgnet (-) beeindruckt einem nicht/ [...] also MICH würde es jetzt eher NERven; (Aria_Szenario, Pos. 952-959)

Im *Stimulated-Recall*-Interview mit Nancy zeigte sich, dass auch sie die Antwortoption mit dem Regen ausgeschlossen hatte, obwohl sie recht viel vom Inputtext verstanden hatte:

S: *und dann hat sie gesagt dass es halt am ZWEIten tag geregnet hat (--) und [...] dann habe ich so gedacht JA: aber das kann sie ja nicht beeindrucken*
 MB: warum MEINST dass äh (-) der starke regen sie nicht beeindrucken konnte (--) hast du das gehört oder hast du dir eher (-) LOgisch überlegt?
 S: *also ich habe so (-) sie hat ja gesagt dass es einfach RECHT stark geregnet hat und als sie dann am zweiten (-) am DRITTen tag wieder RAUSgehen konnte [...] und dann habe ich so gedacht ja dass (-) dann hat sie das wohl NICHT so beeindruckt weil (-) [...] ja (-) regen ist ja eh nicht so cool?* (Nancy_Off, Pos. 623-639)

Diese Evidenz aus den Verbalprotokollen spricht dafür, dass dieses Item überarbeitet werden müsste, weil es nicht die Absicht der Testautor/innen war, das Item durch die gewählte Formulierung absichtlich zu erschweren. Zusätzliche damit verbundene Schwierigkeiten lagen wohl auch darin, dass es sich dabei um ein Item handelte, das 1) das Hörverstehen testete, was von den Lernenden generell als schwieriger als das Leseverstehen empfunden wurde

(u.a. deswegen, weil man den Inputtext nur einmal hören konnte); 2) Inferenzen verlangte, d.h. die Antwort stand nicht wortwörtlich im Text, sondern musste erschlossen werden⁹⁹.

Strategie „Auswahl einer anders aussehenden Antwortoption“ (TW06)

Die *Stimulated-Recall*-Interviews mit den Off-Szenario-Schüler/innen Omega und Picard zeigten, dass sie sich beim Bearbeiten der Aufgabe „Walkies“, Seite 1 für eine (in diesem Fall richtige) Antwortoption entschieden, weil sie grafisch anders als die restlichen Antwortoptionen aussah. In dieser Aufgabe mussten die Lernenden mit Bezug auf eine Gruppenarbeit in der Schule die Frage „Was ist deine Aufgabe?“ beantworten und zusätzlich angeben, in welcher Nachricht sie die Antwort gefunden hatten. Zur Wahl standen ihnen drei Antwortmöglichkeiten: a) Eine Tabelle mit Vor- und Nachteilen zusammenstellen; b) Einen Slogan für den Prospekt erfinden; und c) Einen Text über die Erfinderin der „Walkies“ schreiben (die richtige Antwort). Weil die Frage „Was ist **deine** Aufgabe?“ lautete und im Inputtext die Antwortoption mit dem Wort „Walkies“ in der anders aussehenden „ich“-Sprechblase präsentiert wurde, gingen Omega und Picard davon aus, dass sich darin die richtige Antwort befinden müsse (vgl. Abbildung 19).

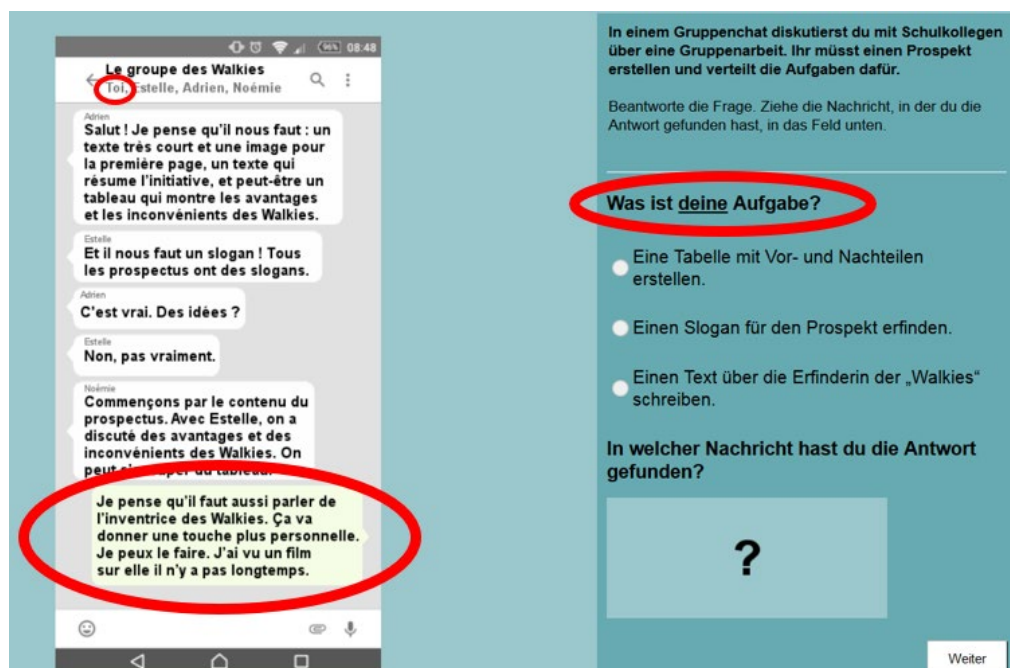


Abbildung 19: Das Design der Aufgabe „Walkies“, Seite 1.

⁹⁹ Die entsprechende Textpassage lautet: „Pendant les deux premiers jours, il a plu du matin au soir, sans arrêt. Je ne voyais même pas les bâtiments de l'autre côté de la rue. Je pense que je n'ai jamais vu une pluie pareille!“ („Während der ersten beiden Tage hat es von morgens bis abends ununterbrochen geregnet. Ich konnte nicht einmal die Gebäude auf der anderen Strassenseite sehen. Ich glaube, so einen Regen habe ich noch nie gesehen!“). Aus dem Text ging ausserdem hervor, dass die zwei anderen Antwortoptionen falsch waren.

Die beiden sagten dazu im Interview:

ausserdem ist das (---) ähm (2) wie [...] meine sprechblase da ist et/
die meinung da steht deine aufgabe und so dann- (Omega_Off,
Pos. 417-419)

da stehen ja die nAmen- von den leuten die geschrieben haben- und dann
(-) ICH bin ja (-) das grüne` u:nd em (-) da steht ja "was ist DEIne
aufgabe" also mEIne- und ICH bin grün und dann habe ich hier gelEsen
(-) wAs: ich mache (-) oder was ich machen soll- (Picard_Off,
Pos. 252-257)

Diese Aufgabe müsste folglich überarbeitet werden, denn die richtige Antwort soll hier nicht aufgrund der grafischen Darstellung eines Textteils, sondern wegen seines Inhalts gewählt werden.

Strategie „Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats“ (TW07)

Aus Verbalisierungen einiger Lernender ging hervor, dass sie eine Antwortoption auswählten, weil sie die Aufgabeninstruktionen nicht aufmerksam genug lasen oder auch bestimmte Vorstellungen davon hatten, wie (Sprach-)Tests normalerweise aufgebaut seien. Die Strategie „Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats“ wurde codiert, wenn die Verbalisierungen der Schüler/innen darauf hindeuteten, dass sie 1) bei Szenario „Stadttrip“, Task C und Off-Szenario „Berufswahl“ die Aufgabenstellung nicht richtig lasen und dachten, dass alle drei Rezensionen einer der drei Fragen zugeordnet werden müssen; oder 2) davon ausgingen, dass eine Antwort negativ oder positiv sein müsse, wenn die früheren Antworten negativ oder positiv waren.

In den Lautdenkprotokollen wurde die Strategie 7 Mal codiert, darunter bei Böbu beim Lösen von Szenario „Stadttrip“, Task C Seite 2. In dieser Aufgabe mussten aufgrund von Inputtexten von drei Personen drei Fragen mit „ja“ oder „nein“ beantwortet werden. Es musste auch angegeben werden, in welchem Text die Antwort gefunden wurde, wobei in der Instruktion darauf hingewiesen wurde, dass sich im gleichen Text auch Antworten auf mehrere Fragen befinden können. Diese Bemerkung in der Instruktion schien u.a. Böbu nicht wahrgenommen oder auch vergessen zu haben. Er ordnete den Text von Lyle der Frage „Gibt es im Hotel Zimmer unter 35 € pro Nacht?“ zu, obwohl er in diesem Text nichts über den Preis fand. Das verunsicherte ihn auch etwas:

*und dann muss DAS noch (2) °h "avis de lyle" ((zieht 1 nach C2_3rev))
(-) muss dann noch (--) (das_sein_und)(---) DAS (hier)- äh jetzt geht
es aber nicht auf;* (2.0) ah (2) aber ich habe nur NEIN* (3.0) ((wählt
C2_3b an)) (---) *ja einfach nur ich habe keinen PREI:S DA; °h hm;*
((klickt auf "weiter")) (Boebu_Szenario, Pos. 269-275)

Er klickte dann „nein“ an, weil er in diesem Text keine Informationen zum Preis fand, was auch als Strategie „Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern“ codiert wurde.

Auch bei Omega wurde diese Strategie codiert. Beim Beantworten der Frage „Muss man im Beruf „représentant commercial“ selbst neue Produkte testen?“ in der Aufgabe „Berufswahl“, Seite 2 sagte er nämlich:

```
*keine ahnung; (---) (unv.) ((schaut auf die antwortmöglichkeiten))  
ich sage in dem fall ne:in- (2.0) dann haben wir nicht das gleiche wie  
vorhin- (--) vorher ((meint die frühere bildschirmseite)) hatte ich  
auch nein-((wählt C2_1b an)) (---) ((schaut auf den text und die  
aufgabenstellung hin und her)) kann ja nicht zweimal das gleiche sein;*(  
Omega_Off, Pos. 147-152)
```

Der Schüler verstand nicht viel vom Inputtext und wählte seine Antwort aufgrund dessen, was er auf der früheren Bildschirmseite angeklickt hatte. Er vermutete, dass es eher unplausibel sei, dass zweimal nacheinander zur gleichen Frage etwas Negatives komme und klickte deswegen bei der zweiten Bildschirmseite „ja“ an.

In den *Stimulated-Recall*-Interviews wurde diese Strategie 11 Mal codiert. Dustin gab z.B. im Gespräch an, dass sie bei der Aufgabe „Berufswahl“, Seite 2 eine der Fragen u.a. deswegen mit „nein“ beantwortet hatte, weil ihre Antwort auf die zwei anderen Fragen bereits „ja“ gelautet hatte:

```
*das mit viel reisen (-) da war ich nicht sicher (--) [...] eigentlich  
dort habe ich (-) NICHT wirklich gecheckt gehabt (-) und danach habe  
ich NEIN angekreuzt (-) weil ich schon zweimal ja hatte und nein noch  
nie* ((lacht)) (Dustin_Off, Pos. 722-726)
```

Auch Mogli setzte diese Strategie bei der Aufgabe „Berufswahl“, Seite 1 ein: Er hatte den letzten der drei Inputtexte einer Frage zugeordnet, weil er (wie Böbu) davon ausgegangen war, dass jeder der drei Texte einer Frage zugeordnet werden müsse:

```
und der LETze (-) habe ich dann (.) AUch noch schnell halt/ also DAS  
musste ich NICHT mehr überlegen- weil ich ja (-) zWEI hatte und dann  
war es das drITte- (Mogli_Off, Pos. 1035-1037)
```

Die Erkenntnisse aus den Verbalprotokollen von u.a. Böbu und Mogli sprechen dafür, dass in der Instruktion der Satz „Manchmal findest du im gleichen Text mehrere Antworten“ stärker hervorgehoben werden sollte.

Resümee

In der vorliegenden Studie konnten insgesamt 7 *test-wiseness*-Strategien beobachtet und codiert werden. Sie entsprechen etwa 8% der in den Lautdenkprotokollen und etwa 16% der in den *Stimulated-Recall*-Interviews codierten Strategien. Generell wurden diese Strategien dann codiert, wenn die Schüler/innen Probleme mit dem Verständnis der Inputtexte hatten oder wenn sie die Inputtexte eher oberflächlich zur Kenntnis nahmen, um die Aufgabe möglichst schnell zu lösen. Dank der Identifizierung der Stellen, an denen *test-wiseness*-Strategien zum Einsatz kamen, konnten Hinweise zu den Teilen der Items und Inputtexte gefunden werden, die in den IFB-Aufgaben überarbeitet werden sollten, weil sie für den Gebrauch dieser Strategien anfällig waren.

In Kap. 4.2.1 und 4.2.2 wurde über die Codierung der einzelnen Lerner- und Teststrategien über alle 20 Schüler/innen hinweg berichtet. Nicht alle Schüler/innen griffen aber bei der Bearbeitung der Testaufgaben auf die gleichen Strategien und im gleichen Ausmass zurück. Im nächsten Kapitel wird daher auf die individuellen Unterschiede beim Strategieneinsatz zwischen den einzelnen Schüler/innen eingegangen.

4.2.3 INDIVIDUELLE UNTERSCHIEDE BEIM STRATEGIENEINSATZ

Um zu illustrieren, wie sich die individuellen Unterschiede zwischen den 20 Testteilnehmenden manifestieren, wird in Abbildung 20 dargestellt, wie oft die Schüler/innen die sieben Strategien¹⁰⁰ einsetzten, die in den Lautdenkprotokollen (LD) und *Stimulated-Recall*-Interviews (SR) am häufigsten codiert wurden.¹⁰¹ Zu diesen Strategien gehören:

- Testmanagementstrategien
 - TM01: „Instruktionen und/oder Antwortoptionen (erneut) lesen“,
 - TM08: „Einsatz von Verstandem bei Auswahl/Ausschluss der Antwortoption(en)“,
 - TM06: „Fokus auf relevante Textteile“,
- *test-wiseness*-Strategie TW02: „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“.

¹⁰⁰ Es wurde entschieden, in Abbildung 20 nur die sieben häufigsten Strategien zu zeigen, weil die Grafik bei der Darstellung einer grösseren Anzahl der Strategien unlesbar wäre.

¹⁰¹ Es wird dabei vom Gesamtpool der Strategien ausgegangen.

- Lernerstrategien
 - L08: „Intensive Auseinandersetzung mit Teilen des Inputtextes“,
 - L03: „Einen Teil des Inputtextes (erneut bzw. genauer) lesen“, sowie
 - L01: „Sorgfältiges Lesen“.

Die in Abbildung 20 präsentierten Zahlen wurden aufgrund einer Kreuztabelle ermittelt, die in MAXQDA (VERBI Software, 2019) mithilfe des Tools „Code-Relations-Browser“ (Art der Analyse: Überschneidung von Codes am Segment) erstellt wurde. Aus technischen Gründen beziehen sich diese Zahlen auf die totale Anzahl der codierten Strategien, d.h. die Codierungen mit den speziellen Gewichtungen „2“ und „25“ wurden hier nicht abgezogen¹⁰² (vgl. Anmerkungen in Kap. 4.2). Deswegen wurde auch in der Abbildung 20 auf die Darstellung der Strategie L10 „Unbekannte Wörter, Satz- bzw. Textteile identifizieren“ verzichtet: Sie wurde zwar im *Stimulated-Recall*-Teil insgesamt 213 Mal, also sehr oft codiert, in 180 Fällen handelte sich aber um Codierungen mit der Gewichtung „25“ (d.h. um Codierungen, die bei der mündlichen Zusammenfassung der Inputtexte vorgenommen wurden).

¹⁰² Für den Abzug der Codierungen mit den speziellen Gewichtungen wäre es nötig gewesen, jedes einzelne codierte Element manuell zu öffnen und zu überprüfen, was mit einem erheblichen Aufwand verbunden gewesen wäre. Weil die Zahlen zu den codierten Elementen, über die in dieser qualitativen Arbeit berichtet wird, lediglich der besseren Transparenz und der Illustration der Trends in den Daten dienen, wurde entschieden, hier auf diese aufwändige Arbeit zu verzichten.

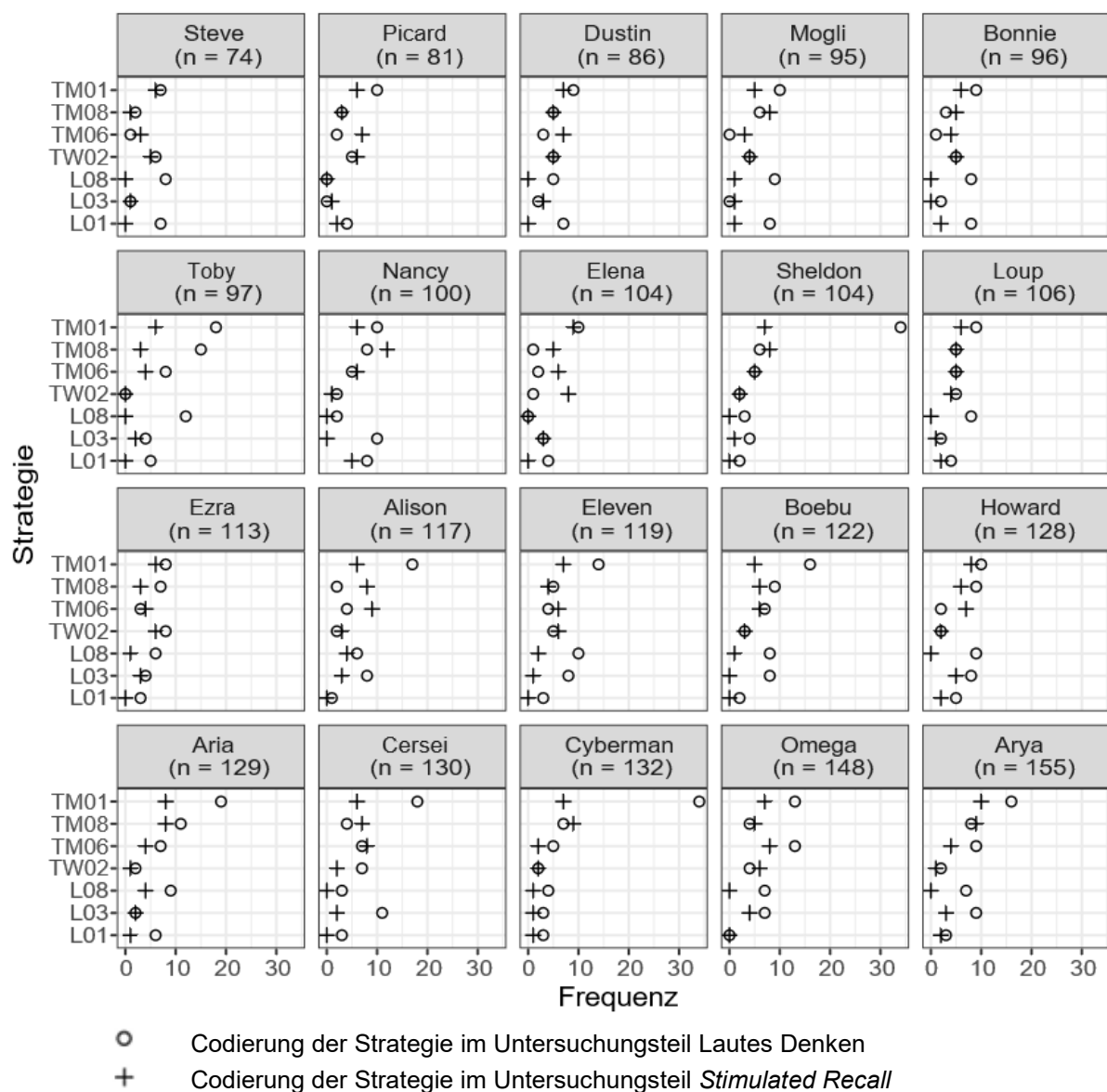


Abbildung 20: Gebrauch der sieben am häufigsten codierten Strategien pro Schüler/in.

Der obigen Abbildung ist zu entnehmen, dass die Lernenden auf die Strategien allgemein unterschiedlich oft zurückgriffen: Bei Arya wurden sie z.B. mit 155 Codierungen mehr als zweimal so häufig beobachtet wie bei Steve (n=74). Allen Schüler/innen war gemeinsam, dass sie sich mit den Instruktionen und/oder den Antwortoptionen z.T. auch wiederholt auseinandersetzten (TM01) und dass sie bei der Textrezeption ihren Fokus auf relevante Teile der Inputtexte lenkten (TM06). Wenn man den Einsatz der Strategie TM08 („Einsatz von Verstandem bei Auswahl/Ausschluss der Antwortoption(en)“) mit den Testergebnissen der Schüler/innen zu den analysierten Testaufgaben vergleicht, liegt die Vermutung nahe, dass diese Strategie von denjenigen Schüler/innen häufiger angewandt wurde, die im Test bessere Resultate erzielten. Sie wurde z.B. häufiger bei Toby (12/12 Punkten im Test), Aria (8/12 Punkten) oder Arya (8/12 Punkten) als bei Steve (1/10 Punkten), Picard (1/10 Punkten) oder Dustin (5/10 Punkten) codiert.

Auch in Bezug auf den Gebrauch der Lernerstrategien unterschieden sich die Lernenden teilweise erheblich voneinander: Während z.B. Omega die Inputtexte nie sorgfältig las (Strategie L01), nahmen sich z.B. Nancy, Bonnie, Mogli oder Dustin häufiger dafür Zeit. Der Beitrag dieser Strategie zum Testserfolg scheint hier aber vor allem davon abhängig zu sein, wie viel die Lernenden vom Gelesenen tatsächlich verstanden. So konnte z.B. neben Omega auch bei Alison kaum eine sorgfältige Lektüre beobachtet werden, der Punktezahl (8/12) und den Verbalprotokollen von Alison ist jedoch zu entnehmen, dass sie die Inhalte der Inputtexte alles in allem trotzdem recht gut verstand. Anders bei Steve, der die Inputtexte zwar sorgfältig rezipierte, aber im Test nur 1 von 10 Punkten holen konnte, dies auch nur unter Einsatz der *test-wiseness*-Strategie „Raten“ (TW01)¹⁰³: Seinen Verbalisierungen und mündlichen Zusammenfassungen der Inputtexte war zu entnehmen, dass seine Französischkenntnisse für die erfolgreiche Lösung der Testaufgaben nicht ausreichten. Gewisse Lernende wie Cersei (5/12 Punkten), Arya (8/12 Punkten) oder Loup (3/10 Punkten) lasen auch Teile des Inputtextes erneut bzw. genauer durch (Strategie L03), während andere wie Steve (1/10 Punkten), Picard (1/10 Punkten) oder Mogli (3/10 Punkten) entweder nie oder selten Gebrauch von dieser Strategie machten – vielleicht deswegen, weil sie bereits beim ersten Lesen nicht viel von den Inputtexten verstanden. Die intensive Auseinandersetzung mit dem Inputtext (L08) konnte bei Elena und Picard nie, bei Toby, Mogli, Bonnie, Steve, Loup, Eleven oder Howard wiederum recht oft beobachtet werden. Da diese Strategie dann codiert wurde, wenn die Schüler/innen die Teile der Inputtexte wiederholten, paraphrasierten, übersetzten oder zusammenfassten, kann nicht ausgeschlossen werden, dass ihr Gebrauch zumindest teilweise auch mit der (unterschiedlich ausgeprägten) Bereitschaft der Lernenden zum lauten Aussprechen ihrer Gedanken zusammenhing.

Die Codierungen der *test-wiseness*-Strategie „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“ (TW02) kamen oft bei Schüler/innen vor, die im Test eine niedrige Punktezahl erreichten und die auch gemäss den Verbalprotokollen eher wenig von den rezipierten Texten verstanden. So konnte der wiederholte Einsatz dieser Strategie bei Steve (1/10 Punkten), Picard (1/10 Punkten), Eleven (2/10 Punkten), Mogli (3/10 Punkten), Loup (3/10 Punkten), Elena (4/12 Punkten) und Cersei (5/12 Punkten); aber auch bei Ezra (5/10 Punkten), Omega (6/10 Punkten) oder Bonnie (7/10 Punkten) beobachtet werden. Auffällig ist, dass die Schüler/innen mit den besten Resultaten, d.h. Toby (12/12 Punkten), Cybermen (9/12 Punkten), Howard (9/12 Punkten), Aria (8/12 Punkten), Arya (8/12 Punkten) und Sheldon (8/12 Punkten) entweder nie oder selten auf diese Strategie zurückgriffen. Da diese Lernenden tendenziell viel von den

¹⁰³ Im *Stimulated-Recall*-Interview sagte Steve zu dieser Stelle: da war ich mir gAr nicht sicher; habe ich geraten; (Steve_Off, Pos. 631-633)

Inputtexten verstanden, hatten sie es wohl auch nicht nötig. Interessanterweise wurde diese Strategie ebenfalls beim Schüler Böbu nicht oft beobachtet, obwohl sein Testresultat (3/12 Punkten) und seine Verbalisierungen darauf hindeuten, dass er mit dem Verständnis der Inputtexte eher Mühe hatte. Dafür setzte aber Böbu andere *test-wiseness*-Strategien ein (TW03 und TW07), die andere Lernenden wiederum nicht häufig brauchten.

Die Analyse der Lautdenkprotokolle und der *Stimulated-Recall*-Interviews bestätigte darüber hinaus die aus der Strategieforschung bekannte und in Kap. 2.2.2 erwähnte Tatsache, dass Strategien selten separat, sondern oft in Clustern oder in Ketten auftreten (Oxford, 2017, S. 41–42). Auf die von den 20 Schüler/innen eingesetzten Strategiecluster und -ketten soll hier nicht im Detail eingegangen werden. Wie unterschiedlich die in den Daten anzutreffenden Strategienkombinationen jedoch sein können, wird hier stellvertretend an zwei Beispielen illustriert (Abbildung 21 und Abbildung 22). In beiden Beispielen wird von den Lernenden die Hörverstehensaufgabe aus dem Szenario „Stadttrip“, Task B, Seite 2 bearbeitet.

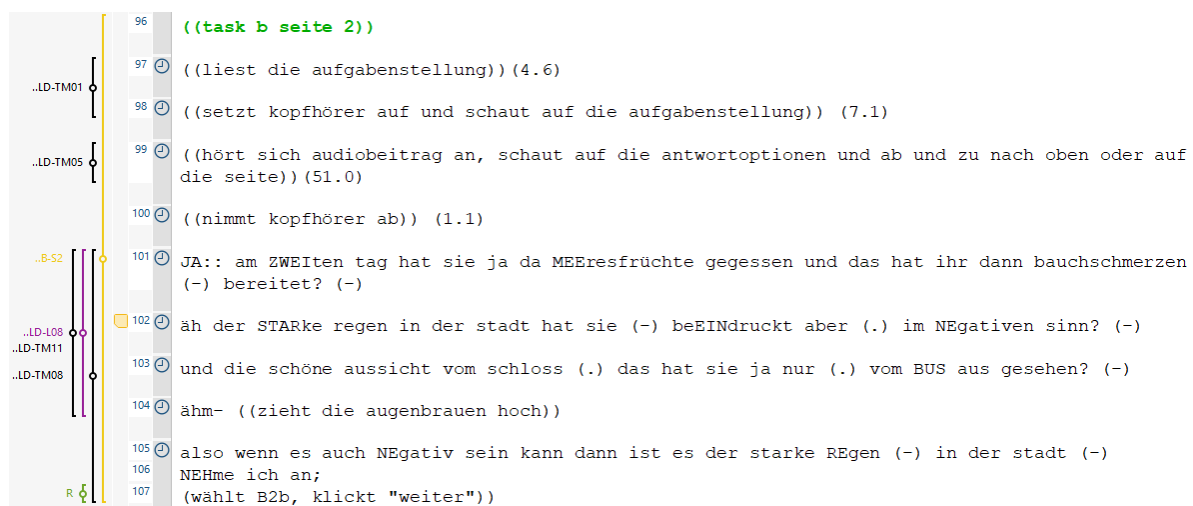


Abbildung 21: Ausschnitt aus dem codierten Lautdenkprotokoll des Szenario-Schülers Toby.

Dem Transkriptausschnitt von Toby ist zu entnehmen, dass er sich zuerst mit der Aufgabenstellung auseinandersetzte (TM01). Danach hörte er den Audiobeitrag an und schaute dabei auf die Aufgabenstellung (TM05). Nach dem Hören paraphrasierte er den Inputtext (L08) und setzte das Verstandene ein (TM08), um sich für eine der drei Antwortoptionen zu entscheiden, die er miteinander verglich (TM11). Toby verstand das Wesentliche vom Inputtext und zog darauf basierend auch die richtigen Schlüsse. Die von ihm gewählte Antwort war daher auch korrekt.

	64	((szenario stadtrip task b seite 2))
	65	((setzt kopfhörer auf und blickt dabei sehr kurz auf die aufgabenstellung, liest diese aber nicht wirklich)) (8.5)
..LD-TM05	66	((hört sich audioblog an und schaut auf die aufgabenstellung, ab und zu auf die seite, nach oben oder nach unten)) (49.1)
..Stadtrip-2	67	((nimmt kopfhörer ab)) (2.7)
	68	ALso sie war KRANK (-) malade (-)
..LD-L08	69	das haben wir gerade in der SCHUle gelernt? (--)
..LD-TM08	70	malade also (-) ja (-) und (-) ähm sie hat (-) ähm
	71	das ESSen (-) hat sie NICHT so (-) LECKER gefunden? (--)
..LD-TM11	72	und (-) das mit dem (.) REgen (-) pleut (-) habe ich gar nicht gehört? (--)
..LD-TV03	73	hm ich weiss jetzt NICHT ähm (-)
..LD-TV01	74	die schöne AUSSicht vielleicht (-) schloss? ((wählt B2c und klickt "weiter" 2x)) (6.0)

Abbildung 22: Ausschnitt aus dem codierten Lautdenkprotokoll der Off-Szenario-Schülerin Loup.

Anders als Toby nahm sich Loup beim Bearbeiten dieser Aufgabe kaum Zeit für die Kenntnisnahme der Aufgabenstellung und der Antwortmöglichkeiten, bevor sie sich den Text anhörte. Dafür schaute sie sich diese beim Hören des Inputtextes an (TM05). Nach dem Hören fasste sie ähnlich wie Toby den Inputtext zusammen (L08) und versuchte das Verstandene einzusetzen (TM08, hier allerdings mit Gewichtung „15“, weil Loup die Textpassage falsch verstand). Die Schülerin verglich dabei auch die Antwortoptionen miteinander (TM11). Sie schloss dann die Antwortoption mit dem Regen aus, weil sie dazu im Inputtext nichts verstand (TV03). Schliesslich schien sie bei der Auswahl der (falschen) Antwortmöglichkeit „An die schöne Aussicht vom Schloss“ geraten zu haben: „ich weiss jetzt NICHT ähm (-) die schöne AUSSicht vielleicht (-) schloss?“, Pos. 73-74, was als TW01 codiert wurde.

Resümee

Den Resultaten zur ersten Forschungsfrage ist zu entnehmen, dass die Schüler/innen viele verschiedene Strategien einsetzten, um die IFB-Testaufgaben zu lösen. Es konnten aber auch teilweise grosse individuelle Unterschiede beim Einsatz der Strategien durch einzelne Schüler/innen festgestellt werden.

Da die Mehrheit der codierten Strategien konstruktrelevant war, ist davon auszugehen, dass die Lernenden in erster Linie die Inputtexte verstehen und sinnvolle Antworten auf die Testfragen geben wollten. Die konstruktirrelevanten *test-wiseness*-Strategien kamen hingegen dann zum Einsatz, wenn das Textverständnis der Lernenden für die sinnvolle Beantwortung der Testfragen nicht ausreichte oder wenn sie die Testaufgaben schnell lösen wollten und die Inputtexte nicht ausreichend gründlich zur Kenntnis nahmen. Durch die Codierungen der *test-wiseness*-Strategien konnten darüber hinaus einige Stellen in den Aufgaben entdeckt werden, die vor einem eventuellen Einsatz der IFB-Aufgaben in Leistungsmessungen überarbeitet und nochmals erprobt werden sollten.

Im nächsten Kapitel werden die Resultate zur zweiten Forschungsfrage dargestellt. Im Zentrum steht dabei die Untersuchung der Zusammenhänge zwischen dem Einsatz der konstruktirrelevanten *test-wiseness*-Strategien und dem erfolgreichen bzw. nicht erfolgreichen Lösen der Testaufgaben.

4.3 ZU FORSCHUNGSFRAGE 2: ZUSAMMENHÄNGE ZWISCHEN KONSTRUKTIRRELEVANTEN STRATEGIEN UND ERFOLGREICHEN BZW. NICHT ERFOLGREICHEN AUFGABENLÖSUNGEN

Beim Erproben von neu erstellten Testaufgaben wie den in der vorliegenden Studie untersuchten ist es wichtig zu ermitteln, ob sie für den Gebrauch der konstruktirrelevanten *test-wiseness*-Strategien anfällig sind, und wenn ja, aus welchem Grund: Sollte sich zeigen, dass bestimmte Eigenschaften der Testaufgaben zum Einsatz dieser Strategien beitragen, müssten die Aufgaben überarbeitet werden, insbesondere dann, wenn der Einsatz von *test-wiseness*-Strategien zum erfolgreichen Lösen der Testaufgaben führte. Die zweite Forschungsfrage lautete daher: „Lassen sich Zusammenhänge zwischen konstruktirrelevanten Strategien und (nicht) erfolgreichen Aufgabenlösungen erkennen?“

Um diese Frage beantworten zu können, wurden in den Lautdenkprotokollen und *Stimulated Recall*-Interviews alle Stellen sorgfältig untersucht, in denen der Einsatz einer *test-wiseness*-Strategie codiert wurde. Pro Item, das unter Einbezug der codierten *test-wiseness*-Strategie(n) bearbeitet wurde, wurde dabei im Sinne der evaluativen qualitativen Inhaltsanalyse eingeschätzt, ob der/die Schüler/in 1) durch diese Strategie(n) eine falsche Antwort wählte, 2) durch diese Strategie(n) die Frage zum Text richtig beantwortete, oder ob er oder sie 3) beim Bearbeiten dieses Items zwar eine oder mehrere *test-wiseness*-Strategie(n) einsetzte, im Endeffekt jedoch das Item aus einem anderen, konstruktrelevanten Grund richtig lösen konnte (z.B. weil er/sie den Inputtext nochmals rezipierte und die ursprüngliche Antwort daraufhin korrigierte).

In Kap. 4.3.1 werden die ermittelten Zusammenhänge zwischen dem Einsatz der *test-wiseness*-Strategien und den erfolgreichen bzw. erfolglosen Lösungen der Szenario-Aufgaben dargestellt. Den erfolgreichen und nicht erfolgreichen Lösungen der Off-Szenario-Aufgaben, bei denen die *test-wiseness*-Strategien eine Rolle spielten, ist Kap. 4.3.2 gewidmet.

4.3.1 TEST-WISENESS-STRATEGIEN UND (NICHT) ERFOLGREICHE LÖSUNGEN DER SZENARIO-AUFGABEN

Tabelle 17 zeigt auf, welche Schüler/innen welche *test-wiseness*-Strategie bei welchen konkreten Items und mit welchem Ergebnis anwandten. Die Codierungen im Lautdenken-Teil und im *Stimulated-Recall*-Interview wurden aus Übersichtsgründen zusammen dargestellt, und zwar pro Strategie und Item. Die Items, bei denen die Schüler/innen aufgrund einer oder mehrerer *test-wiseness*-Strategien eine Antwort wählten und diese Antwort auch richtig war, wurden in der Tabelle rot hinterlegt (rot wurde gewählt, um den Handlungsbedarf für die Testentwicklung zu signalisieren, weil in diesen Fällen die richtigen Lösungen aus den falschen Gründen gewählt wurden). Die Items, bei denen der Einsatz von *test-wiseness*-Strategien zu falschen Resultaten führte, wurden hingegen grün markiert. Die genaue Lektüre der Verbalprotokolle zeigte, dass es auch einige Fälle gab, in denen die Schüler/innen zwar eine *test-wiseness*-Strategie einsetzten, ihre Lösung aber schliesslich doch aus einem konstruktrelevanten Grund wählten, beispielsweise dann, wenn sie den Inputtext nochmals genauer betrachteten und ihre Antwort daraufhin korrigierten. Diese Fälle wurden in der Tabelle gelb markiert.

Tabelle 17: Items, bei denen Szenario-Schüler/innen *test-wiseness*-Strategien einsetzten (gemäss Codierungen im Lautdenkprotokoll und/oder *Stimulated-Recall*-Interview).

Erklärung der Farbsymbolik:

- Grün: Der/die Schüler/in wählte aufgrund der eingesetzten *test-wiseness*-Strategie(n) eine falsche Antwort.
 Gelb: Der/die Schüler/in wählte letztendlich eine richtige Antwort, aber ohne Mitwirkung der codierten *test-wiseness*-Strategie(n).
 Rot: Der/die Schüler/in wählte aufgrund der eingesetzten *test-wiseness*-Strategie(n) eine richtige Antwort.

Szenario „Stadttrip“														
Aufgabe	Task A	Task A	Task A	Task A	Task B	Task B	Task C	Task C	Task C	Task C	Task C	Task C	Anzahl korrekter Antworten mit TW-Strategie(n) pro Schüler/in (Max. 12) ¹⁰⁴	Anzahl falscher Antworten mit TW-Strategie(n) pro Schüler/in (Max. 12)
Bildschirmseite	Seite 1	Seite 1	Seite 2	Seite 2	Seite 1	Seite 2	Seite 1	Seite 1	Seite 1	Seite 2	Seite 2	Seite 2		
Item	Rugby-Ausstellung	Coq Rouge	Carte Avignon	Stadtrundgang	Internet	Beeindruckt in Avignon	WLAN	Stadtzentrum	preiswert	WLAN	Stadtzentrum	preiswert	Anzahl korrekter Antworten mit TW-Strategie(n) pro Schüler/in (Max. 12) ¹⁰⁴	Anzahl falscher Antworten mit TW-Strategie(n) pro Schüler/in (Max. 12)
Alison (8/12) ¹⁰⁵	TW01			TW01, TW02		TW03	TW02				TW02			
Aria (8/12)						TW03, TW05					TW02	TW02	0	2
Arya (8/12)					TW02		TW02	TW02					0	2
Cersei (5/12)	TW03		TW02	TW02		TW03, TW05	TW02				TW07	TW02	0	5
Elena (4/12)	TW03	TW02, TW05	TW01, TW02	TW05	TW02		TW02				TW01, TW02	TW02	2	5
Böbu (3/12)	TW01			TW03				TW02	TW02		TW02	TW04, TW07	0	6
Cyberman (9/12)					TW03	TW02, TW03					TW02		0	3
Howard (9/12)			TW01					TW02			TW02		1	2
Sheldon (8/12)						TW05		TW02			TW02		1	2
Toby (12/12)													0	0
Anzahl korrekter Antworten mit TW-Strategie(n) pro Item ¹⁰⁶	0	0	0	1	1	0	0	1	0	0	2	0		
Anzahl falscher Antworten mit TW-Strategie(n) pro Item	2	1	3	2	2	5	4	2	1	0	6	2		

¹⁰⁴ Ohne Items, in denen die richtigen Antworten letztendlich ohne *test-wiseness*-Strategien gewählt wurden (gelb hinterlegt).

¹⁰⁵ Anzahl erreichte Punkte.

¹⁰⁶ Ohne Items, in denen die richtigen Antworten letztendlich ohne *test-wiseness*-Strategien gewählt wurden (gelb hinterlegt).

Der obigen Zusammenstellung kann entnommen werden, dass bis auf den Schüler Toby alle Szenario-Schüler/innen beim Lösen der Testaufgaben mehrmals auf *test-wiseness*-Strategien zurückgriffen. In den meisten Fällen waren sie damit jedoch nicht erfolgreich: Ein Blick auf die Anzahl der Items, die die Schüler/innen mithilfe der *test-wiseness*-Strategien lösten, zeigt, dass die Schüler/innen damit überwiegend falsche Antworten erzielten. Bei Böbu war dies sogar bei 6 von 12 Items der Fall, Cersei und Elena haben durch den Einsatz der *test-wiseness*-Strategien 5 falsche Antworten ausgewählt, Cybermen und Alison lösten deswegen je 3 und Aria, Arya, Howard und Sheldon je 2 Items falsch. Dies ist aus der Sicht der Testentwickler/innen als positiv zu betrachten, denn es bedeutet, dass die Distraktoren in den meisten Fällen ihre Rolle gut erfüllten (z.B. bei Task C, Seite 1, Item „WLAN“).

Auf den ersten Blick scheinen auch bei Task B, Seite 2 die Distraktoren gut zu funktionieren, weil die *test-wiseness*-Strategien bei allen fünf Szenario-Schüler/innen, die sie beim Lösen dieses Items einsetzten, zum Misserfolg führten. Bei der näheren Betrachtung der codierten Stellen zeigte sich jedoch, dass drei Szenario-Schüler/innen die richtige Antwort „der starke Regen in der Stadt“ deswegen ausschlossen (Strategie TW05), weil sie sich nicht vorstellen konnten, dass dies eine korrekte Antwort auf die Frage „Was hat Lou in Avignon besonders beeindruckt?“ wäre. Dies hatte damit zu tun, dass für sie der Regen während des Urlaubs etwas Negatives wäre und sie das Wort „beeindrucken“ eindeutig mit etwas Positivem assoziierten. Offenbar sorgt dieses Item für unerwünschte Verwirrung bei Testteilnehmenden und sollte daher angepasst werden.

Ausserdem setzten Alison, Aria, Arya, Cersei und Elena zwar beim Bearbeiten bestimmter Items eine oder mehrere *test-wiseness*-Strategien ein, im Endeffekt waren aber ihre richtigen Lösungen dieser Items nicht auf diese Strategien zurückzuführen. Dies hat womöglich damit zu tun, dass die Szenario-Schüler/innen die Lernerstrategie L03 „Einen Teil des Inputtextes (erneut bzw. genauer) lesen“ recht häufig einsetzten (vgl. Kap. 4.4.2), was es ihnen ermöglichte, ihre Fehler zu bemerken und zu korrigieren.

Einer genaueren Betrachtung gebühren hingegen die rot markierten Fälle, in denen die richtigen Lösungen aus den falschen Gründen gewählt wurden. Dies gelang Alison, Elena, Howard und Sheldon. Bei Elena schien Zufall im Spiel gewesen zu sein: Sie schien die Inputtexte nicht richtig gelesen bzw. verstanden zu haben und wählte dann eine Antwort, zu der sie im Inputtext ein passendes Wort fand (Strategie TW02) oder die sie aufgrund ihres Hintergrundwissens für richtig hielt (Strategie TW05). Anders sieht es bei drei weiteren Fällen aus, auf die bereits in Kap. 4.2.2.2 beim Besprechen der Strategie TW02 ausführlich eingegangen wurde und bei denen der Erfolg von Alison, Howard und Sheldon auf gewisse Eigenschaften der Inputtexte zurückgeführt werden kann. Es handelt sich dabei um Task C,

Seite 1, Item „Stadtzentrum“ (Sheldon) und um das analoge Item auf der zweiten Bildschirmseite zu dieser Aufgabe (Alison, Howard). Sheldon ging aufgrund der Stelle „En général, l'hôtel n'était pas mal...“ („Im Allgemeinen war das Hotel nicht schlecht...“) im Inputtext von Curt fälschlicherweise davon aus, dass es sich bei diesem Hotel um ein „Generalhotel“ handle und sich dieses folglich im Stadtzentrum befinde. Im Inputtext wurde zwar tatsächlich erwähnt, dass das Hotel im Zentrum war, aber an einer weiteren Stelle (ausserdem bedeutet „en général“ „im Allgemeinen“). Die Antwort von Sheldon war in dem Fall richtig, aber nicht aus dem richtigen Grund. Bei Alison und Howard konnte wiederum beobachtet werden, dass sie beim Lösen von Task C, Seite 2 deswegen richtig angaben, dass das Hotel im Zentrum war, weil sie nur den Anfang des Inputtextes von Kelly gelesen hatten. Der Textanfang lautete: „L'année dernière, on était dans un hôtel à côté de la Gare Centrale ... qui n'est pas du tout centrale! L'hôtel du Sud est situé juste à côté de la mairie, on a fait tout le centre à pied.“ („Letztes Jahr waren wir in einem Hotel neben dem Hauptbahnhof... der überhaupt nicht zentral ist! Das Hôtel du Sud befindet sich direkt neben dem Rathaus, wir sind durch das ganze Zentrum zu Fuss gegangen.“). Alison und Howard nahmen nur wahr, dass das Hotel neben dem Hauptbahnhof lag, und schlossen daraus, dass das Hotel im Zentrum sein müsse. Ihre Antworten waren richtig, aber ebenfalls nicht aus dem richtigen Grund. Die Stellen in den Inputtexten, die Sheldon, Alison und Howard das richtige Beantworten der Fragen aus falschen Gründen ermöglichten, sollten folglich überarbeitet werden, um die Items für den erfolgreichen Gebrauch dieser *test-wiseness*-Strategie weniger anfällig zu machen.

4.3.2 TEST-WISENESS-STRATEGIEN UND (NICHT) ERFOLGREICHE LÖSUNGEN DER OFF-SZENARIO-AUFGABEN

In Tabelle 18 wird dargestellt, wie und mit welchem Erfolg der Gebrauch der *test-wiseness*-Strategien bei den Off-Szenario-Schüler/innen beobachtet werden konnte. Wie schon bei den Szenario-Schüler/innen werden in der Tabelle die Codierungen im Lautdenken-Teil und im *Stimulated-Recall*-Interview zusammen und pro Item abgebildet. Auch hier wurden diejenigen Items rot markiert, bei denen die Lernenden aufgrund einer oder mehrerer *test-wiseness*-Strategien eine richtige Antwort wählten. Items, bei denen der Einsatz von *test-wiseness*-Strategien zu falschen Resultaten führte, wurden hingegen grün hinterlegt. Im Gegensatz zu den Szenario-Schüler/innen wurde bei den Off-Szenario-Schüler/innen nicht beobachtet, dass sie zwar eine *test-wiseness*-Strategie einsetzten, ihre Lösung aber letztendlich doch aus einem anderen, konstruktrelevanten Grund wählten. Das könnte evtl. damit zu tun haben, dass die Szenario-Schüler/innen die Strategie L03 („Einen Teil des Inputtextes (erneut bzw. genauer) lesen“) fast doppelt so oft einsetzten wie die Off-Szenario-Schüler/innen (vgl. Kap. 4.4.2). Die Off-Szenario-Schüler/innen hatten somit wohl weniger Möglichkeiten, ihre

Fehler zu bemerken und zu korrigieren. Ein weiterer Grund könnte sein, dass die Off-Szenario-Items viel schwieriger waren, und somit die Off-Szenario-Schüler/innen vielleicht auch weniger Gelegenheit hatten, die richtigen Antworten doch noch aus den richtigen, konstruktrelevanten Gründen zu finden.

Tabelle 18: Items, bei denen Off-Szenario-Schüler/innen *test-wiseness*-Strategien einsetzen (gemäss Codierungen im Lautdenkprotokoll und/oder *Stimulated-Recall*-Interview).

Erklärung der Farbsymbolik:

Grün: Der/die Schüler/in wählte aufgrund der eingesetzten *test-wiseness*-Strategie(n) eine falsche Antwort.

Rot: Der/die Schüler/in wählte aufgrund der eingesetzten *test-wiseness*-Strategie(n) eine richtige Antwort.

Szenario	"Walkies"		"Stadttrip"		"Berufswahl"						Anzahl korrekter Antworten mit TW-Strategie(n) pro Schüler/in (Max. 10)	Anzahl falscher Antworten mit TW-Strategie(n) pro Schüler/in (Max. 10)
Aufgabe	Task C	Task C	Task B	Task B	Task C	Task C	Task C	Task C	Task C	Task C		
Bildschirmseite	Seite 1	Seite 2	Seite 1	Seite 2	Seite 1	Seite 1	Seite 1	Seite 2	Seite 2	Seite 2		
Item	<i>deine Aufgabe</i>	<i>Esther</i>	<i>Internet</i>	<i>beeindruckt in Avignon</i>	<i>selbst Produkte testen</i>	<i>mehrere Sprachen sprechen</i>	<i>viel reisen</i>	<i>selbst Produkte testen</i>	<i>mehrere Sprachen sprechen</i>	<i>viel reisen</i>		
Bonnie (7/10) ¹⁰⁷		TW02, TW04		TW01	TW01, TW02	TW02	TW02	TW02, TW07			3	3
Dustin (5/10)	TW02	TW02	TW01		TW01, TW02, TW07		TW02	TW02		TW02, TW07	4	3
Loup (3/10)	TW02	TW02, TW04	TW01	TW01, TW03		TW02	TW02	TW02			0	7
Nancy (4/10)				TW05	TW02		TW02				0	3
Picard (1/10)	TW03, TW06	TW02, TW03	TW01, TW03	TW02, TW03	TW03, TW07	TW02	TW02	TW02	TW02	TW02	1	9
Eleven (2/10)	TW02, TW04	TW02	TW01	TW05	TW01, TW02	TW07		TW02	TW02		0	8
Ezra (5/10)	TW03	TW02			TW01, TW02	TW02	TW02	TW02, TW03		TW02, TW03	2	5
Mogli (3/10)	TW02	TW02, TW03	TW01	TW05	TW03, TW07	TW02	TW02	TW02		TW07	2	7
Omega (6/10)	TW02, TW06	TW02		TW02, TW03	TW01, TW02	TW02		TW02, TW07			2	4
Steve (1/10)	TW02	TW02	TW03, TW04	TW01, TW03	TW07	TW01, TW02	TW02	TW01, TW02	TW01, TW07	TW02	1	9
Anzahl korrekter Antworten mit TW-Strategie(n) pro Item (Max. 10)	4	1	2	2	4	1	0	1	0	0		
Anzahl falscher Antworten mit TW-Strategie(n) pro Item (Max. 10)	4	8	4	6	5	7	8	8	3	5		

¹⁰⁷ Anzahl erreichter Punkte.

Tabelle 18 ist bereits auf den ersten Blick zu entnehmen, dass die Off-Szenario-Schüler/innen die *test-wiseness*-Strategien deutlich häufiger als die Szenario-Schüler/innen einsetzten (vgl. Tabelle 17). Wie in Kap. 4.1 dargestellt, waren jedoch die Items, die die Off-Szenario-Schüler/innen zu lösen bekamen, deutlich schwieriger als diejenigen, die die Szenario-Schüler/innen bearbeiteten. Dies könnte sich durchaus darauf ausgewirkt haben, dass die Off-Szenario-Schüler/innen häufiger von *test-wiseness*-Strategien Gebrauch machten (bzw. machen mussten).

Die Off-Szenario-Schüler/innen waren mit den *test-wiseness*-Strategien in insgesamt 58 von 73 Fällen nicht erfolgreich. Steve und Picard beantworteten trotz ihres Einsatzes von *test-wiseness*-Strategien 9 von 10 Fragen falsch, Eleven 8, Loup und Mogli je 7, Ezra 5, Omega 4, Bonnie, Dustin und Nancy je 3. Dies ist im Prinzip ein gutes Zeichen, weil es bedeutet, dass sich der Test nicht leicht „austricksen“ liess und die Distraktoren gute Arbeit leisteten. Der recht häufige Einsatz der Strategie TW07 („Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats“), der bei der Aufgabe „Berufswahl“ beobachtet werden konnte, deutet allerdings darauf hin, dass gewisse Teile der Instruktionen bei dieser Aufgabe stärker hervorgehoben werden sollten bzw. dass auf solch komplexe Aufgabenstellungen, die wenig intuitiv sind und zuerst länger erklärt werden müssen, in Zukunft lieber verzichtet werden sollte. Es zeigte sich nämlich, dass die Schüler/innen in der Aufgabenstellung (übrigens z.T. auch bei Task C, Szenario „Stadtrip“) die Information „Manchmal findest du in der gleichen Bewertung mehrere Antworten“ oft überlesen und deswegen dachten, dass sie bei der Aufgabe alle drei Inputtexte allen drei Fragen zuordnen mussten, was dann zu falschen Lösungen führte. Ausserdem zeigte sich (wie schon bei der Szenario-Gruppe) auch bei den Off-Szenario-Schüler/innen, dass bei Task B, Seite 2, Szenario „Stadtrip“ die Strategie TW05 von drei Schüler/innen angewandt wurde, um die richtige Antwort „der starke Regen in der Stadt“ auszuschliessen. Auch diese Off-Szenario-Schüler/innen waren der Meinung, dass diese Antwortoption zur Frage „Was hat Lou in Avignon besonders beeindruckt?“ wohl unmöglich passen kann. Wie bereits erwähnt, sollte dieses Item deswegen angepasst werden.

In 15 von 73 Fällen führte der Einsatz von *test-wiseness*-Strategien jedoch zur richtigen Lösung eines Items. So verdankte Dustin sogar 4 ihrer 5 richtigen Antworten den *test-wiseness*-Strategien, Bonnie 3 von 7, Ezra 2 von 5, Omega 2 von 6, und Picard und Steve die jeweils einzigen richtigen Antworten, die sie im Test auswählten. In den meisten Fällen, in denen der Einsatz einer *test-wiseness*-Strategie zu einer richtigen Lösung führte, war der Erfolg dem Zufall zu verdanken, d.h. die Schüler/innen hatten beim Raten einfach Glück oder entschieden sich aufgrund eines Wortes zufällig für die richtige Lösung und nicht für den Distraktor. Die von Omega und Picard bei Item „deine Aufgabe“ (Aufgabe „Walkies“, Seite 1)

mit Erfolg eingesetzte Strategie TW06 „Auswahl einer anders aussehenden Antwortoption“, die bereits ausführlich in Kap. 4.2.2.2 beschrieben wurde, zeigte allerdings, dass diese Aufgabe überarbeitet werden sollte.

Resümee

Für die Beantwortung der zweiten Forschungsfrage wurden alle Codierungen der *test-wiseness*-Strategien in den Transkripten gründlich untersucht. Es wurde dabei eingeschätzt, ob 1) aufgrund der codierten *test-wiseness*-Strategie(n) eine falsche Antwort gewählt wurde, 2) aufgrund der codierten *test-wiseness*-Strategie(n) die Frage zum Text richtig beantwortet wurde, oder 3) beim Bearbeiten dieses Items zwar eine oder mehrere *test-wiseness*-Strategie(n) eingesetzt wurden, das Item aber letztendlich aus einem anderen, konstruktrelevanten Grund richtig gelöst werden konnte.

Die Resultate zeigen, dass bis auf den Szenario-Schüler Toby alle 19 Schüler/innen von den *test-wiseness*-Strategien Gebrauch machten. In den meisten Fällen führte der Einsatz dieser Strategien jedoch zu einer falschen Antwort. Aus der Sicht der Testentwickler/innen ist dies ein positives Zeichen, denn es bedeutet, dass die Testaufgaben gegen den Einsatz von *test-wiseness*-Strategien überwiegend immun waren und die Distraktoren gut funktionierten. In den vergleichsweise seltenen Fällen, in denen die Aufgaben aufgrund der *test-wiseness*-Strategien korrekt gelöst wurden, handelte es sich vor allem um Zufälle. Es konnten aber dank der Analyse der als *test-wiseness*-Strategien codierten Transkriptstellen auch einige Items identifiziert werden, die den Einsatz von *test-wiseness*-Strategien begünstigten und deswegen überarbeitet werden mussten (was sich zu diesen Items auch schon in den Resultaten zur ersten Forschungsfrage zeigte).

Der Vergleich der Codierungen, die in den Transkripten der Szenario- und Off-Szenario-Schüler/innen vorgenommen wurden, zeigt, dass die Off-Szenario-Schüler/innen deutlich häufiger auf *test-wiseness*-Strategien zurückgriffen. Dies könnte darauf zurückzuführen sein, dass die Off-Szenario-Gruppe deutlich schwierigere Aufgaben als die Szenario-Gruppe löste, was erst im Nachhinein – d.h. nach der Auswertung der Daten aus der IFB-Studie – festgestellt werden konnte.

4.4 ZU FORSCHUNGSFRAGE 3: STRATEGIEN BEIM LÖSEN DER SZENARIOBASierten UND NICHT SZENARIOBASierten TESTAUFGABEN

Die Antwort auf die dritte Forschungsfrage soll Erkenntnisse dazu liefern, ob sich Unterschiede beim Einsatz der Strategien zeigen, wenn szenariobasierte und nicht

szenariobasierte Testaufgaben gelöst werden, und wenn ja, worauf dies zurückgeführt werden könnte. Um diese Frage zu beantworten, werden zuerst in Kap. 4.4.1 die Resultate der Analyse der Lautdenkprotokolle und *Stimulated-Recall*-Interviews dargestellt, die daraufhin untersucht wurden, 1) ob die Szenario-Schüler/innen die Szenarioeinbettung der Testaufgaben überhaupt bemerkten und wie sie diese einschätzten, und 2) wie die Off-Szenario-Schüler/innen auf den jeweils anderen inhaltlichen Fokus der von ihnen bearbeiteten Aufgaben reagierten und was sie davon hielten, dass die von ihnen bearbeiteten Aufgaben einen jeweils anderen inhaltlichen Fokus hatten. Anschliessend wird in Kap. 4.4.2 und 4.4.3 die quantitative Evidenz zu den Strategien dargestellt, die in den Verbalprotokollen der Szenario- und Off-Szenario-Schüler/innen codiert wurden. In Kap. 4.4.4 wird (ebenfalls aus einem quantitativen Blickwinkel) illustriert, wie sich der Gebrauch der häufigsten Strategien bei der Bearbeitung der äquivalenten Szenario- und der Off-Szenario-Aufgaben voneinander unterschied.

Bevor im Folgenden die Ergebnisse präsentiert werden, sei an dieser Stelle nochmals daran erinnert, dass die Quantifizierungen der codierten Strategien lediglich Tendenzen in den Daten illustrieren sollen und nicht den gleichen Stellenwert wie die Zahlen in quantitativen Studien haben. Ausserdem muss hier auf die in Kap. 4.1 dargestellten Resultate zur Auswertung der untersuchten Testaufgaben zurückverwiesen werden: Der Vergleich der nachträglich ermittelten Itemschwierigkeitswerte der Szenario- und Off-Szenario-Aufgaben zeigte, dass die Off-Szenario-Schüler/innen schwierigere Aufgaben zu bewältigen hatten, als dies bei den Szenario-Schüler/innen der Fall war, was sich durchaus auf die angewandten Strategien ausgewirkt haben könnte.

4.4.1 WAHRNEHMUNG DER SZENARIO- UND OFF-SZENARIO-TESTAUFGABEN DURCH DIE SCHÜLER/INNEN

Um zu erfassen, ob die 10 Szenario-Schüler/innen beim Lösen der Testaufgaben die Szenarioeinbettung überhaupt wahrnahmen, wurde für die Codierung der Lautdenkprotokolle die Kategorie „Entdeckung des Szenarios“ vorgesehen (vgl. Kap. 3.7.3.3). Es wurde vermutet, dass eine bewusste Aufnahme des Szenarios eventuell einen Einfluss auf die Herangehensweisen der Schüler/innen beim Lösen der Testaufgaben haben könnte.

Bei der Analyse der Lautdenkprotokolle wurde nur im Transkript der Szenario-Schülerin Aria eine Stelle gefunden, die auf die Wahrnehmung des Szenarios beim Lesen der Einführungsseite zu Task D hindeutet:

```
((liest laut vor)) am wochenende in dem du in avignon bist findet= *es
geht ja wieder um das das gleiche* ((liest weiter vor)) (22:5)
(Aria_Szenario, Pos. 243-244)
```

Beim Durchführen der *Stimulated-Recall*-Interviews wurde in einem dreistufigen Verfahren zu ermitteln versucht, ob die Schüler/innen die Einbettung der Aufgaben in ein Szenario zur Kenntnis nahmen. Die Lernenden wurden als Erstes gefragt, ob ihnen an den Aufgaben etwas Spezielles aufgefallen sei. Die Analyse zeigte, dass die Szenarioeinbettung in keiner der Antworten auf diese Frage erwähnt wurde. Dafür wurde die Szenarioeinbettung von Aria, die schon beim Lauten Denken den inhaltlichen Zusammenhang der Aufgaben entdeckte, spontan im Interview angesprochen. Dies wurde mit dem Code „Szenario beim Test wahrgenommen“ gekennzeichnet:

```
S: *aber (sonst) glaube ich das geht jetzt alles mehr um EIN thema`
Avignon;* [...]
MB: (das ist) dir Aufgefallen` dass das (--) alles um das gleiche thema
geht-
S: ja; (Aria_Szenario, Pos. 746-749)
```

Auf die Frage der Forscherin, wie sie diesen thematischen Zusammenhang empfand, antwortete Aria, dass es für sie „irgendwie besser“ war. Sie begründete es wie folgt:

```
*man bekommt irgendwie das gefühl man verSTEht mehr- weil man ein pAAr
sachen schon von vorher gehÖRt hat [...] Oder (--) die einem vielleicht
bekANNt vorkommen oder weil man sich die dinge besser (vorstellt)*
(Aria_Szenario, Pos. 753-757)
```

Die neun restlichen Szenario-Schüler/innen wurden im Anschluss an die erste Frage zu Auffälligkeiten der Testaufgaben gefragt, ob sie am Thema der Testaufgaben etwas Besonderes bemerkt hatten. Wenn sie daraufhin das Szenario (bzw. Avignon, Reisen etc.) erwähnten, wurden ihre Antworten mit der Kategorie „Szenario nach einem Hinweis wahrgenommen“ und Gewichtung „0“ codiert. Dies war bei sechs der neun Schüler/innen der Fall. Arya sagte in ihrer Antwort auf die Frage der Forscherin beispielsweise:

```
MB: okay (-) gut (--) und (-) wegen dem thEma ist dir hier etwas [...]
vielleicht aufgefallen`oder nicht unbedingt`
S: em: es gab ja verschiedene themen also mal HOtel- es gab SEHR viel
eben zu zu dieser stadt- ich glaube beinahe alles war von der stadt-
und das finde ich eigentlich noch cOol; denn man wa:r wirklich die
ganze zeit wie in einem thema drin- man muss nicht immer switchen [...]
und das fand ich ANgenehm; (Arya_Szenario, Pos. 1905-1913)
```

Wenn in der Antwort auf die Frage, ob den Schüler/innen etwas vom Thema her an den Aufgaben aufgefallen war, keine Stichwörter wie „Reisen“, „Avignon“, „Planung“ etc. genannt wurden, wurden die Lernenden letztlich direkt gefragt, ob sie wahrgenommen hatten, dass es sich bei allen Aufgaben um die Planung einer Schulreise nach Avignon handelte. Die so gestellte Frage wurde dann auch von den drei restlichen Schüler/innen bejaht, was mit der

Kategorie „Szenario nach einem Hinweis wahrgenommen“ und Gewichtung „30“ codiert wurde. Mit diesem Code wurde u.a. die folgende Aussage von Cersei erfasst:

```
MB: ist dir (-) vom THEma her etwas aufgefallen?  
S: ja einfach LEsen (-) und eben mit den personen sie SAgen das (-)  
ihre MEInungen [...] ja-  
MB: okey (-) ähm (-) ich FRAge (-) weil eigentlich geht es in ALlen  
diesen aufgaben um avi/ um avignon.  
S: ja avignon [ja geNAU;]  
MB: [es kommt immer wieder] (-) es kommt immer wieder vor so als (-)  
oberTHEma (-) ist das dir AUFGefallen oder-  
S: ja (-) eigentlich schon dass immer (-) avignon (-) aber [...] mir ist  
mal nicht in den sinn gekommen dass das das gleiche (-) war;  
MB: also du hast es nicht (-) beWUSST;  
S: nein (-) aber ich habe schon immer ah avignon avignon (-) ja;  
MB: und wie FINdest du [...] dass es in diesen aufgaben ein thema gibt  
dass in allen (-) vier aufgaben vorkommt-  
S: das finde ich eigentlich gut (-) ja;  
MB: hast du das LIEber so (-) als wenn du ähm (-) in jeder aufgabe  
wieder ein anderes thema hättest?  
S: nein ich fand es gut dass es halt immer ein OBerthema gibt aber  
trotzdem verschiedene sachen die man dazu machen kann;  
(Cersei_Szenario, Pos. 1359-1379)
```

Aus den Gesprächen ging hervor, dass die meisten der 10 befragten Schüler/innen die Einbettung der Aufgaben in ein übergreifendes Szenario – ähnlich wie Arya und Cersei – als gut einschätzten. Howard argumentierte, dass man so die Zusammenhänge leichter erkennen könne:

```
das finde ich GUT weil man (-) [...] nicht in zwei verschiedenen welten  
dann sozusagen den test muss machen weil man einfach immer das (-)  
GLEICHe (--) ja (-) THEma hat und (-) man dann auch zusa/ zuSAMMenhänge  
finden kann (Howard_Szenario, Pos. 2012-2015)
```

Drei Schüler/innen scheint die Szenarioeinbettung aber auch egal gewesen zu sein: Cybermen hatte dazu keine feste Meinung, weil er die Stadt Avignon nicht kannte; Böbu kam nach einer kurzen Überlegung zu dem Schluss, dass er im Prinzip auch nichts gegen Testaufgaben mit unterschiedlichen Themen hätte, und Sheldon war es schlicht gleichgültig.

Die Off-Szenario-Schüler/innen wurden ihrerseits gefragt, wie sie dazu standen, dass es in allen fünf von ihnen bearbeiteten Aufgaben jeweils um ein anderes Thema ging. Alle 10 Schüler/innen fanden es gut und abwechslungsreich. Bonnie z.B. nahm den Themenwechsel beim Bearbeiten der Testaufgaben nicht wirklich wahr, fand es aber im Nachhinein doch spannend:

```
*also (-) ich habe eigentlich den unterschied nicht gross gemerkt (-)  
[...] also es ist eigentlich (-) irgendwie auch SPANNend gewesen (-) man  
löst nicht immer das GLEICHe (-) muss nicht immer auf das gleiche  
befassen sondern hat mal etwas anderes.* (Bonnie_Off, Pos. 1456-1461)
```

Omega schätzte die Abwechslung ebenfalls und vermutete sogar, dass es schwieriger wäre, wenn sich alle Testaufgaben um das gleiche Thema drehen würden:

```
es war (-) gute abwechslung' und (.) [...] wenn es jetzt IMMER die gleichen themen gewesen wären hätte man vielleicht die einzelnen durcheinander gebracht (Omega_Off, Pos. 1506-1509)
```

Loup vermutete wiederum, dass es uninteressant wäre, wenn alle Aufgaben das gleiche Thema behandeln würden:

```
wäre es NUR ein thema wäre es wahrscheinlich LANGweilig (Loup_Off, Pos. 1551)
```

Eine ganz nüchterne Einschätzung kam schliesslich von Mogli, der darauf hinwies, dass er sich bei einem Test doch weder die Themen der Aufgaben noch ihre eventuelle Einbettung aussuchen kann:

```
MB: und was sagst du dazu dass: em (---) die (--) tHEmen in den aufgaben unterschiedlich waren`  
S: pfu:: *also ich finde das stö/ also (--) ist mir eigentlich egal;*  
es kommt ja (-) wenn ich ein tEst mache kann ich ja *SOWIEso* nicht entscheiden (--) welche themen kommen; (Mogli_Off, Pos. 1699-1704)
```

Zusammenfassend lässt sich feststellen, dass die Szenarioeinbettung der IFB-Aufgaben bis auf Aria wohl von niemanden aus der Szenario-Gruppe bewusst wahrgenommen wurde. Man könnte also vermuten, dass ein eventueller, positiver Effekt der Szenarioeinbettung auf die Auseinandersetzung mit den Testaufgaben nur bei dieser Schülerin vorliegen würde. Ob dies aber tatsächlich der Fall war und ob die Szenarioeinbettung vielleicht auch dann einen positiven Einfluss auf die Beschäftigung mit den Testaufgaben gehabt hätte, wenn sie nicht bewusst wahrgenommen wurde, lässt sich aufgrund der vorliegenden Daten nicht sagen. Immerhin wurde aber der thematische Zusammenhang der Szenario-Aufgaben von den meisten Schüler/innen im Nachhinein positiv eingeschätzt.

Die Off-Szenario-Schüler/innen schienen sich ihrerseits nicht dadurch benachteiligt gefühlt zu haben, dass sie Aufgaben zu unterschiedlichen Themen zu lösen bekamen. Einige vermuteten sogar, dass das Gegenteil nicht ideal wäre.

Weitere Indizien zum eventuellen Einfluss der Szenarioeinbettung kann der Vergleich der Lerner- und Teststrategien ans Licht bringen, die beim Lauten Denken und beim *Stimulated Recall* von den Szenario- und Off-Szenario-Schüler/innen eingesetzt wurden. Der Vergleich der Strategiecodierungen in beiden Gruppen wird in Kap. 4.4.2 (Lernerstrategien) und 4.4.3 (Teststrategien) dargestellt. In Kap. 4.4.4 wird zusätzlich veranschaulicht, welche Strategien bei den vergleichbaren Szenario- und Off-Szenario-Aufgaben am häufigsten beobachtet werden konnten.

4.4.2 LERNERSTRATEGIEN DER SZENARIO- UND OFF-SZENARIO-SCHÜLER/INNEN

Die Unterschiede, die sich aufgrund der Anzahl Codierungen der Lernerstrategien zwischen der Szenario- und Off-Szenario-Gruppe feststellen lassen, werden in Tabelle 19 dargestellt.

Der Vergleich der Gesamtzahl der codierten Lernerstrategien pro Gruppe zeigt, dass in beiden Gruppen insgesamt eine ähnliche Zahl von Strategien codiert wurde (227 und 223 Codierungen in den Lautdenkprotokollen sowie 120 und 107 Codierungen im *Stimulated Recall*). Die genauere Betrachtung der einzelnen Codierungen pro Strategie und Gruppe weist jedoch auf einige kleine Unterschiede zwischen den Schüler/innen-Gruppen hin. Aus den Codierungen geht beispielsweise hervor, dass die Off-Szenario-Schüler/innen die Texte häufiger sorgfältig lasen (Strategie L01), als dies die Schüler/innen der Szenario-Gruppe machten. Letztere nahmen wiederum häufiger einzelne Teile der Inputtexte wiederholt oder genauer zur Kenntnis (Strategie L03). Auch wurde beim Lauten Denken in der Gruppe der Off-Szenario-Schüler/innen öfters beobachtet, dass sie auf unbekannte Wörter, Satz- oder Textteile stiessen (Strategie LD-L10). Sie schätzten sich bzw. ihr Textverständnis auch häufiger kritisch ein (Strategie LD-L18), als es bei den Schüler/innen der Szenario-Gruppe zu beobachten war.

Tabelle 19: Häufigkeit der Lernerstrategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“¹⁰⁸ und „25“¹⁰⁹.

Kurzbezeichnung Code	Code LD	Szenario-SuS	Off-Szenario-SuS	Total	Code SR	Szenario-SuS	Off-Szenario-SuS	Total
Sorgfältiges Lesen	LD-L01	34 (40% ¹¹⁰)	52 (60%)	86	SR-L01	6 (33%)	12 (67%)	18
Schnelles Lesen	LD-L02	28 (60%)	19 (40%)	47	SR-L02	20 (54%)	17 (46%)	37
Einen Teil des Inputtextes (erneut bzw. genauer) lesen	LD-L03	60 (63%)	36 (38%)	96	SR-L03	22 (59%)	15 (41%)	37
Weiterlesen, um Nichtverstandenes zu verstehen	LD-L04	0	0	0	SR-L04	3 (50%)	3 (50%)	6
Bisheriges Verständnis hinterfragen	LD-L05	3 (75%)	1 (25%)	4	SR-L05	0	0	0
Bisheriges Verständnis korrigieren	LD-L06	9 (82%)	2 (18%)	11	SR-L06	2 (67%)	1 (33%)	3
Bisheriges Verständnis bestätigen	LD-L07	1 (100%)	0	1	SR-L07	0	0	0
Intensive Auseinandersetzung mit Teilen des Inputtextes	LD-L08	61 (49%)	63 (51%)	124	SR-L08	10 (71%)	4 (29%)	14
Informationen verknüpfen	LD-L09	1 (33%)	2 (67%)	3	SR-L09	3 (19%)	13 (81%)	16
Unbekannte Wörter, Satz- bzw. Textteile identifizieren	LD-L10	16 (39%)	25 (61%)	41	SR-L10	18 (55%)	15 (45%)	33
Unbekannte Wörter ausblenden	LD-L11	1 (100%)	0	1	SR-L11	11 (69%)	5 (31%)	16
Wortbedeutung identifizieren	LD-L12	3 (43%)	4 (57%)	7	SR-L12	0	5 (100%)	5
Morphosyntaktische Regeln anwenden	LD-L13	1 (33%)	2 (67%)	3	SR-L13	10 (91%)	1 (9%)	11
Interpunktionskenntnisse nutzen	LD-L14	0	1 (100%)	1	SR-L14	1 (100%)	0	1
Kognaten identifizieren und nutzen	LD-L15	4 (67%)	2 (33%)	6	SR-L15	11 (42%)	15 (58%)	26
Prosodie beim Hören nutzen	LD-L16	0	0	0	SR-L16	2 (100%)	0	2
Relevantes Hintergrundwissen einbeziehen	LD-L17	2 (67%)	1 (33%)	3	SR-L17	1 (50%)	1 (50%)	2
Selbsteinschätzung	LD-L18	3 (19%)	13 (81%)	16	SR-L18	0	0	0
Total		227 (50%)	223 (50%)	450		120 (53%)	107 (47%)	227

¹⁰⁸ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

¹⁰⁹ Codierungen der Strategie L10 („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“), die bei den mündlichen Zusammenfassungen der Inputtexte vorgenommen wurden.

¹¹⁰ Die Prozentangaben wurden überall abgerundet.

Solche Gruppenvergleiche haben den Nachteil, dass darin die individuellen Unterschiede zwischen den einzelnen Schüler/innen nicht zum Vorschein kommen: Es kann sein, dass einige wenige Personen pro Gruppe, die gewisse Strategien besonders häufig anwenden, zur verzerrten Wahrnehmung der Gruppenergebnisse führen. Um die Unterschiede zwischen dem Einsatz der Lernerstrategien durch einzelne Schüler/innen der Szenario- (SZ) und Off-Szenario-Gruppe (OFF) sichtbarer zu machen, werden in Abbildung 23 sieben ¹¹¹ Lernerstrategien pro Schüler/in dargestellt, die in beiden Gruppen am häufigsten codiert wurden. Die Zahlen in Abbildung 23 beziehen sich, wie bereits weiter oben, aus technischen Gründen auf die totale Anzahl der codierten Lernerstrategien, d.h. die Codierungen mit den speziellen Gewichtungen „2“¹¹² und „25“¹¹³ wurden hier nicht abgezogen (vgl. Erklärung in Kap. 4.2.3). Aus diesem Grund wurde die Strategie L10 auch aus Abbildung 23 ausgeschlossen, bei der die verhältnismässig sehr häufig codierte Gewichtung „25“ nicht abgezogen werden konnte.

¹¹¹ In der Grafik werden nur die sieben häufigsten Strategien dargestellt, weil die Abbildung einer höheren Zahl codierter Strategien die Grafik unlesbar machen würde.

¹¹² Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

¹¹³ Codierungen der Strategie L10 („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“), die bei den mündlichen Zusammenfassungen der Inputtexte vorgenommen wurden.

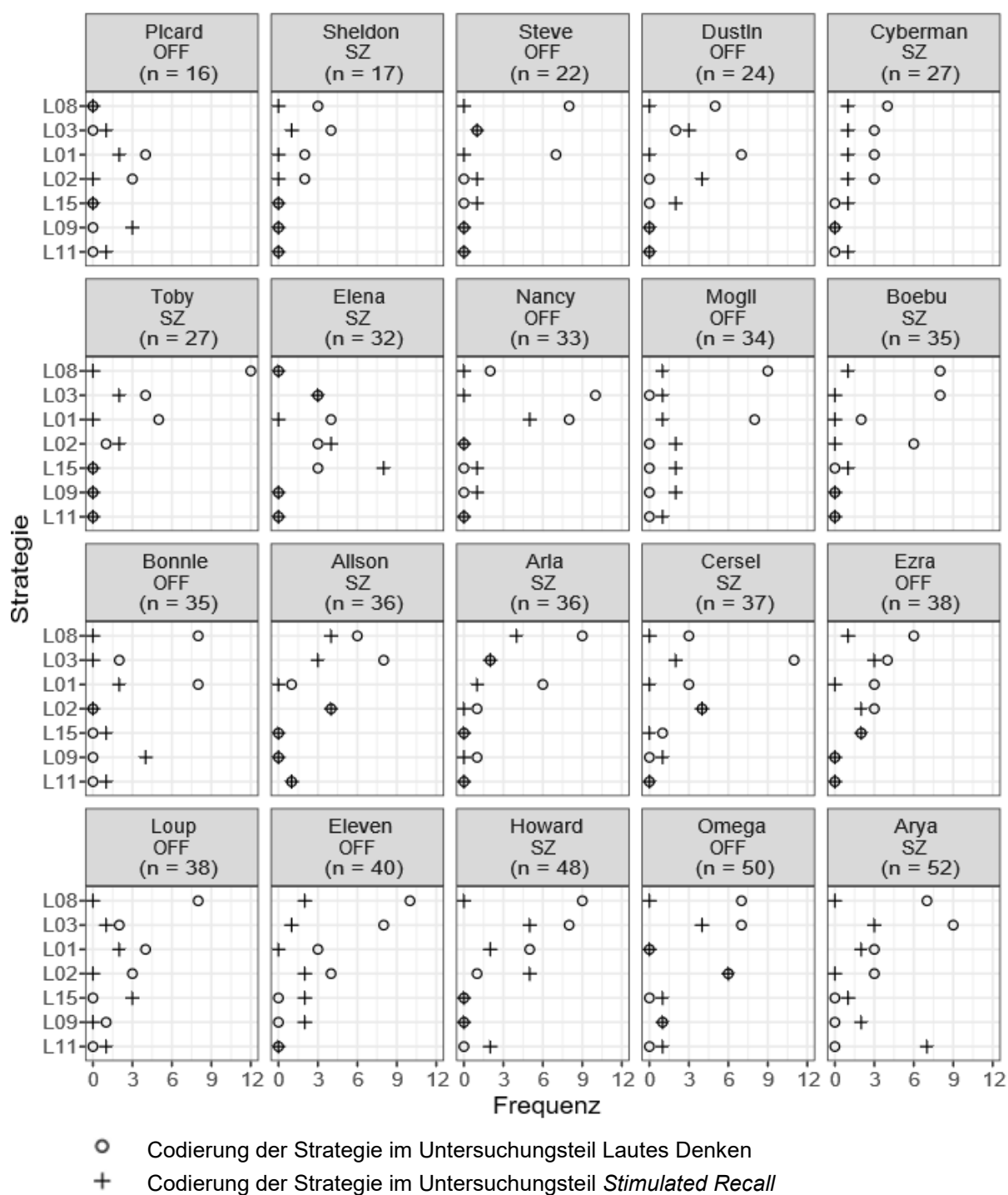


Abbildung 23: Sieben am häufigsten codierte Lernerstrategien pro Szenario- (SZ) und Off-Szenario-Schüler/innen (OFF), ohne Strategie L10.

In Abbildung 23 wurden die Schüler/innen nach der Anzahl der von ihnen eingesetzten Lernerstrategien geordnet. Es zeigt sich, dass die Lernenden der beiden Gruppen in Bezug auf die Anzahl der Codierungen einer bestimmten Lernerstrategie wie auch auf die totale Anzahl der eingesetzten Strategien bunt durchmischt sind. Dies legt die Vermutung nahe, dass der Einsatz von Lernerstrategien wohl eher mit individuellen Präferenzen der einzelnen Schüler/innen und womöglich auch mit ihrer Französischkompetenz zu tun haben könnte, als

mit ihrer Zugehörigkeit zur Szenario- oder Off-Szenario-Gruppe bzw. mit den bearbeiteten Aufgaben.

4.4.3 TESTSTRATEGIEN DER SZENARIO- UND OFF-SZENARIO-SCHÜLER/INNEN

Zu den Teststrategien gehören die Testmanagement- und die *test-wiseness*-Strategien. Dem Vergleich der codierten Testmanagementstrategien der Szenario- und Off-Szenario-Schüler/innen ist Kap. 4.4.3.1 gewidmet. In Kap. 4.4.3.2 werden die Unterschiede in den Codierungen der *test-wiseness*-Strategien in beiden Gruppen dargestellt.

4.4.3.1 TESTMANAGEMENTSTRATEGIEN DER SZENARIO- UND OFF-SZENARIO-SCHÜLER/INNEN

Tabelle 20 zeigt die Unterschiede auf, die zwischen den codierten Testmanagementstrategien der Szenario- und Off-Szenario-Gruppe festgestellt werden konnten.

Bereits der erste Blick auf die Gesamtzahl der codierten Testmanagementstrategien pro Gruppe zeigt, dass die Szenario-Schüler/innen diese Strategien deutlich häufiger als die Off-Szenario-Schüler/innen einsetzten (445 vs. 276 Codierungen beim Lauten Denken und 327 vs. 244 Codierungen beim *Stimulated Recall*). Die Szenario-Schüler/innen lasen die Instruktionen bzw. Antwortoptionen (Strategie TM01) beim Lautdenken-Teil fast doppelt so oft wie die Off-Szenario-Schüler/innen (192 vs. 99 Codierungen). Häufiger bzw. deutlich häufiger machten sie auch von Strategien wie Fokus auf relevante Textteile (Strategie TM06) Gebrauch, setzten das Verstandene bei der Auswahl oder beim Ausschluss der Antwortoptionen ein (Strategie TM08) oder stützten sich dabei auf Inferenzen (TM12). Ausserdem wurde die Testmanagementstrategie „Kenntnisse über das Itemformat nutzen“ (TM15) nur bei Szenario-Schüler/innen codiert, was allerdings darauf zurückzuführen ist, dass die Codierung dieser Strategie für ein bestimmtes Itemformat vorgesehen war, das nur in Szenario „Stadttrip“, Task A zum Einsatz kam. Vergleicht man die Codierungen der Testmanagementstrategien in den beiden Gruppen im *Stimulated-Recall*-Teil, sind die Unterschiede zwischen den beiden Gruppen zwar immer noch sichtbar, aber deutlich kleiner.

Tabelle 20: Häufigkeit der Testmanagementstrategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“¹¹⁴.

Kurzbezeichnung Code	Code LD	Szenario-SuS	Off-Szenario-SuS	Total	Code SR	Szenario-SuS	Off-Szenario-SuS	Total
Instruktionen und/oder Antwortoptionen (erneut) lesen	LD-TM01	192 (66% ¹¹⁵)	99 (34%)	291	SR-TM01	72 (54%)	62 (46%)	134
Paraphrasieren und/oder Übersetzen der Aufgabe	LD-TM02	14 (44%)	18 (56%)	32	SR-TM02	9 (56%)	7 (44%)	16
Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren	LD-TM03	0	2 (100%)	2	SR-TM03	5 (45%)	6 (55%)	11
Vorhersage machen	LD-TM04	2 (100%)	0	2	SR-TM04	4 (80%)	1 (20%)	5
Frage und Antwortoptionen beim Hören anschauen	LD-TM05	19 (49%)	20 (51%)	39	SR-TM05	3 (100%)	0	3
Fokus auf relevante Textteile	LD-TM06	56 (60%)	37 (40%)	93	SR-TM06	55 (51%)	53 (49%)	108
Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären	LD-TM07	4 (80%)	1 (20%)	5	SR-TM07	1 (33%)	2 (67%)	3
Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)	LD-TM08	72 (60%)	48 (40%)	120	SR-TM08	69 (58%)	51 (43%)	120
Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden	LD-TM09	1 (33%)	2 (67%)	3	SR-TM09	11 (73%)	4 (27%)	15
Richtigkeit der Antwortoption überprüfen	LD-TM10	11 (79%)	3 (21%)	14	SR-TM10	3 (75%)	1 (25%)	4
Antwortoptionen miteinander vergleichen	LD-TM11	19 (49%)	20 (51%)	39	SR-TM11	13 (39%)	20 (61%)	33
Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen	LD-TM12	20 (77%)	6 (23%)	26	SR-TM12	19 (58%)	14 (42%)	33
Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war	LD-TM13	18 (69%)	8 (31%)	26	SR-TM13	14 (64%)	8 (36%)	22
Auswahl einer Antwortoption, weil andere nicht plausibel sind	LD-TM14	9 (64%)	5 (36%)	14	SR-TM14	14 (67%)	7 (33%)	21
Kenntnisse über das Itemformat nutzen	LD-TM15	6 (100%)	0	6	SR-TM15	8 (100%)	0	8
Zeitmanagement	LD-TM16	0	0	0	SR-TM16	3 (100%)	0	3
Selbstkorrektur F-R	LD-TM17	9 (82%)	2 (18%)	11	SR-TM17	3 (100%)	0	3
Selbstkorrektur R-F	LD-TM18	3 (60%)	2 (40%)	5	SR-TM18	0	0	0
Selbstkorrektur F-F	LD-TM19	0	3 (100%)	3	SR-TM19	0	0	0
Selbstkorrektur beim SR	-	-	-	-	SR-TM20	21 (72%)	8 (28%)	29
Total		455 (62%)	276 (38%)	731		327 (57%)	244 (43%)	571

¹¹⁴ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

¹¹⁵ Die Prozentzahlen wurden überall abgerundet.

Abbildung 24 zeigt, wie die sieben am häufigsten codierten Testmanagementstrategien über einzelne Schüler/innen hinweg verteilt sind. Wie schon in Abbildung 23 (Kap. 4.4.2) erklärt, beziehen sich die hier dargestellten Zahlen auf die totale Anzahl der codierten Lernerstrategien, d.h. die Codierungen mit der Gewichtung „2“¹¹⁶ wurden hier nicht abgezogen (vgl. Erklärung in Kap. 4.2.3).

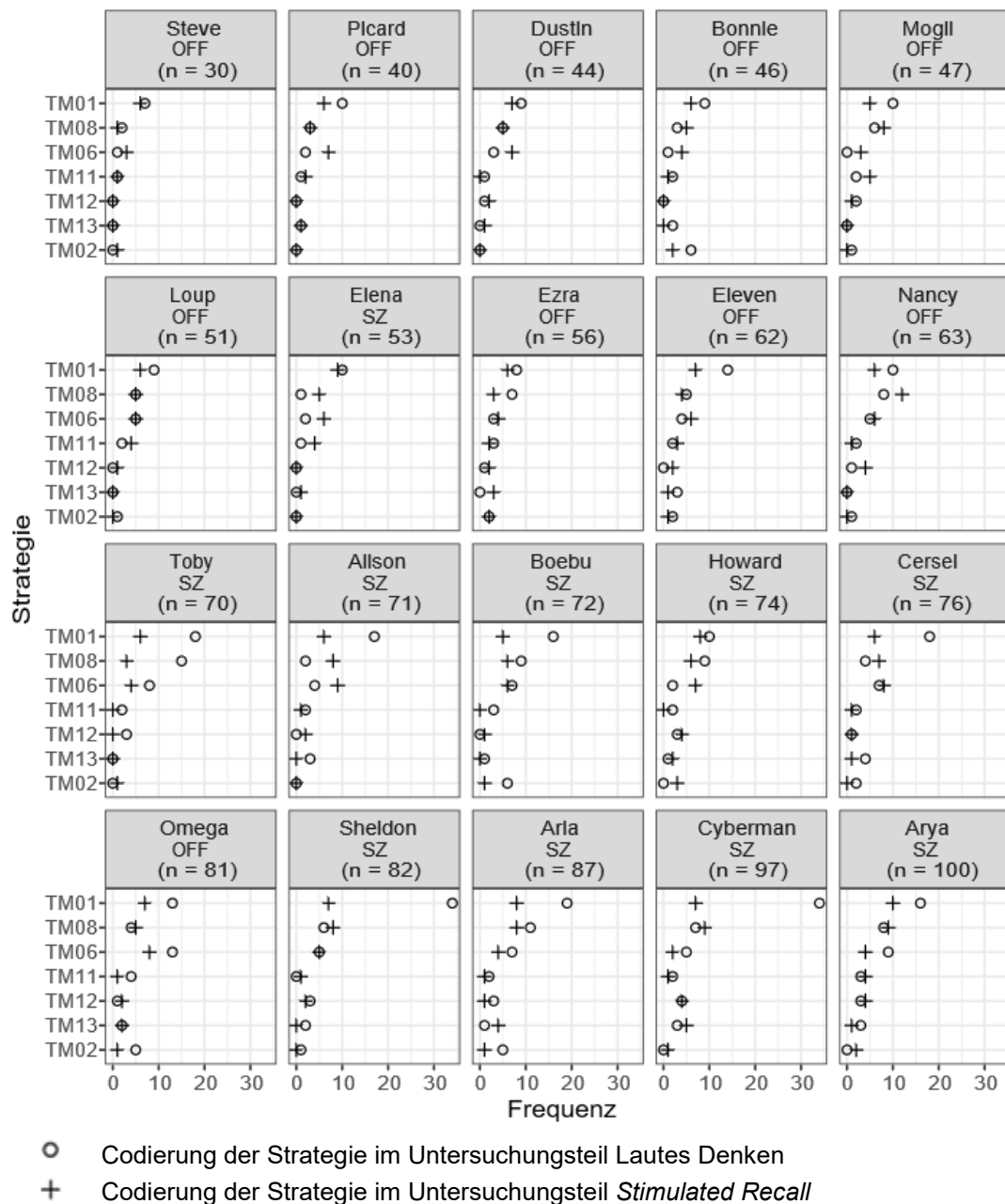


Abbildung 24: Sieben am häufigsten codierte Testmanagementstrategien pro Szenario- (SZ) und Off-Szenario-Schüler/innen (OFF).

Abbildung 24 ist zu entnehmen, dass die Testmanagementstrategien bei den Off-Szenario-Schüler/innen generell seltener und bei den Szenario-Schüler/innen generell häufiger

¹¹⁶ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

beobachtet und codiert wurden. Es gab dabei zwei Ausnahmen: Die Szenario-Schülerin Elena setzte weniger Strategien als die restlichen Lernenden der Szenario-Gruppe ein, und beim Off-Szenario-Schüler Omega wurden 18 Strategien mehr codiert als bei Nancy, bei der die Testmanagementstrategien in der Off-Szenario-Gruppe am zweithäufigsten codiert wurden.

Die oben dargestellten Resultate lassen vorsichtig vermuten, dass der Einsatz der Testmanagementstrategien durchaus mit den unterschiedlichen Aufgaben zusammenhängen könnte, die von den Schüler/innen der beiden Gruppen gelöst wurden. Es bleibt allerdings offen, auf welche konkreten Eigenschaften der jeweiligen Aufgaben dies zurückgeführt werden könnte. Der nachträgliche Vergleich der Itemschwierigkeitswerte der Szenario- und Off-Szenario-Aufgaben zeigte, dass die Schüler/innen der Szenario-Gruppe nicht nur thematisch verbundene, sondern auch einfachere Aufgaben als die Schüler/innen der Off-Szenario-Gruppe lösten (vgl. Kap. 4.1). Es wäre daher nicht richtig anzunehmen, dass die Unterschiede im Einsatz der Testmanagementstrategien der Szenario- und der Off-Szenario-Schüler/innen lediglich auf die thematische Einbettung der Szenario-Aufgaben und deren Fehlen bei den Off-Szenario-Aufgaben zurückzuführen sind. Ausserdem wurde zwar im Task A, Szenario „Stadtrip“ und in der Off-Szenario-Aufgabe „Walkies“ die gleiche digitale Textsorte „Messenger-Chat“ angewandt, die Itemformate waren jedoch in beiden Aufgaben anders. Somit lässt sich aufgrund der vorliegenden Daten nicht mit Sicherheit sagen, ob die Unterschiede in den Codierungen der Testmanagementstrategien in beiden Gruppen auf die (im Fall der Off-Szenario-Aufgaben fehlende) thematische Einbettung der Tasks zurückzuführen sind; ob die unterschiedlichen Herangehensweisen der Szenario- und Off-Szenario-Schüler/innen vielleicht eher mit der Itemschwierigkeit der gelösten Items zu tun hatten; ob dies eventuell von den Itemformaten beeinflusst wurde; ob sich hier die individuellen Unterschiede zwischen den Schüler/innen auswirkten; oder ob sogar mehrere oder alle diese Faktoren dabei eine Rolle gespielt haben könnten.

4.4.3.2 TEST-WISENESS-STRATEGIEN DER SZENARIO- UND OFF-SZENARIO-SCHÜLER/INNEN

In Tabelle 21 werden die Differenzen zwischen der Szenario- und Off-Szenario-Gruppe veranschaulicht, die sich in Bezug auf die Codierung der *test-wisness*-Strategien in den beiden Gruppen feststellen lassen.

Tabelle 21: Häufigkeit der *test-wiseness*-Strategien der Szenario- und Off-Szenario-Schüler/innen nach Abzug der Codierungen mit Gewichtung „2“¹¹⁷.

Kurzbezeichnung Code	Code LD	Szenario-SuS	Off-Szenario-SuS	Total	Code SR	Szenario-SuS	Off-Szenario-SuS	Total
Raten	LD-TW01	2 (25% ¹¹⁸)	6 (75%)	8	SR-TW01	6 (24%)	19 (76%)	25
Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war	LD-TW02	22 (31%)	49 (69%)	71	SR-TW02	24 (33%)	48 (67%)	72
Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern	LD-TW03	9 (53%)	8 (47%)	17	SR-TW03	13 (48%)	14 (52%)	27
Auswahl einer Antwortoption mit bekannten Wörtern	LD-TW04	0	0	0	SR-TW04	0	4 (100%)	4
Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen	LD-TW05	4 (67%)	2 (33%)	6	SR-TW05	2 (33%)	4 (67%)	6
Auswahl einer anders aussehenden Antwortoption	LD-TW06	0	0	0	SR-TW06	0	2 (100%)	2
Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats	LD-TW07	2 (29%)	5 (71%)	7	SR-TW07	4 (36%)	7 (64%)	11
Total		39 (36%)	70 (64%)	109		49 (33%)	98 (67%)	147

¹¹⁷ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

¹¹⁸ Die Prozentzahlen wurden überall abgerundet.

Der obigen Tabelle kann entnommen werden, dass die *test-wiseness*-Strategien in der Gruppe der Off-Szenario-Schüler/innen in beiden Untersuchungsteilen etwa doppelt so oft wie in der Gruppe der Szenario-Schüler/innen codiert wurden (70 vs. 39 Codierungen im Lautdenken-Teil und 98 vs. 49 Codierungen beim *Stimulated-Recall*). Der grösste Unterschied zwischen den beiden Schüler/innen-Gruppen betrifft den Gebrauch der Strategie TW02 („Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“). Diese Strategie, die unter den *test-wiseness*-Strategien auch mit Abstand am häufigsten codiert wurde, konnte in den Lautdenkprotokollen der Off-Szenario-Schüler/innen mehr als doppelt so häufig wie bei der Szenario-Gruppe beobachtet werden (49 vs. 22 Codierungen); sie wurde auch in den *Stimulated-Recall*-Interviews beider Gruppen in einem ähnlichen Verhältnis codiert (48 vs. 24 Codierungen). Etwas häufiger griffen die Off-Szenario-Schüler/innen beim Lösen der Testaufgaben auch auf die *test-wiseness*-Strategie „Raten“ (TW01) zurück (6 Codierungen in den Lautdenkprotokollen vs. 2 Codierungen in der Szenario-Gruppe). In den *Stimulated-Recall*-Interviews wurde diese Strategie noch häufiger codiert, und zwar 19 Mal in der Off-Szenario-Gruppe und 6 Mal in der Szenario-Gruppe.

In Abbildung 25 werden alle sieben *test-wiseness*-Strategien dargestellt, die bei den einzelnen Schüler/innen codiert wurden. Die Zahlen in der Abbildung beziehen sich, wie auch in Abbildung 23 und Abbildung 24, auf die totale Anzahl der codierten Strategien; die Codierungen mit der Gewichtung „2“¹¹⁹ wurden hier also nicht abgezogen (vgl. Erklärung in Kap. 4.2.3).

¹¹⁹ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

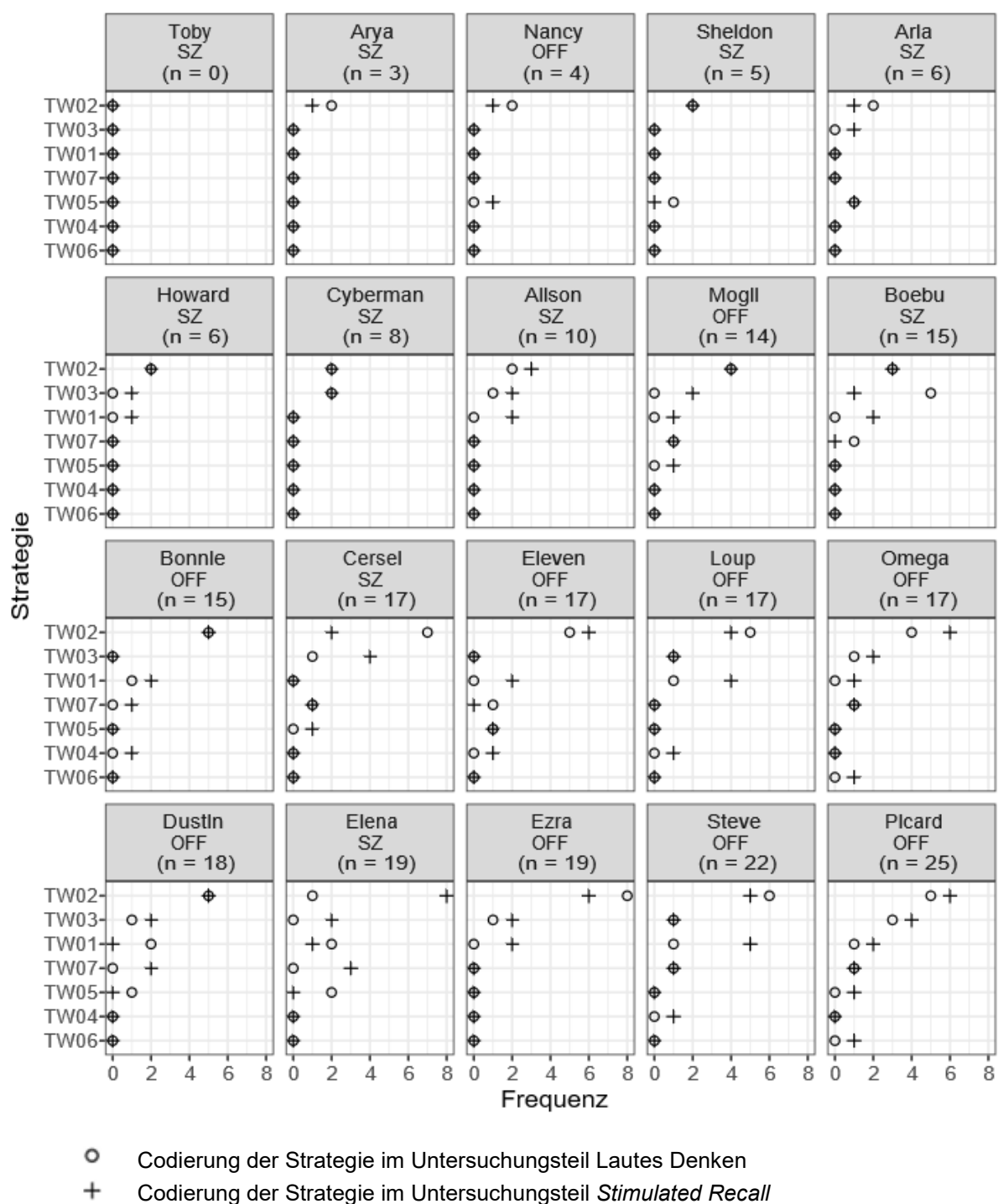


Abbildung 25: Alle codierten *test-wiseness*-Strategien pro Szenario- (SZ) und Off-Szenario-Schüler/innen (OFF).

Aus der obigen Darstellung geht hervor, dass die meisten Off-Szenario-Schüler/innen die *test-wiseness*-Strategien generell häufiger einsetzten, als dies bei den Szenario-Schüler/innen der Fall war. Bei den einzelnen Lernenden zeigen sich dabei interessante individuelle Unterschiede, die in beiden Gruppen wahrscheinlich auf die unterschiedlich ausgeprägten Französischkompetenzen der Schüler/innen zurückgeführt werden können: Bei stärkeren Schüler/innen wie Toby, der alle 12 Items korrekt löste, oder Arya mit 8 von 12 korrekten Antworten, wurden keine (Toby) bzw. wenige *test-wiseness*-Strategien codiert (Arya mit zwei

Codierungen der Strategie TW02 im Lauten Denken und einer Codierung der gleichen Strategie im *Stimulated-Recall*). Bei den zwei schwächsten Schüler/innen Steve und Picard, die jeweils nur 1 von 10 Testantworten richtig lösten, wurden hingegen die meisten *test-wiseness*-Strategien codiert (insgesamt 22 bei Steve und 25 bei Picard).

4.4.4 STRATEGIEN, DIE VON SZENARIO- UND OFF-SZENARIO SCHÜLER/INNEN BEI DER BEARBEITUNG VON ÄQUIVALENTEN TASKS EINGESETZT WURDEN

In den Kap. 4.4.2 und 4.4.3 wurde quantitative Evidenz zu den Unterschieden präsentiert, die bezüglich der Strategien in den Verbalprotokollen der Szenario- und Off-Szenario-Schüler/innen codiert wurden. Der Fokus lag dort also auf den beiden Schüler/innen-Gruppen. Im Zentrum dieses Kapitels stehen nun weiterhin Unterschiede zwischen dem Einsatz von Strategien beim Lösen der Szenario-Aufgaben und beim Lösen der Off-Szenario-Aufgaben. Es wird hier aber nicht auf die Schüler/innen, sondern auf die gleiche Aufgabe bzw. auf die vergleichbaren Aufgaben fokussiert, die von den beiden Gruppen gelöst wurde(n). Es wird hier also der Frage nachgegangen, welche Unterschiede sich zwischen der gleichen Aufgabe bzw. den vergleichbaren Szenario- und Off-Szenario-Aufgaben in Bezug auf die Strategien feststellen lassen, die bei ihrer Bearbeitung eingesetzt wurden. Der Vergleich kann vorgenommen werden, weil 1) die Hörverstehensaufgabe Task B, Szenario „Stadttrip“ von den Schüler/innen beider Gruppen gelöst wurde und 2) bei der Auswahl der sonstigen Off-Szenario-Aufgaben darauf geachtet wurde, dass die Handlungsziele, die Formate und die Gewichtung der getesteten Fertigkeiten aus den Off-Szenario-Aufgaben möglichst denen aus dem Szenario „Stadttrip“ entsprechen (vgl. Kap. 3.4.2). So hatten der Task A des Szenarios „Stadttrip“ und die Off-Szenario-Aufgabe „Walkies“ die gleiche Textsorte „Messenger-Chat“ als Inputtext, wenn auch mit unterschiedlichen Itemformaten. Task C im Szenario „Stadttrip“ und die Off-Szenario-Aufgabe „Berufswahl“ hatten ebenfalls die gleiche Textsorte „Beitrag aus einem Internet-Forum“ und darüber hinaus auch das gleiche Itemformat (vgl. Anhang 2 und 3). Der Vergleich dieser drei äquivalenten Aufgaben wird aufgrund von sieben Lerner- und Teststrategien gemacht, die bei den Szenario- und Off-Szenario-Aufgaben am häufigsten codiert wurden und in Abbildung 26 dargestellt werden.

Abbildung 26 zeigt die Gesamtsumme der Strategien auf, die beim Bearbeiten von beiden Bildschirmseiten pro jeweilige Aufgabe und Schüler/innen-Gruppe (Szenario- vs. Off-Szenario-Schüler/innen) codiert wurden, und zwar in der Reihenfolge ihrer Häufigkeit. Die dargestellten Zahlen beziehen sich aus technischen Gründen auf die totale Anzahl der

codierten Strategien, d.h. die Codierungen mit der Gewichtung „2“¹²⁰ wurden hier nicht abgezogen (vgl. Erklärung in Kap. 4.2.3). Beim Betrachten der Grafik ist darüber hinaus die unterschiedliche Darstellung der Häufigkeit von Strategien zu berücksichtigen, die pro Szenario- und Off-Szenario-Aufgaben codiert wurden: Auf der rechten Seite von Abbildung 26 geht die x-Achse nur bis 50, auf der linken Seite hingegen über 80.

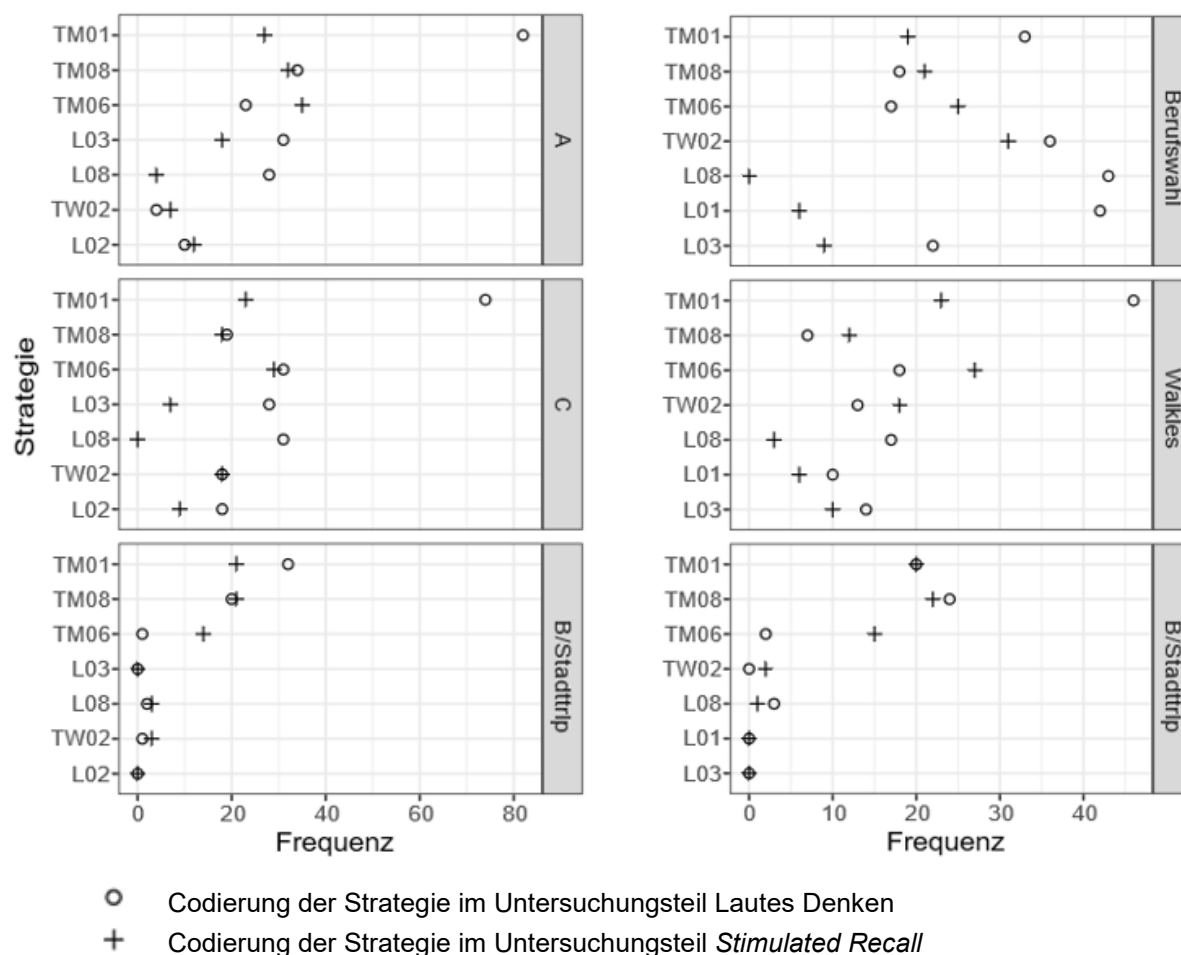


Abbildung 26: Sieben am häufigsten codierte Strategien pro Task (links: Szenario-Aufgaben, rechts: Off-Szenario-Aufgaben).

Aus Abbildung 26¹²¹ geht hervor, dass bei der Bearbeitung der vergleichbaren Szenario- und Off-Szenario-Aufgaben teilweise andere Strategien am häufigsten codiert wurden. Am meisten wurden in beiden Gruppen die Strategien TM01 („Instruktionen und/oder

¹²⁰ Wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde.

¹²¹ Der Grafik kann ebenfalls entnommen werden, dass sowohl die Szenario- wie auch die Off-Szenario-Schüler/innen beim Bearbeiten der unterschiedlichen Aufgaben aus ihrem jeweiligen Testbooklet (Task A, B und C im Fall der Szenario-Schüler/innen und Tasks „Berufswahl“, „Walkies“ und Task B (Szenario „Stadttrip“) im Fall der Off-Szenario-Schüler/innen) unterschiedlich gerne auf die sieben am häufigsten codierten Strategien zurückgriffen. Auf diese Unterschiede, die sowohl auf die getesteten Fertigkeiten und auf den anderen Fokus der Aufgaben zurückgeführt werden können, wird an dieser Stelle nicht weiter eingegangen.

Antwortoptionen (erneut) lesen“), TM08 („Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“ und TM06 („Fokus auf relevante Textteile“) beobachtet. Den vierten Platz belegt bei den Szenario-Aufgaben die Strategie L03 („Einen Teil des Inputtextes (erneut bzw. genauer) lesen“), die bei den Off-Szenario-Aufgaben zwar auch vorkommt, aber erst an der siebten Stelle. Danach kommt bei den Szenario-Aufgaben die Strategie L08 vor („Intensive Auseinandersetzung mit Teilen des Inputtextes“), die bei den Off-Szenario-Aufgaben ebenfalls den fünften Platz einnimmt. Die sechste Stelle belegt bei den Szenario-Aufgaben Strategie TW02 („Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“), die bei den Off-Szenario-Aufgaben am vierthäufigsten codiert wurde. Bei den Off-Szenario-Aufgaben steht hingegen an sechster Stelle Strategie L01 („Sorgfältiges Lesen“), die unter den sieben häufigsten Strategien bei den Szenario-Aufgaben nicht vorkommt. Den letzten Platz nimmt bei den Szenario-Aufgaben die Strategie L02 („Schnelles Lesen“) ein.

Task B, Szenario „Stadttrip“ wurde, wie gesagt, von beiden Schüler/innen-Gruppen gelöst. Ein Blick auf die Schwierigkeitswerte der beiden Items aus dieser Aufgabe zeigt, dass das eine Item mittelschwer und das andere schwer war (vgl. Kap. 4.1). Beim Betrachten der Codierungen zu Task B fällt auf, dass die Szenario-Schüler/innen beim Lauten Denken die Strategie TM01 etwas häufiger einsetzten als die Off-Szenario-Schüler/innen. Die Strategien TM08, TM06, L03, L08 und TW02 wurden hingegen in beiden Gruppen beim Lauten Denken vergleichbar oft codiert.

Der Vergleich der Codierungen in Task A, Szenario „Stadttrip“ und in der Off-Szenario-Aufgabe „Walkies“ im Lautdenken-Teil zeigt, dass die Szenario-Schüler/innen die Strategien TM01, TM08 und auch L03 deutlich häufiger anwandten. Die Off-Szenario-Schüler/innen griffen wiederum häufiger auf die *test-wiseness*-Strategie TW02 und auf die Lernerstrategie L08 zurück. Die Testmanagementstrategie TM06 wurde in beiden Gruppen vergleichbar oft codiert. Die Unterschiede in den Codierungen, die in beiden Aufgabengruppen vorgenommen wurden, können nicht nur mit dem unterschiedlichen Itemtyp, sondern auch mit der Anzahl und der Schwierigkeit der Items zusammenhängen, die in diesen Aufgaben vorkamen: Task A, Szenario „Stadttrip“ beinhaltete insgesamt vier Items, wovon zwei einfach und zwei mittelschwer waren. Die Off-Szenario-Aufgabe „Walkies“ beinhaltete hingegen zwei Items, die beide schwierig waren (vgl. Kap. 4.1).

Der Vergleich der Codierungen im Lautdenken-Teil in Task C, Szenario „Stadttrip“ und in der Off-Szenario-Aufgabe „Berufswahl“ zeigt, dass die Szenario-Schüler/innen etwa doppelt so oft wie die Off-Szenario-Schüler/innen von den Strategien TM01 und TM06 Gebrauch machten. Dafür wurden die Strategien TW02 und L08 häufiger von den Off-Szenario-

Schüler/innen und die Strategie L03 öfter von den Szenario-Schüler/innen eingesetzt. Strategie TM08 konnte wiederum bei beiden Gruppen etwa gleich oft beobachtet werden. Es ist auch hier nicht auszuschliessen, dass die Unterschiede in den Codierungen, vor allem die häufigere Codierung der *test-wiseness*-Strategie TW02 bei den Off-Szenario-Aufgaben, auf die unterschiedlichen Schwierigkeitswerte der Items aus diesen Aufgaben zurückzuführen sind. Unter den sechs Items, die die Off-Szenario-Schüler/innen bei der Aufgabe „Berufswahl“ lösten, befanden sich ein mittelschweres und fünf schwierige Items. Die Szenario-Schüler/innen lösten wiederum in Task C ein einfaches Item sowie zwei mittelschwere und zwei schwierige Items.¹²²

Resümee

Die Analyse der Verbalprotokolle zeigte, dass nur 1 von 10 Szenario-Schüler/innen die thematische Einbettung der Aufgaben bewusst bemerkte. Von 6 weiteren Schüler/innen wurde die Szenarioeinbettung angesprochen, nachdem sie gefragt wurden, ob sie etwas am Thema der Aufgaben bemerkten. Den 3 restlichen Schüler/innen wurde letztlich die direkte Frage gestellt, ob ihnen aufgefallen war, dass es in allen Aufgaben um die Planung eines Ausflugs nach Avignon ging, was sie dann auch bejahten. Die Einbettung der Aufgaben in ein übergreifendes Szenario fand die Mehrheit der befragten Szenario-Schüler/innen gut (7 von 10 Schüler/innen), den anderen Lernenden war es hingegen egal. Die Befragung der Off-Szenario-Schüler/innen zeigte wiederum, dass es niemanden aus dieser Gruppe störte, dass die von ihnen gelösten Aufgaben jeweils ein anderes Thema behandelten, und dass sie die Off-Szenario-Aufgaben abwechslungsreich fanden.

Der Vergleich der Codierungen von Lernerstrategien, die in der Szenario- und Off-Szenario-Gruppe vorgenommen wurden, zeigte, dass in beiden Gruppen im Prinzip gleich viele Strategien codiert wurden. Ein Blick auf die sieben am häufigsten codierten Lernerstrategien pro Schüler/in legt die Vermutung nahe, dass beim Einsatz dieser Strategien eher individuelle Vorlieben der Lernenden als die Einbettung oder Nicht-Einbettung der Aufgaben in ein Szenario eine Rolle spielten. Beim analogen Vergleich der codierten Testmanagementstrategien konnte festgestellt werden, dass die Szenario-Schüler/innen sowohl als Gruppe als auch individuell deutlich öfter auf diese Strategien zurückgriffen als die Off-Szenario-Schüler/innen. Bei den *test-wiseness*-Strategien war es gerade umgekehrt: Diese wurden von den Off-Szenario-Schüler/innen deutlich häufiger eingesetzt. Dabei zeigte sich auch, dass diese Strategien öfter bei den schwächeren Französischlernenden codiert wurden. Die Darstellung der sieben am häufigsten codierten Strategien pro Aufgabengruppe

¹²² Für ein Item wurde keine Itemschwierigkeit berechnet, weil das Item schlecht differenzierte, vgl. Kap. 4.1.

veranschaulichte darüber hinaus die Unterschiede, die beim Einsatz der Strategien bei der Bearbeitung der vergleichbaren Szenario- und Off-Szenario-Aufgaben festgestellt werden konnten. So zeigte sich, dass bei der Bearbeitung des gleichen Tasks bzw. der vergleichbaren Tasks, die von Szenario- und Off-Szenario-Schüler/innen gelöst wurden, teilweise andere Strategien am häufigsten codiert wurden.

Den in diesem Kapitel vorgestellten Ergebnissen ist zu entnehmen, dass der Einsatz der Strategien beim Bearbeiten der Szenario- und Off-Szenario-Aufgaben teilweise unterschiedlich ausfiel. Für die Erklärung dieser Unterschiede gibt es in den Daten zwar einige Indizien, auf die in Kap. 5 ausführlicher eingegangen wird. Es lässt sich aufgrund der vorliegenden Evidenz jedoch nicht eindeutig sagen, welche Faktoren für den unterschiedlichen Strategieneinsatz durch die Szenario- und Off-Szenario-Schüler/innen ausschlaggebend waren.

4.5 WEITERE AUSGEWÄHLTE ERGEBNISSE

In diesem Kapitel wird auf weitere Ergebnisse der Datenauswertung eingegangen, die sich zwar nicht direkt auf die Forschungsfragen beziehen, sich aber für das weitere Umfeld derselben dennoch als relevant erweisen. In Kap. 4.5.1 wird im Hinblick auf die Methodendiskussion über die Meinungen der Lernenden zum Lauten Denken beim Bearbeiten der Testaufgaben berichtet. Ein weiteres Kapitel (4.5.2) geht der Frage nach, wie die Schüler/innen die digitalen Textsorten empfanden, die in den IFB-Testaufgaben vorkamen. Zuletzt werden die aufschlussreichen und nebenbei gemachten Äusserungen der Schüler/innen zu den Sprachtestaufgaben rekapituliert, die nicht nur für Testentwickler/innen, sondern auch für Fremdsprachendidaktiker/innen und Fremdsprachenforschende von Interesse sein könnten (4.5.3).

4.5.1 EINSCHÄTZUNG DES LAUTEN DENKENS DURCH DIE SCHÜLER/INNEN

Die Lernenden wurden in den Gesprächen gefragt, ob sie das Laute Denken beim Lösen der Testaufgaben gestört hatte. Die Antworten der Lernenden (n=19)¹²³ wurden mithilfe der Kategorien „Lautes Denken hat (eher) gestört“ (n=6), „Lautes Denken hatte einen positiven Einfluss auf das Testverhalten“ (n=2, darunter der Schüler Steve, den das Laute Denken gleichzeitig störte) und „Lautes Denken war (eher) egal“ (n=12) codiert.

¹²³ Aus Versehen wurde diese Frage im Interview mit Schüler Cybermen nicht gestellt.

Die 6 Lernenden, die das Laute Denken eher störte, nannten dafür unterschiedliche Gründe. Loup erlebte es beispielsweise als ermüdend:

```
es ist SCHON noch anstrengend also (-) LAUT zu denken (-) es ist
ANstrengend wirklich (-) so mit sich SELBer zu reden. (Loup_Off,
Pos. 265-266)
```

Picard wies ihrerseits darauf hin, dass das Laute Denken für sie schwierig war, u.a. deswegen, weil sie darin keine Übung hatte. Es hatte sie auch etwas verwirrt, dass sie beim Lösen einer Aufgabe gleichzeitig mit der Aufgabe und mit dem Lauten Denken beschäftigt gewesen war:

```
es war komplizierter weil (--) wenn man (-) dENKt ohne etwas/ also ohne
es zu sAgen (--) dann kommt es einfach- aber wenn man es (.) erzÄhlen
muss oder (--) LAUt denken muss das ist halt UNGewohnt- aber würde man
das immer machen wäre es glaube ich nicht so kompliziert; [...] und das
ist halt ein bisschen (-) verwirrend manchmal wenn man noch LESeN und
HÖren muss- dann erzählen muss was man geDACHt hat und so- (Picard_Off,
Pos. 1175-1191)
```

Die anschliessende Frage der Forscherin, ob Picard das Gefühl hatte, durch das Laute Denken vielleicht im Test schlechter als sonst abgeschnitten zu haben, verneinte die Schülerin jedoch. Sie meinte aber, dass sie durch Lautes Denken etwas mehr Zeit für den Test brauchte.

```
MB: hast du das gefühl dass du jetzt [...] in diesem test vielleicht
schlechter abgeschnitten hast` wegen dem-
S: m: nEin aber (-) es hat ein bisschen länger gedauert (-)
wahrscheinlich-und es war ein bisschen em (-) herausfordernder- aber-
(na) ((S nickt)) (Picard_Off, Pos. 1192-1197)
```

Den Schüler Ezra hat das Laute Denken wiederum gestört bzw. gestresst, wenn er etwas nicht „begründen“ konnte:

```
MB: hat es dich geSTÖRt oder gestresst`
S: es ge:ht; manchmal manchmal (1.5) also manchmal wenn ich: ES
begründen konnte (-) war es EINFach aber manchmal (-) halt (einfach)
nicht; also- ja;
MB: auch wenn es nicht darum geht (-) dass du alles begründest?
S: ja.
MB: okay ((beide lachen)) (Ezra_Off, Pos. 398-405)
```

Auch Steve gab zu, dass ihn das Laute Denken gestört hatte und er es „komisch“ fand. Zugleich meinte er, dadurch die Aufgaben anders, und zwar gründlicher gelöst zu haben. Er empfand die Methode interessanterweise für ihn persönlich als „schlecht“, weil ungewöhnlich, meinte aber, dass sie besser „für die Lehrer“ sei, vielleicht weil er sich dadurch mehr Mühe gab:

```
MB: hast du das gefühl dass du das ANders gelöst hast als wenn du (--)
jetzt NIchT laut denken musstest?
S: ja.
MB: [...] was war anders?
S: ich hätte (--) nicht so viel überlegt- (.) ich hätte einfach gemacht;
```


MB: [...] war das jetzt also bEsser für dich` oder nICht (-) unbedingt-
 S: man kann es von <<lachend>beiden seiten> sehen; ((MB lacht)) für
 die lehrer ist es besser und für mich schlechter; also schlechter weil
 es (--) ungewöhnlich war- ja; (Steve_Off, Pos. 165-179)

Die positive Auswirkung des Lauten Denkens wurde auch von Böbu erwähnt: Das Aussprechen der Gedanken hatte ihm geholfen, sich mit den Inhalten der Inputtexte vertiefter auseinanderzusetzen.

[...] *es war noch gut also (-) so habe ich es wohl besser aufgenommen
 für mich- [...] weil [...] ich habe es gelesen (-) dann habe ich es nochmal
 wie wiederholt;* (Boebu_Szenario, Pos. 486-494)

Die Mehrheit der befragten Schüler/innen (n=12) war jedoch der Meinung, dass ihnen das Laute Denken eher egal war (auch wenn es für die meisten etwas Ungewöhnliches darstellte), d.h. es hatte weder einen positiven noch einen negativen Einfluss auf ihre Vorgehensweise bei der Bearbeitung der Testaufgaben. Omega sagte dazu beispielsweise:

am anfang musste man sich daran gewöhnen- aber nachher- (--) wa/ (-)
 ging es (unv.) so automatisch (Omega_Off, Pos. 298-299)

Auch andere Schüler/innen fanden das Laute Denken am Anfang etwas speziell. Sie waren aber der Meinung, dass es an ihrer Vorgehensweise nichts änderte, wie beispielsweise Dustin:

MB: okay (-) hat es dich geSTÖRT dass du [...] LAUT denken musstest?
 S: *ich fand es etwas speZIELL;* ((lacht))
 MB: GLAUBST du dass du deswegen ANders äh (-) vorgegangen bist oder
 wärst du genau so;
 S: *NEIN ich wäre glaub ich auch so irgendwie;* (Dustin_Off,
 Pos. 1057-1062)

4.5.2 EINSATZ VON DIGITALEN TEXTSORTEN IN EINEM SPRACHTEST¹²⁴

Damit die im Projekt IFB entwickelten, computerbasierten Lese- und Hörverstehensaufgaben authentischer erscheinen, wurden darin Inputtexte in Form von digitalen Textsorten eingesetzt, die im Alltag der Schüler/innen längst üblich sind (Chats, Forumsbeiträge oder Podcasts). In den *Stimulated-Recall*-Interviews wurden die Schüler/innen u.a. gefragt, wie sie diese digitalen Textsorten empfanden.

Obwohl die digitalen Textsorten und Testaufgaben bei den meisten Schüler/innen sehr gut ankamen, zeigte sich, dass sich die Begeisterung einiger von ihnen merklich in Grenzen hielt. Einzelne Aufgaben (vor allem Task D im Szenario „Stadttrip“ und Off-Szenario-Task „Schulfest“) wurden wegen unbekannter und als kompliziert wahrgenommener Formate

¹²⁴ In diesem Kapitel werden ausgewählte Resultate dargestellt, über die bereits in Barras et al. (2020) ausführlicher berichtet wurde.

kritisiert. Andere Schüler/innen wie Amy, Aria oder Sheldon fanden wiederum Task D im Szenario „Stadttrip“ deswegen wenig attraktiv, weil sie sich weder für Comics noch für Comicfestivals interessierten, über die in den Inputtexten zu dieser Aufgabe berichtet wurde.

Das realistische Design der Inputtexte wurde von vielen Schüler/innen gelobt. Sheldon fand aber ausgerechnet diese Eigenschaft der Testaufgaben „nervig“, denn

es hat ausgesehen als wäre [es] im internet aber es ist halt nicht so richtig (-) tEstArtig- (Sheldon_Szenario, Pos. 639-640)

Eine ähnliche Beobachtung wurde für die insgesamt gelobten Leseverstehenstexte in Form von Smartphone-Chats gemacht. Bonnie fand die Chats „eigentlich gut“, fügte aber auch hinzu, dass ihr dabei der vertraute Aufbau der „normalen“ Texte fehlte:

teilweise ist es (-) halt dann schwieriger zum verstehen weil [...] in einem normalen text hast du (-) geNAU so einen ablauf also (-) quasi hauptteil schlussteil und so (Bonnie_Off, Pos. 319-327)

Die meisten Lernenden reagierten auf die digitalen Textsorten jedoch sehr positiv. Mehrere Schüler/innen gaben an, dass ihnen die Textsorten vertraut sind und dass dies dazu führte, dass man sich während des Tests „wohler“ (Howard) fühlen konnte. Auch konnte man sich dadurch besser in die Situation reinversetzen, wie Picard betonte:

das ist etwas was man ja im leben sELbst auch hat und dann konnte man (-) sich damit ehm (-) vOrstellen; (Picard_Off, Pos. 894-895)

Einige Schüler/innen erwähnten darüber hinaus, dass die digitalen Textsorten und das attraktive, realistische Design ihre Motivation beeinflussen könnten. Arya fand es beispielsweise „echt cool“, dass das Leseverstehen in Form eines Gruppenchats dargestellt war, denn ihrer Meinung nach kommt es dann

einfach viel em (1.0) alltagstauglicher herüber als wenn einfach ein trockener text steht und es motiVIERT auch viel mehr wenn man [...] eine schöne darstellung [hat] und man kennt es aus dem ALLtag und dadurch verstehe ich es auch viel besser als wenn jetzt einfach nur ein text geschrieben wäre; (Arya_Szenario, Pos. 566-571)

An einer weiteren Stelle im Gespräch fügte sie noch hinzu:

ich habe mehr motivation wenn ich etwas sehe und ich finde DOCH vielleicht kann ich das noch geBRAUchen im späteren leben- (Arya_Szenario, Pos. 1374-1376)

Eine interessante Aussage wurde vom Schüler Omega gemacht. Als die Forscherin nachfragte, ob digitale Texte, die er soeben als „witzig“ und „cool“ bezeichnete, ihn in einer Testsituation etwas mehr motivieren würden, antwortete er:

[ja als/ ja] eh nein` es macht (-) also es ist zuerst eine motivation` und dann (irgendwie) <<:-)> kommt dir in den sinn dass es ein test ist`> ((lacht)) dann denkst du JA ist doch witzig aber jetzt muss ich mich wieder konzentrieren. (Omega_Off, Pos. 474-478)

Auch Bonnie war der Ansicht, dass der Inputtext nicht primär interessant, sondern vor allem verständlich sein muss, denn sonst

am schluss hat man so das gefühl ja (-) es ist zwar ein TEXT der mich vielleicht interessieren würde aber ich habe nichts verstanden. (Bonnie_Off, Pos. 1084-1086)

Die *Stimulated-Recall*-Interviews mit weiteren Schüler/innen deuten ebenfalls darauf hin, dass die positive Wirkung der digitalen Textsorten aufgrund der Testsituation eher eingeschränkt ist. So sagte z.B. Eleven über die Höraufgabe in Form eines Podcast:

es ist halt etwas ANders gemacht oder aber (-) es kommt auf das GLEICHE schlussendlich (Eleven_Off, Pos. 1255-1256)

Ezra seinerseits bemerkte gar nicht, dass eine der Höraufgaben in Form eines Podcast präsentiert wurde, während Steve nicht nur die Aufgabenform „egal“ war, sondern auch das Französische allgemein.

Die oben zusammengefassten Beobachtungen lassen vermuten, dass sich die Motivation – zumindest der hier untersuchten Lernenden – in einer Testsituation nur bedingt durch die Auswahl der Aufgabenstellung und der Texte erhöhen liess. Viel wichtiger schien es den Schüler/innen zu sein, dass sie die Testaufgaben bewältigen können, und zwar mit Erfolg.

4.5.3 SPRACHTESTAUFGABEN IN DEN AUGEN DER SCHÜLER/INNEN

Obwohl in den *Stimulated-Recall*-Interviews die Vorgehensweisen der Schüler/innen beim Lösen der IFB-Testaufgaben im Zentrum standen, machten manche Lernende beiläufige Bemerkungen, die darauf schliessen lassen, wie die ihnen sonst aus dem Unterricht bekannten Sprachtestaufgaben aussehen könnten. In diesem Kapitel werden ausgewählte Äusserungen der Schüler/innen vorgestellt, die im Hinblick auf die Entwicklung von standardisierten Sprachtestaufgaben, aber auch auf die Gestaltung der Beurteilungsanlässe im alltäglichen schulischen Fremdsprachenunterricht von Interesse sein könnten.

In den Interviews mit einigen Lernenden konnte festgestellt werden, dass sie überrascht waren, dass in den Inputtexten Elemente aller Antwortoptionen (d.h. auch der Distraktoren) erwähnt wurden. Arya berichtete z.B. im Interview sichtlich irritiert:

für mich war es ein wenig (-) also eigentlich !SEHR! verwirrend weil irgendwie (-) sie hat ALLES erwähnt was hier in diesen antworten stand; wirklich ALLES; (Arya_Szenario, Pos. 921-923)

Eleven bemerkte zu seiner grossen Überraschung erst im Interview, dass in der Aufgabe „Berufswahl“, Seite 2 eigentlich zwei Inputtexte Wörter beinhalten, die zum Item „selbst neue Produkte testen“ grundsätzlich passen könnten (wohingegen de facto jeweils nur ein Inputtext die richtige Antwort beinhaltete). Jetzt war er nicht mehr sicher, ob er im Test die richtige Antwort gewählt hatte, weil in seinen Augen

beide (-) könnten richtig sein. (Eleven_Off, Pos. 1093)

Auch Alison, die generell viel von den Inputtexten verstand und bei der zweimal die *test-wiseness*-Strategie „Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war“ beobachtet wurde, merkte erst im Interview erstaunt, dass sie den ganzen Inputtext hätte lesen sollen, um die Frage richtig zu beantworten.

S: man muss eigentlich fast den gAnzen text lesen (-) wenn man eine (-) frage beantworten will; ja. habe ich gemerkt; ((MB und S lachen)) weil/
MB: erst jetzt`
S: ja; (Alison_Szenario, Pos. 1180-1184)

Ähnlich berichtete Toby, dass er erst nach dem ersten Hörverstehensitem in Task B, Szenario „Stadttrip“ realisiert hatte, dass in den Inputtexten alle Informationen aus den drei Antwortoptionen in irgendeiner Form vorkommen würden. Deswegen hatte er während des Tests seine ursprüngliche Herangehensweise umgestellt und nicht mehr sofort die erste Antwortoption gewählt, zu der er im Inputtext ein passendes Wort hörte. Er hatte stattdessen genauer hingehört, um herauszufinden, was im Text gesagt würde:

S: ich habe (-) MICH darauf eingestellt dass ALLe drei vorkommen werden (--) und dass ich wirklich (-) äh mich achten muss (-) WAS genau sie meint. (-) [...]
MB: hm_hm (--) HEISST das dass du dann bei den früheren NICHT so eingestellt warst?
S: nein ich war eigentlich drauf eingestellt dass EInes vorkommt wie beim ersten und dann das EInTe gehört (-) GUT (-) das IST es (-) und dann (-) FERTig? (-) aber DA habe ich dann gemerkt ja das ist (.) IMMer ein bisschen (-) das ALLe drankommen dass man wirklich (-) ZUhören muss bis zum SCHLUSS? (-) (Toby_Szenario, Pos. 946-959)

Auf die Frage der Forscherin, ob das bei den schulischen Fremdsprachentests anders ist, antwortete Toby, dass die ihm bekannten Distraktoren normalerweise leichter von der richtigen Antwort zu unterscheiden sind:

S: also es kommt (-) EHer selten vor dass es so präzise? (-) also es ist MEISTens dass irgendetwas GROB anders ist oder wirklich (-)

KOMplett anders ist und nicht so (-) NAHEliegende sachen; (-)
(Toby_Szenario, Pos. 965-967)

Aus den Interviews ging ausserdem hervor, dass die Schüler/innen die IFB-Testaufgaben u.a. deswegen gut fanden, weil sie anders als die ihnen bisher bekannten Aufgaben waren. Teilweise war dies darauf zurückzuführen, dass die IFB-Aufgaben am Computer bearbeitet wurden. Die Äusserungen von Dustin und Howard zeigen aber, dass die Aufgaben nicht nur deswegen geschätzt wurden, sondern auch weil sie gut erklärt und interessant waren:

also [es war] ANders als wir es SONST jeweils haben (-) weil wir jetzt halt (-) alles am compUTer (--) und (--) es ist halt gut erklärt [...] (---) und auch (--) also es ist nicht einfach so LANGweilig also so TEXTe und so sondern auch dass es im INternet ist und (-) chat und so (-) ja das fand ich noch gut; (Dustin_Off, Pos. 1047-1056)

MB: ist dir etwas speZIELLes über diesen test aufgefallen?

S: so ein bisschen moDERNer und einfach (--) [...] es ist MEHR auch für den SCHÜLer gemacht als dann für den lehrer (-) weil der lehrer macht es ja meistens SELber und (-) der macht dann dass er es für SICH gut bewerten kann und alles (-) aber für die schüler halt ist es (-) der test dan halt manchmal LANGweilig oder eben NICHT interessant und das ist extrem schwierig oder (-) UNverständlich (-) und das ist eben dann halt (-) das ist VIEL besser. (--) es ist mehr so an den SCHÜLern angepasst [...]

MB: vom THEma her oder von der aufgabe;

S: ja (-) AUFBau (-) THema alles (-) ja. (Howard_Szenario, Pos. 1979-1992)

Die Interviews mit Böbu und Sheldon lassen wiederum vermuten, dass sich die IFB-Hörverstehensaufgaben deutlich von dem unterscheiden, was den beiden bisher aus dem Französischunterricht vertraut war. Böbu erwähnte, dass er das Multiple-Choice-Format der IFB-Hörverstehensaufgaben nicht kannte und dass das Hörverstehen allgemein selten im Unterricht trainiert wird:

MB: war das jetzt anders als das was ihr in der schule macht`

S: *ä:hm (.) ja also (--) ist SCHON anders; aber (3) ja (.) wir haben nie so mit ankreuzen eigentlich- (--) [eigentlich selten]*

MB: [okay und was macht ihr] dann` zum hören?

S: ö:h (---) zum hö/ (---) öhm: *also wir haben RECHT (.) selten also so hörverständnis im französisch;*

MB: <<erstaunt> hm;> (---) echt-

S: <<nickend> ja->

(Boebu_Szenario, Pos. 1257-1266)

Im Französischunterricht von Sheldon läuft hingegen das Hörverstehen gemäss seiner Schilderung so ab, dass die Lehrperson einen Tonträger laufen lässt und die Schüler/innen zuhören müssen, ohne dass sie dabei auch einen (visuellen) Kontext dazu haben. Deswegen fand er auch die visuelle Aufmachung von Task B im Szenario „Stadttrip“, die in Form eines Podcasts auf einer Internetseite präsentiert wurde, „nervig“ und „nicht so richtig testartig“:

S: ich fand (jetzt) dAs hier oben ein bisschen nErvig- [...] also es hat ausgesehen als wäre im internet aber es ist halt nicht so richtig (-) tEStArtig- [...]
 MB: was meinst du mit testARTig` (die tests sollen)/
 S: also die test/ jA: also Hörverständnistest haben wir halt NIE solche` sondern Eher einfach dass die LEhrerin (lässt/liest) etwas ab und dann hören wir sO:- (Sheldon_Szenario, Pos. 633-650)

Aus den oben dargestellten Aussagen der Schüler/innen kann geschlossen werden, dass sich die Sprachtestaufgaben, die diese Lernenden aus ihrem Fremdsprachenunterricht kennen, von den in den IFB-Aufgaben eingesetzten und auch sonst aus den standardisierten Sprachtests bekannten Aufgabenformaten teilweise deutlich unterscheiden. Auch scheinen die Aufgaben, die den Schüler/innen aus dem Unterricht bekannt sind, Themen zu behandeln, die für die Jugendlichen nicht besonders interessant sind oder die auf sie nicht wirklich motivierend wirken. Des Weiteren scheinen die Hörverstehensanlässe im Französischunterricht – zumindest gemäss Böbu und Sheldon – nicht oft vorzukommen bzw. sich von der Aufmachung her deutlich von den Hörverstehensaufgaben zu unterscheiden, die im Projekt IFB entwickelt wurden und in standardisierten, computerbasierten Sprachtests eingesetzt werden.

Resümee

Dieses Kapitel war weiteren ausgewählten Resultaten gewidmet, die das weitere Umfeld der Forschungsfragen beleuchten. Dazu gehörten die Meinungen und Einschätzungen der Schüler/innen 1) zum Lauten Denken als Datenerhebungsmethode, 2) zum Einsatz von digitalen Textsorten in einem Sprachtest und 3) zu den Sprachtests. Auf diese Erkenntnisse wird auch im nächsten Kapitel (Kap. 5) eingegangen, in dem alle Befunde aus dem vorliegenden Projekt nochmals zusammengefasst, interpretiert und diskutiert werden.

5 DISKUSSION

In diesem Kapitel werden zunächst die Ergebnisse der Untersuchung pro Forschungsfrage zusammengefasst und interpretiert (Kap. 5.1). Anschliessend wird auf die Limitationen der Studie und auf die Frage der Generalisierbarkeit der Resultate eingegangen (Kap. 5.2). In Kap. 5.3 wird die Relevanz der Studie diskutiert, und zwar im Hinblick auf die empirischen, theoretischen und methodologischen Erkenntnisse sowie auf die möglichen Implikationen der gewonnenen Erkenntnisse für den Fremdsprachenunterricht.

5.1 ZUSAMMENFASSUNG UND INTERPRETATION DER ERGEBNISSE

5.1.1 STRATEGIEN DER SCHÜLER/INNEN BEIM LÖSEN DER TESTAUFGABEN

Um die erste Forschungsfrage zu beantworten, wurde untersucht, welche Strategien bei den Schüler/innen beim Lösen der IFB-Aufgaben beobachtet werden konnten. Dafür wurden in den Lautdenkprotokollen und *Stimulated-Recall*-Interviews die beobachteten Lerner-, Testmanagement- und *test-wiseness*-Strategien codiert, ausgewertet und in Kap. 4.2 detailliert dargestellt.

Die Resultate der Datenauswertung zeigen, dass die Schüler/innen beim Bearbeiten der Testaufgaben eine breite Palette an Strategien einsetzten und dass es sich bei der überwiegenden Mehrheit der codierten Strategien – ähnlich wie in den Studien von A. D. Cohen und Upton (2006), L. Gu und So (2017) und Suvorov (2018) – um konstruktrelevante Lerner- und Testmanagementstrategien handelte. In den Daten konnten insgesamt 18 Lerner-, 19 Testmanagement- und 7 *test-wiseness*-Strategien ermittelt werden. Die codierten Lernerstrategien machen knapp 35% der in den Lautdenkprotokollen und ca. 24% der in den Interviews gefundenen Strategien aus; die Testmanagementstrategien ergeben etwa 57% der in Lautdenkprotokollen und ca. 60% aller im *Stimulated-Recall*-Teil codierten Strategien; und bei ca. 8% der in Lautdenkprotokollen und ca. 16% der in den *Stimulated-Recall*-Interviews beobachteten Strategien handelt es sich um konstruktirrelevante *test-wiseness*-Strategien.¹²⁵

¹²⁵ Die Prozentzahlen beziehen sich auf die Anzahl der codierten Strategien nach Abzug der Codierungen mit den Spezialgewichtungen „2“ (wiederholte Aussage, die bei der gerade besprochenen Bildschirmseite bereits codiert wurde) und „25“ (Codierungen der Strategie L10 („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“), die bei den mündlichen Zusammenfassungen der Inputtexte vorgenommen wurden).

Diese Resultate sind aus der Sicht der Testentwickler/innen als positiv zu bewerten, weil sie zeigen, dass sich die Lernenden beim Lösen der IFB-Testaufgaben vor allem darum bemühten, die französischen Inputtexte unter dem Einsatz von Lernerstrategien zu verstehen und mithilfe von Testmanagementstrategien sinnvolle Antworten auf die Testfragen zu geben. Ob der Einsatz dieser konstruktrelevanten Strategien aber auch dazu führte, dass die Schüler/innen richtige Antworten auf die Testfragen finden konnten, hing vor allem davon ab, *wie gut* sie die Inhalte der Inputtexte tatsächlich verstehen und im Zusammenhang mit den Testfragen interpretieren konnten. So halfen den Schüler/innen gewisse konstruktrelevante Strategien wie sorgfältiges Lesen (L01), Einsatz von Kognaten (L15) oder Anwendung von morphosyntaktischen Regeln (L13) wenig, wenn sie vom Inputtext nicht genug verstanden. Dies war vor allem bei den Lernenden mit eher schwach entwickelten Französischkenntnissen wie Steve oder Elena deutlich zu beobachten.

Diese Resultate stimmen mit den bisherigen Ergebnissen der Strategie- und Teststrategieforschung überein. Diese zeigen, dass der Einsatz einer bestimmten Strategie bzw. mehrerer Strategien nicht automatisch zum Erfolg führen muss, weil erfolgreiche und nicht erfolgreiche Lernende bzw. Testteilnehmende oft von den gleichen Strategien Gebrauch machen (vgl. u.a. Anderson, 1991; A. D. Cohen & Upton, 2006; Griffiths, 2018). Anderson brachte dies aufgrund der Ergebnisse seiner Studie zu Teststrategien beim Lösen von Leseverstehensaufgaben bereits vor drei Jahrzehnten folgendermassen auf den Punkt:

The most significant finding from these data suggests that there is no single set of processing strategies that significantly contributes to success [...]. Readers scoring high and those scoring low appear to be using the same kinds of strategies while reading and answering the comprehension questions on either measure. This seems to indicate that strategic reading is not only a matter of knowing what strategy to use, but also the reader must know how to use a strategy successfully and orchestrate its use with other strategies. [...] However, strategy use may also be a matter of vocabulary control and general background knowledge. Beginning level language learners may know what strategies to use but because of a lack of vocabulary, or other schema related information, they may not have a strong enough language foundation to build on (1991, S. 468–469).

In der vorliegenden Studie konnte in Übereinstimmung mit der soeben zitierten Vermutung von Anderson tatsächlich – und wie auch schon u.a. bei Gordon (1987; zit. in A. D. Cohen, 2014), Tian (2000, zit. in A. D. Cohen, 2012) oder Nikolov (2006) – beobachtet werden, dass die Sprachkenntnisse der schwächeren Französischlernenden für eine erfolgreiche Bewältigung der Aufgaben nicht genügten. Bei diesen Lernenden konnten wohl deswegen vor

allem diejenigen Lernerstrategien beobachtet werden, die sich auf Prozesse auf den unteren Verarbeitungsebenen wie z.B. Worterkennung beziehen (vgl. Kap. 2.1.1 und 2.1.2). In vielen Fällen reichten ihre Sprachkompetenzen (noch) nicht aus, um darauf aufbauend dank der Prozesse auf den höheren Verarbeitungsebenen (wie z.B. das Ziehen von richtigen Schlussfolgerungen aus dem gelesenen Text) zu einem adäquaten Verständnis der gelesenen oder gehörten Texte zu kommen. Dies ist keine überraschende Erkenntnis: Einschlägige Forschung zeigt, dass das Worterkennen zu den wichtigsten Prozessen auf den unteren Verarbeitungsebenen gehört (vgl. z.B. Grabe, 2009; Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Perfetti, 2007; Schmitt et al., 2017; van Zeeland & Schmitt, 2013; Vandergrift & Baker, 2018) und dass Versuche, die Defizite auf den unteren Verarbeitungsebenen mithilfe von *top-down*-Prozessen auszugleichen, oft zu einem falschen Textverständnis führen (vgl. u.a. Alderson et al., 2015; Buck, 2001; Grabe, 2009; Grotjahn, 2012). Flüssiges Lesen und Hören ist also erst dann möglich, wenn die Lernenden grosse Wortschatzmengen rapide und vor allem automatisch erkennen können, ohne dabei auf die Unterstützung der Informationen aus dem Kontext angewiesen zu sein. Dies war nur bei den wenigsten Schüler/innen der qualitativen Stichprobe der Fall (z.B. Toby, Alison), wie neben den Testergebnissen auch ihre mündlichen Zusammenfassungen der Inputtexte zeigten.

Der Einsatz von *test-wiseness*-Strategien konnte in der vorliegenden Studie vor allem dann beobachtet werden, wenn die Lernenden die Inputtexte nicht richtig oder nicht genügend verstanden, wie z.B. Steve beim Lösen der Aufgabe „Berufswahl“, Seite 2. Interessanterweise konnten aber die konstruktirrelevanten Strategien auch bei Schüler/innen festgestellt werden, die zwar allgemein über recht gute Französischkenntnisse verfügten, aber die Testaufgaben zu schnell und zu oberflächlich zu lösen versuchten, wie z.B. Alison beim Bearbeiten von Task C, Seite 2 im Szenario „Stadttrip“. Aufgrund dieser Erkenntnisse kann vermutet werden, dass es zumindest einen Versuch wert wäre, optimale Vorgehensweisen und nützliche Testmanagementstrategien beim Lösen von (standardisierten) Sprachtestaufgaben im Fremdsprachenunterricht zu thematisieren. Schüler/innen wie Alison könnten vermutlich stark davon profitieren, wenn man ihnen nahelegen würde, die Inputtexte nicht überwiegend oberflächlich zu scannen und darin nach den Wörtern aus den Antwortoptionen zu suchen, sondern die Texte – natürlich je nach der zu lösenden Aufgabe – auch gründlicher zur Kenntnis zu nehmen.

Die Tatsache, dass in der vorliegenden Studie vor allem Teststrategien (d.h. Testmanagementstrategien wie auch in kleinerem Ausmass *test-wiseness*-Strategien) und weniger oft Lernerstrategien beobachtet werden konnten, spricht dafür, dass die Schüler/innen beim Lösen der Testaufgaben vor allem an der erfolgreichen Lösung dieser Aufgaben interessiert waren und weniger daran, die Inputtexte inhaltlich wahrzunehmen oder

etwas daraus zu lernen. Zu ähnlichen Schlüssen kamen in ihren Studien auch schon A. D. Cohen und Upton (2006), Rossa (2012) und Suvorov (2018). Die vorliegenden Erkenntnisse bilden folglich die Strategien ab, von denen die Schüler/innen im Rahmen einer Testsituation Gebrauch machen, und es lässt sich aufgrund dieser Evidenz nicht sagen, ob die Lernenden bei der Rezeption der gleichen Inputtexte im Rahmen einer Lernsituation eventuell auf andere Strategien zurückgreifen würden. Dies ist ein generelles Problem, mit dem die Sprachtestforschung konfrontiert ist (vgl. Phakiti, 2007, S. 154; Reed, 2014, S. 891) und zu dem Alderson (2000, S. 27) vor bereits über zwei Jahrzehnten treffend anmerkte:

When we read 'normally', we are not being assessed. Thus, knowing that we are being assessed when reading creates a different event, and it is difficult to extrapolate from 'performance' in one event to 'performance' in the other. The riposte is partly empirical: let's see what difference it makes. But it is also partly philosophical: society needs assessment. Assessment is a different event from many other activities, but it is valued by society for its own sake, and is therefore justified. The ability to extrapolate from assessment to the real world is still important, but it is equally important not to confuse the assessed event with the 'real thing'.

Des Weiteren konnte auch in den vorliegenden Daten beobachtet werden, dass die Strategien selten alleine, sondern überwiegend in Clustern und Ketten vorkommen, was eine aus der Strategieforschung mittlerweile bestens bekannte und auch für die Analyse herausfordernde Tatsache ist (vgl. A. D. Cohen, 2011a; A. D. Cohen & Upton, 2006; Nikolov, 2006; Oxford, 2017). Wie auch schon bei A. D. Cohen und Upton (2006), Nikolov (2006) oder Suvorov (2018) zeigte sich weiterhin, dass dabei grosse individuelle Unterschiede zwischen den einzelnen Schüler/innen festgestellt werden können (vgl. Kap. 4.2.3).

In der vorliegenden Studie konnten letztlich dank der beobachteten Strategien der Schüler/innen (ähnlich wie in der Studie von Rossa, 2012) auch Hinweise auf Elemente der Items oder der Inputtexte gefunden werden, die überarbeitet werden müssten. Die Aufgaben mit diesen Elementen waren für den Gebrauch von *test-wiseness*-Strategien anfällig und/oder führten zur unnötigen und unerwünschten Verwirrung der Testteilnehmenden. Dazu gehörte z.B. der Inputtext von Kelly im Szenario „Stadttrip“, Task C, Seite 2, wo der gleiche Text ungünstigerweise eine richtige Antwort und einen Distraktor beinhaltete, oder Task B im Szenario „Stadttrip“, in dem in den Augen der Testteilnehmenden auf die Frage „Was hat Lou in Avignon besonders beeindruckt?“ eindeutig eine positive Sache folgen musste, die richtige Antwortoption aber „Der starke Regen in der Stadt“ lautete, was als negativ angesehen wurde. Dadurch zogen viele Schüler/innen diese korrekte Option gar nicht in Betracht. Diese

Testelemente müssten folglich bei der Testrevision angepasst und anschliessend nochmals erprobt werden.

5.1.2 ZUSAMMENHÄNGE ZWISCHEN KONSTRUKTIRRELEVANTEN STRATEGIEN UND (NICHT) ERFOLGREICHEN AUFGABENLÖSUNGEN

Die zweite Forschungsfrage ging den Zusammenhängen zwischen konstruktirrelevanten *test-wiseness*-Strategien und erfolgreichen bzw. nicht erfolgreichen Aufgabenlösungen nach. Um diese Forschungsfrage zu beantworten, wurden die Äusserungen der Schüler/innen zu jedem Item, bei dem eine *test-wiseness*-Strategie entweder im Lautdenkprotokoll oder im *Stimulated-Recall*-Interview beobachtet und codiert wurde, im Sinne der evaluativen qualitativen Inhaltsanalyse (Kuckartz, 2018) gründlich untersucht. Diese Vorgehensweise erwies sich als zielführend, denn sie ermöglichte das Ermitteln der Fälle, in denen die Lernenden 1) eine richtige Antwort aufgrund einer *test-wiseness*-Strategie wählten, 2) eine falsche Antwort aufgrund einer *test-wiseness*-Strategie wählten, oder auch 3) zwar beim Lösen eines Items von einer *test-wiseness*-Strategie Gebrauch machten, dann aber ihre im Endeffekt richtige Antwort aus einem anderen, konstruktrelevanten Grund wählten. Die in Kap. 4.3 dargestellten Ergebnisse zeigen, dass durch das gewählte Vorgehen ein nuanciertes Bild des tatsächlichen Einflusses der *test-wiseness*-Strategien auf die Erfolge bzw. Misserfolge beim Lösen der Testaufgaben erfasst werden konnte.

Den Resultaten ist zu entnehmen, dass *test-wiseness*-Strategien von fast allen Szenario-Schüler/innen (Ausnahme: Toby) und von allen Off-Szenario-Schüler/innen angewandt wurden, von den Letzteren jedoch deutlich häufiger. Dies könnte evtl. mit der in Kap. 4.1 dargestellten höheren Schwierigkeit der von den Off-Szenario-Schüler/innen gelösten Items zusammenhängen: Weil die Off-Szenario-Schüler/innen die Aufgaben nicht mithilfe der Lerner- und Testmanagementstrategien lösen konnten, griffen sie auf *test-wiseness*-Strategien zurück. Generell konnte wie bei Rossa (2012), Suvorov (2018), L. Gu und So (2017) beobachtet werden, dass von den *test-wiseness*-Strategien in der vorliegenden Studie vor allem die schwächeren Schüler/innen Gebrauch machten. Daher kann hier vermutet werden, dass die eingesetzten Aufgaben (vor allem im Off-Szenario-Setting) für gewisse Schüler/innen sehr schwierig waren und dass es ihnen nicht ausreichend gelang, die Französischkompetenzen derjenigen Lernenden genauer zu erfassen, deren Französischkenntnisse noch am schwächsten entwickelt waren. Dies war im Projekt IFB aber auch nicht beabsichtigt und hatte mit dem Konstrukt zu tun: Weil die im Projekt erstellten Aufgaben mit Blick auf die ÜGK den Bereich um A2.2/B1.1 abdecken sollten, ist es

naheliegend, dass die Schüler/innen, deren Französischkompetenz tiefer lag, diese Aufgaben nicht erfolgreich bewältigen konnten.

Die Schüler/innen, die ein Item mithilfe der *test-wiseness*-Strategien beantworteten, waren in den meisten Fällen nicht erfolgreich. Dies ist aus der Sicht der Testentwickler/innen erfreulich, weil es darauf hindeutet, dass die Items nicht einfach zu „überlisten“ waren und die Antwortoptionen mit den Distraktoren ihre Rolle wie erwartet erfüllten. Ähnliches konnte auch schon in der Studie von Rossa (2012) beobachtet werden.

Mittels der von den Schüler/innen erfolglos angewandten *test-wiseness*-Strategien konnte in der vorliegenden Untersuchung darüber hinaus entdeckt werden, dass das bereits im obigen Kapitel erwähnte Item „beeindruckt in Avignon“ (Task B, Szenario „Stadttrip“, Seite 2) überarbeitet werden sollte, weil das positiv konnotierte Wort „beeindrucken“ in den Augen der Schüler/innen unmöglich zur (richtigen) Antwortoption „Der starke Regen in der Stadt“ passte, was sie unabsichtlich in die Irre führte. Auch sollte der Hinweis, dass sich manchmal in Szenario „Stadttrip“, Task C und in der Off-Szenario-Aufgabe „Berufswahl“ die richtigen Antworten im gleichen Text befinden können (und dass folglich nicht alle drei Texte allen drei Fragen zugeordnet werden müssen), stärker hervorgehoben werden, um die Chance zu erhöhen, dass die Schüler/innen diesen zur Kenntnis nehmen.

Bei der Untersuchung der – im Vergleich weniger häufigen – Fälle, in denen ein Item mithilfe einer *test-wiseness*-Strategien richtig gelöst werden konnte, zeigte sich, dass dabei oft der Zufall eine Rolle spielte. Das ist bei geschlossenen Antwortformaten nicht zu vermeiden und auch nicht anders zu erwarten: Das einzige, was die Testentwickler/innen dagegen unternehmen können, wenn sie Aufgaben mit geschlossenen Antwortformaten erstellen (müssen), ist die Entwicklung von möglichst guten Distraktoren, ihre gründliche Erprobung und die Durchführung von psychometrischen Auswertungen (z.B. IRT) ¹²⁶, die solch zufällig richtig ausgewählte Antworten zu entdecken helfen. Dank der Untersuchung der *test-wiseness*-Strategien der Schüler/innen konnten in der vorliegenden Studie drei Items identifiziert werden, die überarbeitet werden müssten, weil man sie aufgrund bestimmter Eigenschaften der Inputtexte (Szenario „Stadttrip, Task C, Item „Stadtzentrum“, Seite 1 und Seite 2) oder auch u.a. aufgrund der grafischen Darstellung der Aufgabe aus falschen Gründen richtig lösen konnte (Off-Szenario-Aufgabe „Walkies“, Seite 1).

¹²⁶ Im Projekt IFB wurde die Itemanalyse mit einer 2-PL-Modellierung durchgeführt, wie schon in Kap. 4.1 näher ausgeführt wurde.

5.1.3 UNTERSCHIEDE BEIM EINSATZ DER STRATEGIEN BEIM LÖSEN DER SZENARIO- UND OFF-SZENARIO-AUFGABEN

Die dritte Forschungsfrage lautete, ob Unterschiede beim Einsatz der Strategien festgestellt werden können, wenn szenariobasierte und nicht szenariobasierte Testaufgaben gelöst werden, und wenn ja, worauf dies zurückgeführt werden könnte. Um eine Antwort auf diese Frage zu finden, wurden die vorhandenen Daten aus einem qualitativen (Kap. 4.4.1) und aus einem quantitativen Blickwinkel analysiert (Kap. 4.4.2 - 4.4.4).

In Kap. 4.4.1 wurden die Ergebnisse der Verbalprotokollanalyse dargestellt, die im Hinblick auf die Wahrnehmung der Szenarioeinbettung von Testaufgaben durch die Szenario-Schüler/innen untersucht wurden. Es zeigte sich, dass nur 1 von 10 Szenario-Schüler/innen den thematischen Zusammenhang der bearbeiteten Aufgaben bewusst wahrnahm. Sechs Szenario-Schüler/innen erwähnten die Szenarioeinbettung, nachdem sie gefragt wurden, ob ihnen am Thema der Aufgaben etwas aufgefallen sei, und drei weitere bejahten die ihnen letztlich direkt gestellte Frage, ob ihnen aufgefallen sei, dass sich alle Aufgaben um die Planung einer Schulreise nach Avignon drehten. Sieben Schüler/innen fanden die Einbettung der Aufgaben in ein übergreifendes Szenario gut, drei weiteren Lernenden war der thematische Zusammenhang der Aufgaben wiederum gleichgültig. Es kann vermutet werden, dass ein eventueller, positiver Einfluss der Szenarioeinbettung nur bei der einen Schüler/in, von der diese Einbettung bemerkt wurde, vorliegen könnte. Dies lässt sich aber nicht mit Sicherheit behaupten. Ausserdem kann nicht ausgeschlossen werden, dass die Szenarioeinbettung von den restlichen Schüler/innen unbewusst wahrgenommen wurde, was sich vielleicht doch in irgendeiner Form auf ihre Herangehensweise beim Lösen der Aufgaben ausgewirkt haben könnte.

Die Off-Szenario-Schüler/innen wurden wiederum im Interview gefragt, wie sie es empfanden, dass alle Aufgaben ein anderes Thema behandelten. Alle 10 Schüler/innen fanden es gut und abwechslungsreich, einige von ihnen vermuteten sogar, dass das Lösen von mehreren Aufgaben zum gleichen Thema ungünstig sein könnte (was aber von keiner und keinem der Szenario-Schüler/innen behauptet wurde).

Die Off-Szenario-Schüler/innen vermissten also die Szenario-Einbettung nicht; vermutlich auch deswegen, weil sie diese Art der Testdarbietung bisher nicht kannten. Den thematischen Zusammenhang der Aufgaben nahmen ausserdem fast alle Szenario-Schüler/innen kaum bewusst wahr. An dieser Stelle könnte also die berechtigte Frage auftauchen, ob es sich überhaupt lohnt, bei der Konstruktion von Sprachtestaufgaben den erheblichen Mehraufwand zu leisten, der mit der Entwicklung solcher Aufgaben verbunden ist. Diese Frage lässt sich

aufgrund der vorliegenden Evidenz nicht eindeutig beantworten, da in dieser Studie nur ein einziges Szenario vertiefter untersucht wurde, in dem die Einbettung der Aufgaben vielleicht auch nicht stark genug zum Vorschein kam. Es müssten weitere Untersuchungen mit verschiedenen Szenarien durchgeführt werden, um eine klarere Antwort auf diese Frage zu finden.

Um weitere Hinweise zum eventuellen Einfluss der Szenarioeinbettung auf die Vorgehensweisen der Lernenden beim Testlösen zu sammeln, wurde in Kap. 4.4.2 die Anzahl der Codierungen der Lernerstrategien der Szenario- und Off-Szenario-Schüler/innen und in Kap. 4.4.3 die Anzahl der codierten Teststrategien (d.h. Testmanagement- und *test-wiseness*-Strategien) pro Schüler/innen-Gruppe miteinander verglichen. In Kap. 4.4.4 wurden zudem die sieben bei den Szenario- und Off-Szenario-Aufgaben am häufigsten codierten Strategien im Vergleich dargestellt.

In Bezug auf die Lernerstrategien zeigte sich, dass in der Szenario- und Off-Szenario-Gruppe grundsätzlich ähnlich viele Strategien codiert wurden. Der Vergleich der sieben am häufigsten codierten Lernerstrategien pro einzelner Schülerin bzw. pro individuellem Schüler lässt jedoch vermuten, dass der Einsatz von Lernerstrategien wahrscheinlich eher auf individuelle Präferenzen der Schüler/innen und womöglich auch auf ihre Französischkompetenz zurückgeführt werden kann, als auf ihre Zugehörigkeit zur Szenario- oder Off-Szenario-Gruppe.

Bei den Testmanagementstrategien konnte beobachtet werden, dass die Szenario-Schüler/innen diese Strategien sowohl als Gruppe als auch individuell deutlich häufiger als die Off-Szenario-Schüler/innen einsetzten. Dies lässt vermuten, dass der Einsatz der Testmanagementstrategien mit den unterschiedlichen und auch unterschiedlich schwierigen Aufgaben zu tun haben könnte, die diese beiden Schüler/innen-Gruppen lösten. Es lässt sich jedoch aufgrund der vorliegenden Daten nicht mit Sicherheit sagen, welche konkreten Eigenschaften der Aufgaben sich auf diese Unterschiede auswirkten. Wahrscheinlich spielten dabei auch die individuellen Unterschiede zwischen den Schüler/innen eine Rolle.

Im Fall der *test-wiseness*-Strategien konnte beobachtet werden, dass die Off-Szenario-Schüler/innen diese Strategien etwa doppelt so häufig wie die Szenario-Schüler/innen einsetzten, was sich generell auch bei der Betrachtung des individuellen Einsatzes dieser Strategien pro Schüler/in beobachten liess. Es zeigte sich, dass der Einsatz der *test-wiseness*-Strategien mit den unterschiedlich ausgeprägten Französischkompetenzen der Lernenden zusammenhängen könnte: Stärkere Schüler/innen setzten diese Strategien generell seltener

ein als schwächere Lernende. Diese Beobachtung wurde auch schon in den Studien von Rossa (2012), L. Gu und So (2017) und Suvorov (2018) festgehalten.

Aus dem Vergleich der jeweils sieben am häufigsten codierten Strategien pro Aufgabengruppe ging hervor, dass bei der Bearbeitung der vergleichbaren Szenario- und Off-Szenario-Aufgaben z.T. andere Strategien am häufigsten eingesetzt wurden. Bei dem von beiden Schüler/innen-Gruppen gelösten Task B, Szenario „Stadttrip“ zeigte sich, dass die Mehrheit der Strategien vergleichbar oft in beiden Gruppen codiert wurden. Eine Ausnahme bildete die Strategie TM01 („Instruktionen und/oder Antwortoptionen (erneut) lesen“), die häufiger von den Szenario-Schüler/innen eingesetzt wurde. Anders sah dies beim Vergleich von 1) Task C, Szenario „Stadttrip“ und der Off-Szenario-Aufgabe „Berufswahl“ sowie 2) Task A, Szenario „Stadttrip“ und der Off-Szenario-Aufgabe „Walkies“ mit der gleichen Textsorte beim Inputtext aus: Dabei konnte beobachtet werden, dass die Anzahl der codierten Strategien pro Aufgabenpaar teilweise erheblich voneinander abwich. Es ist daher zu vermuten, dass gewisse Strategien durch bestimmte Eigenschaften der Testaufgaben (wie z.B. die von ihnen intendierten *types of reading*) hervorgerufen wurden – eine Erkenntnis, über die bereits auch Nikolov (2006), A. D. Cohen & Upton (2006) und Suvorov (2018) berichtet hatten.

Die hier zusammengefassten Ergebnisse zeigen, dass es in der Tat Unterschiede beim Einsatz der Strategien gab, die beim Bearbeiten der Szenario- und Off-Szenario-Aufgaben beobachtet und codiert werden konnten. Für die Ursachen dieser Unterschiede sind in den Daten einige Indizien vorhanden. Worauf der unterschiedliche Einsatz der Strategien durch Szenario- und Off-Szenario-Schüler/innen jedoch letztlich zurückzuführen ist, kann aufgrund der vorliegenden Evidenz nicht mit Sicherheit konstatiert werden. Es könnte sein, dass sich die thematische Einbettung der Szenario-Aufgaben bzw. ihr Fehlen bei den Off-Szenario-Aufgaben auf die unterschiedlichen Herangehensweisen zumindest einiger Lernenden beim Lösen der Testaufgaben auswirkte. Diese Unterschiede könnten aber auch damit zusammenhängen, dass die Aufgaben, die die Szenario- und die Off-Szenario-Schüler/innen lösten, unterschiedliche Itemschwierigkeitswerte zugunsten der (relativ einfacheren) Szenario-Aufgaben aufwiesen. Auch waren die Itemtypen und die Anzahl Items bei den Szenario- und Off-Szenario-Aufgaben nicht gleich. Schliesslich wurden diese Unterschiede bestimmt auch durch die unterschiedlichen Vorlieben für gewisse Strategien wie auch durch die unterschiedlich ausgeprägten Französischkompetenzen der jeweiligen Schüler/innen mitbestimmt. Es ist also nicht ausgeschlossen, dass die hier ermittelten Unterschiede beim Einsatz der Strategien durch das Zusammenspiel mehrerer oder aller hier erwähnten Faktoren erklärt werden können.

5.2 LIMITATIONEN DER STUDIE UND GENERALISIERBARKEIT DER RESULTATE

Auf diejenigen Limitationen der Studie, die mit dem gewählten Datenerhebungs- und -auswertungsverfahren zusammenhängen, wurde bereits ausführlich in Kap. 3.7.4 eingegangen, weswegen sie an dieser Stelle nur kurz zusammengefasst werden. Die methodischen Erkenntnisse, die in dieser Studie durch den Einsatz des Lauten Denkens und der *Stimulated-Recall*-Interviews gewonnen wurden, werden darüber hinaus noch vertiefter in Kap. 5.3.2 diskutiert.

Zu den Einschränkungen der Studie gehört, wie üblich bei qualitativen Studien, die kleine Zahl der Studienteilnehmenden (zu denen hier allerdings umfangreiche und vielfältige Daten vorliegen). Unter anderem aus diesem Grund wird hier die Generalisierbarkeit der Resultate nicht anvisiert, was jedoch in qualitativen Studien ohnehin nicht angestrebt wird (Caspari, 2016, S. 18). Dadurch aber, dass hier der Vergleich der Resultate der Schüler/innen aus der qualitativen Stichprobe mit den Resultaten der Schüler/innen aus der quantitativen IFB-Stichprobe für Französisch möglich war (vgl. Kap. 3.4.3.2), kann vorsichtig vermutet werden, dass die Erkenntnisse aus der qualitativen Studie zumindest die Tendenzen in der IFB-Französischstichprobe abbilden könnten.

Darüber hinaus konnte in dieser Studie nur eine kleine Auswahl an Aufgaben aus dem Projekt IFB vertieft untersucht werden. Es stellte sich dabei erst nach der Datenerhebung heraus, dass die für die qualitative Studie ausgewählten Szenario- und Off-Szenario-Aufgaben unterschiedliche Schwierigkeitswerte zugunsten der Szenario-Aufgaben aufwiesen. Ferner gab es auch Unterschiede in den Aufgabenformaten, die in den beiden Schüler/innen-Gruppen zum Einsatz kamen, obwohl bei der Konzipierung der Studie auf eine möglichst grosse Vergleichbarkeit der ausgewählten Aufgaben geachtet wurde.

Weitere Limitationen der Studie hängen mit den eingesetzten introspektiven Methoden zusammen: Bei der Interpretation der Lautdenkdaten ist zu beachten, dass sich diese Methode nicht zur Erfassung von automatisch ablaufenden Prozessen eignet und dass „prinzipiell mehr gedacht wird als simultan ausgesprochen werden kann und nur die Gedankeninhalte mit der höchsten Salienz verbalisiert werden“ (Heine, 2014, S. 127). Dies führt dazu, dass die mit dieser Methode erhobenen Daten bezüglich der Gesamtheit der ablaufenden kognitiven Prozesse immer unvollständig sind (ebd.). Ausserdem fühlten sich trotz eines vorherigen, standardisierten Trainings nicht alle Forschungsteilnehmende beim Einsatz dieser Methode gleich wohl, sodass einige von ihnen deutlich weniger verbalisierten als andere. Beim Sichten der *Stimulated-Recall*-Daten muss wiederum mitbedacht werden,

dass sie erst nach der beforschten Tätigkeit erhoben wurden und bei den Verbalisierungen auf die Erinnerungen aus dem Langzeitgedächtnis zurückgegriffen wurde. Dies hatte zur Folge, dass sich die Forschungsteilnehmenden u.U. nicht an alle Gedanken erinnern konnten, die ihnen beim Lösen der Testaufgaben durch den Kopf gegangen waren. Ausserdem könnten sich ihre Erinnerungen mit den Einstellungen und nachträglichen Interpretationen des Geschehens vermischt haben (vgl. u.a. Heine & Schramm, 2016). In der vorliegenden Studie wurde deshalb versucht, mithilfe der Triangulation der beiden Methoden die Nachteile der einen Methode durch die andere Methode mindestens teilweise auszugleichen. Bei der Interpretation der *Stimulated-Recall*-Interviews ist weiter auch zu bedenken, dass der Verlauf der Interviews nicht nur durch die Proband/innen, sondern auch zu einem nicht unerheblichen Teil durch die Forschenden mitbestimmt wird. Dadurch, dass alle 30 Datensitzungen nach einem möglichst ähnlichen Ablaufschema und von der gleichen Forscherin geleitet wurden, wurde hier eine möglichst grosse Vergleichbarkeit der erhobenen Daten angestrebt. Da die Datenerhebung recht lange dauerte und auch intensiv war, kann hier allerdings nicht ausgeschlossen werden, dass die Müdigkeit der Forschungsteilnehmenden gegen Ende der Datenerhebung einen Einfluss auf die Qualität der erhobenen Daten gehabt hat. Trotz all dieser Punkte konnten in dieser Studie dennoch viele wertvollen Einblicke in die Testlösungsprozesse und Überlegungen der Schüler/innen gewonnen werden.

Wie die Datenerhebung wurde auch die Codierung und Analyse der erhobenen Daten von der gleichen Forscherin vorgenommen. Dies erwies sich in der vorliegenden Studie als Vorteil, weil die Forscherin die Daten schon zu Beginn der Datenauswertung bereits bestens kannte (vgl. dazu auch Y. Gu, 2014). Auch wenn im Sinne der Intracoder-Reliabilität mindestens sechs systematische Codiervorgänge vorgenommen wurden, ist hier natürlich nicht auszuschliessen, dass eine andere Forscherin oder ein anderer Forscher bei der Analyse der gleichen Daten evtl. zu abweichenden Resultaten kommen würde. Alle codierten und uncodierten Transkripte (auch in Form der MAXQDA-Projekte (VERBI Software, 2019)) sowie der Kategorienleitfaden werden deshalb interessierten Forschenden für weitere Analysen im KFM-Forschungsdatenarchiv ¹²⁷ zur Verfügung gestellt. Ausserdem wurde die Vorgehensweise bei der Analyse in der vorliegenden Dissertation transparent und möglichst detailliert dokumentiert. Es wäre interessant zu erfahren, ob eine Zweitanalyse der vorliegenden Daten durch eine andere Person zu ähnlichen Ergebnissen führen würde.

Obwohl die vorliegende Studie qualitativ ausgerichtet war, wurde bei der Darstellung der Ergebnisse auch über die Anzahl der Codierungen pro Strategie, Schüler/in bzw.

¹²⁷ <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000012367>

Schüler/innen-Gruppe und Aufgabentyp berichtet. Ziel war es, 1) dem Postulat von Seale (1999, S. 121) zu entsprechen, der dafür plädierte, auch in der qualitativen Forschung das Zählbare zu zählen, sowie 2) alle in den Daten vorhandenen Informationen darzustellen, wozu auch die Angabe zur Häufigkeit des Auftretens der codierten Segmente gehört (vgl. Kuckartz, 2018, S. 116). Es sollte dabei jedoch beachtet werden, dass in den Daten nur das codiert wurde, was die Lernenden verbalisierten und was von der (einen) Codiererin beobachtet wurde.

Im Zusammenhang mit der Erforschung von Teststrategien wird oft (selbst)kritisch angemerkt, dass die beforschte Vorgehensweise der Testteilnehmenden beim Lösen der Testaufgaben für Forschungszwecke eventuell nicht dem entspricht, was die gleichen Testteilnehmenden beim Lösen von Aufgaben in einem „richtigen“ Test (d.h. im realen Leben) tun würden. Als Grund dafür wird genannt, dass die Beteiligung an einem Forschungsprojekt für das weitere Leben der Testteilnehmenden keine Konsequenzen hat. Bei einem „richtigen“ Test kann hingegen für die Testteilnehmenden – natürlich je nach Test – viel mehr auf dem Spiel stehen, was sich auf ihre Vorgehensweise beim Test auswirken kann (vgl. u.a. A. D. Cohen, 2011b, S. 306; Reed, 2014, S. 891; Suvorov, 2018, S. 18). Bei der vorliegenden Studie trifft diese Einschränkung nicht zu: Sowohl die Teilnahme an der qualitativen und quantitativen Untersuchung der Testaufgaben im Rahmen des Projekts IFB, wie auch die Teilnahme am Bildungsmonitoring ÜGK, in das die Erkenntnisse aus dem Projekt IFB in Zukunft einfließen sollten, hatte bzw. hätte auf das Leben, die Schulnoten oder die Schulkarriere der beteiligten Schüler/innen keinen Einfluss. Die einzigen Unterschiede zwischen der Teilnahme am Bildungsmonitoring und an der qualitativen Datenerhebung waren die Anwesenheit der Forscherin bei der letzten Situation, sowie die Tatsache, dass die Schüler/innen beim Lösen der Aufgaben ihre Gedanken verbalisieren mussten. Darauf, wie sich dieses doch recht ungewöhnliche Setting auf die Vorgehensweise der Lernenden auswirkte bzw. auswirken konnte, wird in Kap. 5.3.2 eingegangen.

Zum Schluss muss auch angemerkt werden, dass die Definition dessen, welche Strategie in einem bestimmten Projekt als konstruktrelevante Testmanagementstrategie oder als konstruktirrelevante *test-wiseness*-Strategie definiert wird, davon abhängig sein muss, welches Konstrukt dem untersuchten Test zugrunde liegt. Dadurch wird die direkte Vergleichbarkeit der Studien zum Einsatz von Strategien beim Lösen von Testaufgaben generell eingeschränkt. Hinzu kommt, dass in all diesen Studien jeweils andere Testaufgaben von einem jeweils anderen Zielpublikum gelöst werden: In der Studie von L. Gu und So (2017) waren es z.B. junge Englischlernende, in der vorliegenden Studie jugendliche

Französischlernende und in den Studien von Suvorov (2018) sowie A. D. Cohen und Upton (2006) erwachsene Englischlernende.

5.3 RELEVANZ DER STUDIE

5.3.1 EMPIRISCHE UND THEORETISCHE ERKENNTNISSE

Dieser Teil ist den wichtigsten Punkten gewidmet, die als Beitrag der vorliegenden Arbeit zur Diskussion der empirischen und theoretischen Erkenntnisse im Bereich der Sprachtestforschung angesehen werden können. Dazu gehören die im Projekt gewonnenen Einsichten 1) zum Einsatz szenariobasierter Testaufgaben für *large scale assessments* in Fremdsprachen (Kap. 5.3.1.1), 2) zur Rolle der qualitativen Evidenz im Prozess der Testvalidierung (Kap. 5.3.1.2), und 3) zur Rolle der strategischen Kompetenz als Teil des Testkonstrukts (Kap. 5.3.1.3).

5.3.1.1 SZENARIOBASIERTE TESTAUFGABEN FÜR *LARGE SCALE ASSESSMENTS* IN FREMDSPRACHEN

In dieser Studie wurden szenariobasierte Testaufgaben beforscht, die im Projekt IFB erstellt und dort überwiegend mit quantitativen Methoden untersucht wurden. Gemäss dem Wissenstand des IFB-Teams zu Beginn des Projekts handelte es sich dabei um einen Pioniersversuch, den von ETS für die Überprüfung von Lesekompetenzen in der Schulsprache Englisch entwickelten szenariobasierten Ansatz (vgl. ETS Research, 2019; O'Reilly & Sabatini, 2013; Sabatini et al., 2014b) für die Überprüfung von fremdsprachlichen Sprachkompetenzen umzusetzen. Die vorliegende Untersuchung zeigte, dass sich der Einfluss der Szenarioeinbettung auf die Vorgehensweisen der Schüler/innen, die diese Testaufgaben lösten, in Grenzen hielt. Es ist dabei nicht auszuschliessen, dass die Szenarioeinbettung der Aufgabe in dem für die Untersuchung ausgewählten Szenario „Stadttrip“ nicht stark genug zum Vorschein kam und dass die Untersuchung eines anderen Szenarios zu anderen Resultaten führen würde.

Dessen ungeachtet konnte hier festgestellt werden, dass sowohl die Szenario- wie auch die Off-Szenario-Schüler/innen die Aufmachung der Aufgaben, die digitalen Textsorten sowie die Themen der Inputtexte im Allgemeinen spannend und interessant fanden (obwohl dabei natürlich auch der Neuigkeitseffekt der computerbasierten Tasks und die spezielle Situation der Datenerhebung eine Rolle spielen könnten). Auch hatten die Lernenden mit der Bedienung der Testaufgaben am Computer keine grösseren Probleme, obwohl sich auch

zeigte, dass die Instruktionen zu bestimmten, den Schüler/innen bisher unbekannten Itemformaten noch expliziter bzw. einfacher formuliert werden sollten.

Es konnte ausserdem festgestellt werden, dass die Französischaufgaben vor allem für die schwächeren Schüler/innen oft (sehr) schwierig waren. Es wäre daher überlegenswert, in zukünftigen Leistungsmessungen im Rahmen der ÜGK zumindest für Französisch als Fremdsprache auch einfachere Aufgaben einzusetzen, um nicht nur das Erreichen der Grundkompetenzen der Schüler/innen zu überprüfen, sondern auch das gesamte Spektrum ihrer Kompetenzen besser abbilden zu können. Es könnte sich nämlich zeigen, dass es angebracht wäre, die in der Schweiz festgelegten Grundkompetenzen für Französisch evtl. nach unten anzupassen oder auch eine höhere Stundendotation für den Französischunterricht vorzusehen, sollte an den bestehenden Grundkompetenzen festgehalten werden. Die letztere Lösung legen auch die Ergebnisse des in der Zentralschweiz durchgeführten Projekts „Fremdsprachenevaluation BKZ“ nahe, die zeigen, dass die Schüler/innen mit einer höheren Stundendotation im Französisch bessere Testergebnisse als die Schüler/innen mit niedrigerer Anzahl Jahreswochenlektionen in Französisch erzielten (Peyer et al., 2016).

Darüber hinaus zeigte sich in dieser Untersuchung, wie schon in den Studien von A. D. Cohen und Upton (2006) und Suvorov (2018), dass die Verwendung von bestimmten Strategien durch das Testdesign evoziert werden kann, die entweder vom Testkonstrukt vorgesehen wurden oder auch diesem zuwiderlaufen können. In der vorliegenden Studie wurde beispielsweise der Einsatz der konstruktrelevanten Testmanagementstrategie „Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war“ oft und wie gewünscht in den Items beobachtet, die *search reading* auf lokaler Ebene verlangten. Es zeigte sich aber auch, dass gewisse Items unerwünschte *test-wiseness*-Strategien hervorbrachten, was natürlich vom Testkonstrukt nicht intendiert war (z.B. wurde die *test-wiseness*-Strategie „Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen“ oft beim Item „beeindruckt in Avignon“ (Szenario „Stadttrip“, Task B, Seite 2) codiert). Es kann hier daher mit Nachdruck Suvorov (2018, S. 17) zugestimmt werden, der zu dem Schluss kommt, dass von Testentwickler/innen sichergestellt werden sollte, dass die durch das Testdesign ausgelösten Strategien auch tatsächlich für das Konstrukt relevant sind.

5.3.1.2 QUALITATIVE EVIDENZ IM PROZESS DER TESTVALIDIERUNG

Trotz der in Kap. 5.2 genannten Limitationen veranschaulicht die vorliegende, explorativ-interpretativ ausgerichtete Studie, wie wichtig und wertvoll es ist, neu erstellte (Test-)Aufgaben nicht nur mit quantitativen, sondern auch mit qualitativen Methoden zu untersuchen (vgl. u.a.

Ercikan & Pellegrino, 2017a; Leung, 2013; Lumley & Brown, 2005). Die Erforschung der Strategien von Testteilnehmenden kann Bestrebungen zur Testvalidierung eindeutig bereichern: Viele Einsichten in die Lösungsprozesse der IFB-Testaufgaben (z.B. konkrete Informationen zu den Items, die überarbeitet werden sollten, weil sie den Einsatz von *test-wiseness*-Strategien begünstigen), die hier mit introspektiven Methoden gewonnen wurden, wären unentdeckt geblieben, wenn nur die psychometrische Evidenz aus der quantitativen Studie des Projekts IFB betrachtet worden wäre. Allgemein kann behauptet werden, dass sowohl die Erkenntnisse aus der vorliegenden Studie wie auch aus dem Projekt IFB insgesamt (Karges, Lenz et al., 2021) zukünftigen Testentwicklungsprojekten für Leistungsmessungen zugutekommen können.

Es ist davon auszugehen, dass es qualitativ ausgerichtete Untersuchungen wie die vorliegende auch in Zukunft immer geben und brauchen wird, sobald neue Testaufgaben erstellt und erprobt werden. In dieser Studie erwiesen sich sowohl 1) die für die Datenerhebung eingesetzten Methoden (Triangulation des Lauten Denkens und *Stimulated-Recall*-Interviews), 2) die für die Analyse der erhobenen Daten gewählte Herangehensweise, d.h. der von A. D. Cohen (u.a. 2006, 2011b, 2014) vorgeschlagene und von Purpura (2014) als *the "Learner Strategies" approach* benannte konzeptuelle Rahmen für die Erfassung der Vorgehensweisen der Testteilnehmenden, als auch 3) die hier angewandte Datenauswertungsmethode der qualitativen Inhaltsanalyse (Kuckartz, 2018) als zielführend und gewinnbringend. Sie können somit für ähnliche Studien empfohlen werden. Selbstverständlich gibt es auch weitere Datenerhebungs- und Datenauswertungsmethoden, die sich für solche Studien anbieten würden (z.B. eine Kombination von *eye-tracking* mit der retrospektiven Befragung wie in der Studie von Suvorov (2018)). Für welche Methode(n) man sich aber auch immer entscheidet – es muss dabei stets bedacht werden, dass „no method provides a fully objective, comprehensive and conclusive picture of cognition“ (Purpura, 2014, S. 1470). Mit Studien wie der vorliegenden können folglich immer nur mehr oder weniger grosse Fragmente des Gesamtbildes erfasst werden.

Für Studien, die sich bei der Festlegung der Kategorien zum Codieren von Strategien an dem Vorschlag von A. D. Cohen (u.a. 2006, 2011b, 2014) orientieren, ist ferner zu beachten, dass die Unterteilung der Strategien in Lerner-, Testmanagement- und *test-wiseness*-Strategien jeweils im Hinblick auf das zugrundeliegende Testkonstrukt vorgenommen werden muss. Dies bedarf einer gründlichen Reflexion und Diskussion darüber, was mithilfe der Testaufgaben erfasst werden soll, d.h. was in einem konkreten Test als konstruktrelevant und was als konstruktirrelevant zu bezeichnen ist.

5.3.1.3 STRATEGISCHE KOMPETENZ ALS TEIL DES TESTKONSTRUKTS

In Übereinstimmung mit ähnlichen Studien zur Untersuchung von Strategien beim Lösen von Testaufgaben (u.a. A. D. Cohen & Upton, 2006; L. Gu & So, 2017; Nikolov, 2006; Rossa, 2012; Suvorov, 2018) zeigte auch diese Studie ganz deutlich, dass die Testteilnehmenden beim Lösen der Testaufgaben auf sehr unterschiedliche Strategien zurückgriffen. L. Gu und So (2017, S. 132) weisen darauf hin, dass die strategische Kompetenz in Modellen kommunikativer Kompetenz (vgl. Kap. 2.2), bereits seit den 1980er Jahren als eigenständige Komponente anerkannt wird (Canale & Swain, 1980; Canale, 1983; Bachman, 1990). Sie postulieren auch aufgrund ihrer Erkenntnisse, die strategische Kompetenz als Teil der Konstruktdefinition für die Entwicklung von Testaufgaben für junge L2-Lernende einzubeziehen.

Aufgrund der Resultate der vorliegenden Studie kann diesem Postulat grundsätzlich zugestimmt werden. Es müsste allerdings erst diskutiert und erforscht werden, welche Rolle die strategische Kompetenz in einem Testkonstrukt haben sollte und wie ihre Erfassung zu operationalisieren wäre.

5.3.2 METHODOLOGISCHE ERKENNTNISSE

In diesem Kapitel werden die wichtigsten Erkenntnisse zu den introspektiven Datenerhebungsverfahren und zur Datenauswertungsmethode dargestellt, die in der vorliegenden Arbeit eingesetzt wurden. Zuerst wird auf das Laute Denken und den *Stimulated Recall* einzeln eingegangen (Kap. 5.3.2.1 und 5.3.2.2), bevor in Kap. 5.3.2.3 die Erkenntnisse aus der Triangulation beider Methoden diskutiert werden. Der für die Auswertung der Daten eingesetzten inhaltlich strukturierenden und evaluativen qualitativen Inhaltsanalyse ist schliesslich Kap. 5.3.2.4 gewidmet.

5.3.2.1 LAUTES DENKEN

Wie in Kap. 4 deutlich wurde, erwiesen sich in der vorliegenden Studie die meisten Lautdenkprotokolle als reichhaltige Informationsquelle. Sie erlaubten es (je nach Schüler/in), informative und detaillierte Einblicke in die Gedanken der Lernenden beim Lösen von Testaufgaben zu gewinnen. Es kann also festgehalten werden, dass sich das Laute Denken als Datenerhebungsmethode bewährte, auch wenn die Erhebung und Auswertung der Daten durchaus als zeit- und ressourcenintensiv bezeichnet werden muss. Angemerkt sei noch, dass die Interpretation der Lautdenkdaten durch die Videoaufnahmen der Gesichter der Schüler/innen sowie die Bildschirmaufnahmen ihrer Aktivitäten beim Lösen der Aufgaben

immens erleichtert wurde: Die Audioaufnahmen der Verbalisierungen alleine hätten oft nicht gereicht, um die Daten im Nachhinein adäquat transkribieren, interpretieren und codieren zu können.

Im Zusammenhang mit dem Lauten Denken wird in der Fachliteratur nicht nur auf die Vorteile, sondern auch auf die Einschränkungen der Methode hingewiesen. Zu den letzteren gehören vor allem 1) ihre Reaktivität, d.h. die möglichen Auswirkungen des lauten Aussprechens von Gedanken auf die Denkprozesse der Forschungsteilnehmenden; 2) die Tatsache, dass beim Lauten Denken nicht alle Gedanken verbalisiert werden; sowie 3) die Tatsache, dass das Laute Denken gewöhnungsbedürftig ist und nicht alle Forschungsteilnehmenden sich für die Datenerhebung mit dieser Methode gleich gut eignen (vgl. u.a. Bowles, 2010; Heine & Schramm, 2016; Leow & Morgan-Short, 2004). Im Folgenden wird auf diese drei Punkte aufgrund der Erfahrungen aus dem vorliegenden Projekt eingegangen.

Um sicherzustellen, dass die von laut denkenden Forschungsteilnehmenden erhobenen Daten valide sind, wird von Leow & Morgan-Short (2004, S. 50) und Bowles (2010, S. 121) empfohlen, eine kleine Kontrollgruppe von Teilnehmenden miteinzubeziehen, die dieselben Aufgaben ohne gleichzeitiges Lautes Denken bewältigen. Sollte sich nämlich zeigen, dass die Ergebnisse in der stillen und in der lauten Denkgruppe statistisch ähnlich sind, könnte man daraus schließen, dass die Verbalisierung keine reaktiven Effekte hatte und dass Lautes Denken die kognitiven Prozesse der Teilnehmenden nicht wesentlich veränderte. Als eine solche (nicht gerade kleine) Kontrollgruppe dienen in der vorliegenden Studie die Teilnehmenden der quantitativen IFB-Stichprobe, die die französischen Versionen der Testaufgaben lösten (n=303). Über den Vergleich der Testresultate der Lautdenk-Schüler/innen und der Schüler/innen der IFB-Französisch-Stichprobe wurde bereits in Kap. 4.1 detailliert berichtet. Die Zusammenstellung der Itemschwierigkeitswerte (die aufgrund der Antworten in der IFB-Französisch-Stichprobe ermittelt wurden) und der Anzahl richtiger und falscher Antworten der laut denkenden Schüler/innen lässt vermuten, dass die Off-Szenario-Aufgaben sowohl für die Lautdenk-Schüler/innen wie für die restlichen Lernenden aus der IFB-Französisch-Stichprobe mittelschwierig bis schwierig waren. Auch waren die Szenario-Aufgaben für die IFB-Französisch-Stichprobe und die Lautdenk-Schüler/innen vergleichbar schwierig, wobei gewisse Items für die Lernenden der Lautdenk-Gruppe scheinbar einfacher als für die Schüler/innen der IFB-Französisch-Stichprobe waren (vgl. Tabelle 12 und Tabelle 13). Das Laute Denken scheint folglich zumindest auf den ersten Blick keine negativen Auswirkungen auf die Testresultate der Lautdenk-Schüler/innen gehabt zu haben.

Untersuchungen zum Lauten Denken zeigen auch, dass Forschungsteilnehmende beim Einsatz dieser Methode generell mehr Zeit für die Bewältigung der zu lösenden Aufgabe brauchen (Bowles, 2010, S. 119). Um Hinweise darauf zu bekommen, inwiefern Lautes Denken die Testbearbeitungszeit in der vorliegenden Studie beeinflusste, wurde die Testlösungszeit der 17 Schüler/innen, die das Szenario „Stadttrip“ bearbeiteten, mit der Testlösungszeit der Schüler/innen aus der IFB-Französisch-Stichprobe verglichen.¹²⁸ Dieser Vergleich war möglich, da die Testauslieferungsplattform EE4CBA (DIPF & Nagarro, 2018a) bei jeder Aktion der Testteilnehmenden die aktuelle Uhrzeit aufzeichnet (z.B. beim Drücken einer Schaltfläche). Diese Daten können u.a. genutzt werden, um den Beginn und das Ende der Bearbeitung einer Aufgabe zu erfassen und so die Testbearbeitungszeit zu berechnen. Die vier untenstehenden Grafiken (Abbildung 27) zeigen die Anzahl der Sekunden, die die einzelnen Schüler/innen aus der qualitativen (blau markiert) und aus der quantitativen Stichprobe (orange markiert) zum Bearbeiten der jeweiligen Tasks brauchten.

¹²⁸ Auf den Vergleich der Bearbeitungszeit der Off-Szenario-Schüler/innen wurde verzichtet. Grund dafür ist, dass diese Lernenden jeweils nur zwei von drei Bildschirmseiten pro Task aus den Szenarien „Walkies“, „Stadttrip“, „Berufswahl“ und „Waitaha“ lösten und so der Vergleich mit den „normalen“ Testbearbeitungszeiten nur bedingt verlässlich ist. Auch sahen diese Schüler/innen keine Szenario-Einbettung und arbeiteten mit Aufgaben aus verschiedenen Szenarien und zu verschiedenen Themen, was womöglich ihre Bearbeitungszeit zusätzlich beeinflusste.

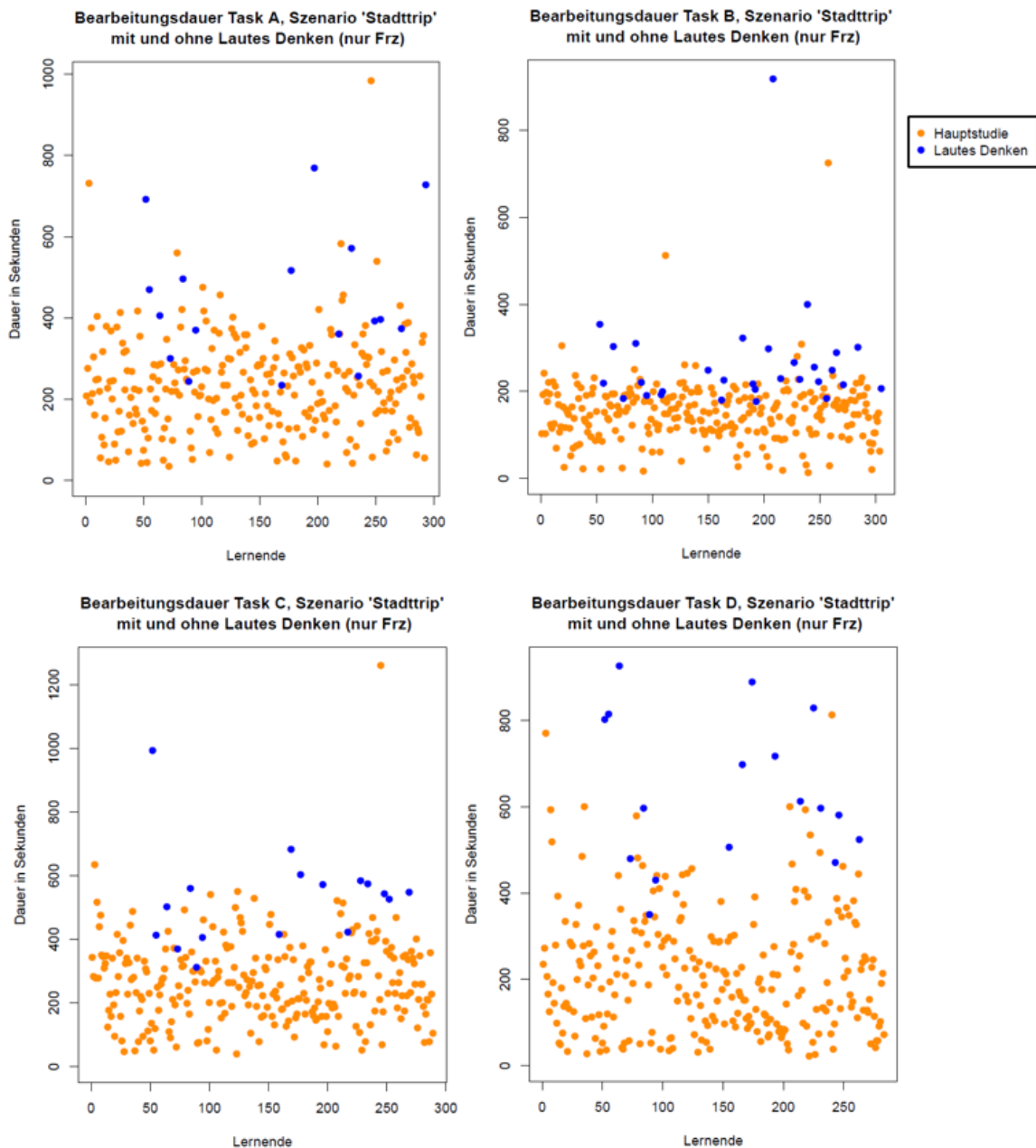


Abbildung 27: Bearbeitungsdauer der Tasks im Szenario „Stadttrip“ mit und ohne Lautes Denken.

Es zeigt sich, dass die Lautdenk-Schüler/innen tendenziell etwas länger an den Aufgaben arbeiteten als die meisten (aber nicht alle) Schüler/innen aus der Gesamtstichprobe. Dies deckt sich mit der Aussage der Schülerin Picard, die auf die Frage nach dem Einfluss des Lauten Denkens auf ihre Auseinandersetzung mit den Testaufgaben antwortete:

es hat ein bisschen länger gedauert (Picard_Off, Pos. 1194)

Auch Steve meinte, er hätte die Aufgaben ohne Lautes Denken schneller (und dadurch wahrscheinlich auch weniger gründlich) gelöst, weil er sich beim Lauten Denken mehr überlegte:

MB: [...] meinst du dass du [ohne lautes denken] ANdere antworten (--)
wählen würdest`
S: <<bejahend>mh_mh>; schneller also (-) weniger überlegen (Steve_Off,
Pos. 181-182)

Die oben dargestellte Evidenz könnte als Indiz dafür interpretiert werden, dass das Laute Denken auch in der vorliegenden Studie einen verlangsamenden Einfluss auf den Testbewältigungsprozess hatte. Es kann aber auch nicht ausgeschlossen werden, dass bereits die Testdurchführungssituation der Lautdenk-Schüler/innen, die die Aufgaben in einem separaten Raum und in Anwesenheit der Forscherin bearbeiteten, sowie das Wissen darüber, dass sie danach zu ihren Gedanken befragt würden, die Vorgehensweise der Schüler/innen insofern beeinflusste, dass sie die Aufgaben sorgfältiger, also womöglich auch langsamer bearbeiteten.

Weitere Hinweise zur Reaktivität des Lauten Denkens als Datenerhebungsmethode lassen sich den Äusserungen der Schüler/innen in den *Stimulated-Recall*-Interviews entnehmen. Wie bereits in Kap. 4.5.1 ausführlich dargestellt, sagten im Interview 6 von 19 dazu befragten Lernenden, dass sie das Laute Denken beim Lösen der Testaufgaben (eher) störte, weil es für sie anstrengend oder ungewöhnlich war. Von zwei Lernenden (darunter von Steve, den das Laute Denken gleichzeitig auch störte) wurde angemerkt, dass das Verbalisieren der Gedanken einen positiven Einfluss auf ihren Testbewältigungsprozess hatte. Der Grund dafür war, dass sie sich dadurch mit den Aufgaben intensiver auseinandersetzten. Der Mehrheit der Schüler/innen (n=12) scheint jedoch das Laute Denken nicht viel ausgemacht zu haben: Sie waren der Meinung, dass es weder einen positiven noch einen negativen Einfluss auf ihre Vorgehensweise im Test hatte. Dieses Ergebnis scheint die Vermutung von Heine zu bestätigen, gemäss welcher anzunehmen sei, dass „kognitive Verarbeitung stark individuell geprägt ist“ und „[d]er Grad der Reaktivität der Methode [...] somit nicht pauschal, sondern immer nur in Bezug auf das jeweilige Individuum betrachtet werden“ kann (2014, S. 130).

Dass beim Lauten Denken nicht alle Gedanken verbalisiert werden, konnte auch in dieser Studie beobachtet werden. Um dieser bekannten Tatsache bereits im Voraus möglichst entgegenzuwirken, wurde die Triangulation der Lautdenkprotokolle mit den *Stimulated-Recall*-Interviews vorgenommen. Der Vergleich der im Lautdenken- und *Stimulated-Recall*-Teil vorgenommenen Codierungen zeigte, dass ohne *Stimulated-Recall*-Interviews die Lernerstrategien „Weiterlesen, um Nichtverstandenes zu verstehen“ und „Prosodie beim Hören nutzen“, die Testmanagementstrategie „Zeitmanagement“ sowie die *test-wiseness*-Strategien „Auswahl einer Antwortoption mit bekannten Wörtern“ und „Auswahl einer anders aussehenden Antwortoption“ unentdeckt geblieben wären.

Bei der Datenerhebung konnte auch beobachtet werden, dass eine sorgfältig formulierte Instruktion zur Vorgehensweise beim Lauten Denken sehr hilfreich ist und dass das Training erheblich zur Validität der erhobenen Daten beitragen kann. Dennoch zeigte sich auch in der vorliegenden Studie, dass die Methode für gewisse Teilnehmende gewöhnungsbedürftig ist. Trotz der Hinweise in der Instruktion, beim Lauten Denken mit sich selbst zu sprechen, die Gedanken frei laufen zu lassen und nicht zu versuchen, jemandem etwas zu erklären, richteten einige Schüler/innen ihre Äusserungen eher an die Forscherin als an sich selbst (was übrigens auch von Nikolov (2006, S. 28) und von Feick (2013) in ihrer Studie zum videobasierten Lauten Erinnern beobachtet werden konnte). Da es sich bei diesen Verbalisierungen um Einblicke in die Gedanken und Vorgehensweisen der Lernenden handelte, die für die Beantwortung der Forschungsfragen ebenfalls interessant waren, wurde entschieden, in dieser Arbeit auch die metakognitiven Äusserungen der Schüler/innen bei der Datenauswertung mitzuberücksichtigen. Weiter musste auch in der vorliegenden Studie festgestellt werden, dass sich nicht alle Forschungsteilnehmende für die Datenerhebung mit dem Lauten Denken gleich gut eignen, was zu Datenverlusten führt. Um dem vorzubeugen ist es ratsam, die Datenerhebung (wie in dieser Studie) bereits im Voraus mit mehr Teilnehmenden zu planen, als man tatsächlich auszuwerten gedenkt.

5.3.2.2 *STIMULATED RECALL*

Der Einbezug von *Stimulated-Recall*-Interviews in das Forschungsdesign erwies sich für die vorliegende Studie als sehr vorteilhaft. In den Interviews konnten die Lernenden gezielt und detailliert zu ihren Vorgehensweisen befragt werden und Auskunft erteilen. Wie bereits in Kap. 5.3.2.1 erwähnt, wurden in diesen Gesprächen von den Lernenden die Strategien „Weiterlesen, um Nichtverstandenes zu verstehen“, „Prosodie beim Hören nutzen“, „Zeitmanagement“, „Auswahl einer Antwortoption mit bekannten Wörtern“ und „Auswahl einer anders aussehenden Antwortoption“ angesprochen, die in den Lautdenkprotokollen nicht beobachtet werden konnten. Vor allem im Vergleich mit den entsprechenden Passagen aus den Lautdenkprotokollen wurden aber auch Einschränkungen der Methode deutlich sichtbar, auf die im Folgenden näher eingegangen wird.

Stimulated-Recall-Interviews werden nach der Tätigkeit durchgeführt, die im Fokus einer Untersuchung steht. Wie bereits in Kap. 3.3 erläutert, wird dabei auf die Erinnerungen zurückgegriffen, die im Langzeitgedächtnis der Forschungsteilnehmenden abgespeichert wurden. Dies hat zur Folge, dass 1) sich die Erinnerungen nicht nur auf die Tätigkeit selbst beziehen, sondern auch mit den Einstellungen der Forschungsteilnehmenden und mit ihrem Selbstbild verknüpft sein können; 2) die Erinnerungen teilweise lückenhaft sein können bzw.

nicht ganz der tatsächlichen Vorgehensweise bei der untersuchten Tätigkeit entsprechen (vgl. Gass & Mackey, 2017; Heine & Schramm, 2016). In der vorliegenden Studie konnten beide oben genannten Punkte beobachtet werden.

Die Vermischung von Erinnerungen der Schüler/innen an das Lösen der Testaufgaben mit ihren Einstellungen und ihrem Selbstbild war in der vorliegenden Studie erwünscht und erwies sich im Hinblick auf die Interpretation der Codierungen aus den Lautdenkprotokollen als bereichernd. Dadurch konnte besser und nicht selten zuverlässiger nachvollzogen werden, aus welchen Gründen sich die Schüler/innen für bestimmte Strategien und bestimmte Antworten im Test entschieden (oder auch nicht entscheiden). Alison, die in den untersuchten Items 8 von 12 Punkten erreichte, gehörte beispielsweise zu den Lernenden, die beim Lauten Denken eher wenig verbalisierten. Aufgrund ihrer Aussagen im *Stimulated-Recall*-Interview und ihrer mündlichen Zusammenfassungen der Inputtexte konnte festgestellt werden, dass sie im Französischen zu den stärksten Schüler/innen der qualitativen Stichprobe gehörte und eigentlich fast alle Aufgaben korrekt hätte lösen können. Sie versuchte jedoch, die Testaufgaben möglichst schnell zu bearbeiten und las die Inputtexte nicht immer aufmerksam oder auch nicht zu Ende durch. Bei der intensiveren Auseinandersetzung mit den Inputtexten im Interview merkte und korrigierte sie die meisten ihrer Fehler selbst. Nur aufgrund der Lautdenkprotokolle wäre es deutlich schwieriger gewesen, die Gedanken von Alison und ihre Gründe für die Wahl bestimmter Antworten im Test adäquat zu rekonstruieren.

Als problematisch erwiesen sich in dieser Studie die unzuverlässigen Erinnerungen der Lernenden an ihre Vorgehensweise bei der Bearbeitung der Testaufgaben. Es zeigte sich, dass die Interviews alleine, auch wenn sie zweifelsohne eine hervorragende Informationsquelle darstellen, teilweise zu falschen Einschätzungen und Codierungen der Vorgehensweisen der Schüler/innen geführt hätten, wenn sie als einzige Informationsquelle vorgelegen hätten. Dies lag wohl daran, dass sich einige Lernende im Gespräch nicht mehr ganz genau an ihre Vorgehensweise beim Lösen der Aufgaben erinnern konnten, obwohl die Interviews so zeitnah nach der Bewältigung der Testaufgaben durch die Schüler/innen durchgeführt wurden, wie es nur möglich war. Die Erinnerungslücken mancher Schüler/innen sind vermutlich u.a. darauf zurückzuführen, dass sie zum Zeitpunkt des Gesprächs etwas müde waren. Auch können die Erinnerungslücken partiell mit Erinnerungen an ähnliche Situationen gefüllt werden. Durch den im zweiten Codierdurchgang vorgenommenen Vergleich der Lautdenkprotokolle mit den entsprechenden Ausschnitten aus den *Stimulated-Recall*-Interviews (vgl. Kap. 3.7.2) konnten die unzuverlässigen Erinnerungen der Schüler/innen entdeckt werden. Die im ersten Codierdurchgang nur aufgrund der

Interviewaussagen vorgenommenen Codierungen konnten folglich entfernt und mit einem entsprechenden Memo versehen werden (vgl. Beispiel in Kap. 3.7.2).

Der Vergleich der Codierungen zwischen dem Lautdenken- und *Stimulated-Recall*-Teil zeigt ausserdem, dass in den Lautdenkprotokollen Strategien beobachtet und codiert wurden, die von den Schüler/innen in den Interviews nicht erwähnt wurden. Wären in der vorliegenden Studie nur die *Stimulated-Recall*-Interviews eingesetzt worden, wären die Lernerstrategien „Bisheriges Verständnis hinterfragen“, „Bisheriges Verständnis bestätigen“ und „Selbsteinschätzung“, sowie die Testmanagementstrategien „Selbstkorrektur R-F (richtig-falsch)“ und „Selbstkorrektur F-F (falsch-falsch)“ unbemerkt geblieben.

Auch muss an dieser Stelle die Tatsache reflektiert werden, dass die Interviewer/innen durch die Art und Weise ihrer Gesprächsführung einen nicht zu unterschätzenden Einfluss auf den Gesprächsverlauf und die generierten *Stimulated-Recall*-Daten haben können. Die Autorin der vorliegenden Studie bemühte sich sehr darum, die Fragen in allen Interviews möglichst gleich und auch möglichst nicht suggestiv zu stellen. Bei der Sichtung der Interviewtranskripte musste sie trotzdem selbstkritisch feststellen, dass ihr das nicht immer gleich gut gelungen war. Je nach Fall wurde dies bei der Codierung im Nachhinein berücksichtigt und in einem Memo festgehalten.¹²⁹ Dieses unerwünschte Verhalten war teilweise der Müdigkeit geschuldet, die sich bei den über zweistündigen und intensiven Datenerhebungssitzungen nicht nur bei den Schüler/innen, sondern auch bei der Forscherin bemerkbar machte. Einen nicht unerheblichen Einfluss hatte aber auch die Stimmung bei der Datenerhebung, die jeweils auch von den Schüler/innen mitbestimmt war. Einige Schüler/innen, wie z.B. Arya, erzählten von sich aus viel und ausführlich über ihre Gedanken und Vorgehensweise beim Lösen der Testaufgaben. Anderen Schüler/innen, u.a. Steve und Alison, mussten hingegen sehr viele Fragen gestellt werden, um von ihnen das Wichtigste zu erfahren. Auch äussere Faktoren wie Lärm oder Störungen durch Lehrpersonen oder Mitschüler/innen wirkten sich in einigen Fällen negativ auf den Verlauf der Datenerhebung aus. Davon war im Übrigen gelegentlich auch das Laute Denken betroffen.

¹²⁹ Cersei wurde beispielsweise im Gespräch gefragt: „MB: [...] warst du dir da SICHer bei äh (-) dieser option oder hast du eher geraten;“ (Cersei_Szenario, Pos. 840-841). Die Antwort der Schülerin wurde im Transkript uncodiert gelassen und das entsprechende Memo lautete: „Hier wurde TW01 [„Raten“] nicht codiert, weil die Frage suggestiv gestellt wurde und die S auch nicht zugibt, dass sie geraten hat, auch wenn sie sagt, dass sie sich nicht sicher war, ob ihre Antwort stimmt oder nicht.“

5.3.2.3 LAUTES DENKEN UND *STIMULATED RECALL* ALS EINANDER ERGÄNZENDE DATENERHEBUNGSMETHODEN

Wie aus den Ausführungen in Kap. 5.3.2.1 und 5.3.2.2 sichtbar wurde, erwiesen sich die beiden introspektiven Datenerhebungsmethoden in der vorliegenden Studie als wertvoll und komplementär zueinander. Die Tatsache, dass einige Strategien beim Lauten Denken nicht verbalisiert wurden, aber dafür beim *Stimulated Recall* zur Sprache kamen (z.B. Strategie L16 „Prosodie beim Hören nutzen“), scheint die Vermutung zu bestätigen, dass die Lernenden beim Lauten Denken nicht alle ihre Gedanken verbalisierten. Andere Strategien wurden wiederum in den *Stimulated-Recall*-Interviews gar nicht angesprochen, wie z.B. Strategie L05 „Bisheriges Verständnis hinterfragen“. Dies ist wahrscheinlich darauf zurückzuführen, dass sich die Lernenden im anschliessenden Gespräch nicht an alle ihre Gedanken bzw. Vorgehensweisen beim Lösen der Aufgaben erinnern konnten. Der Vergleich der Anzahl codierter Strategien in den Lautdenkprotokollen und *Stimulated-Recall*-Interviews zeigte ausserdem, dass die Codierungen der gleichen Strategie in beiden Untersuchungsteilen z.T. stark voneinander abweichen, was – wie bereits in Kap. 4.2 erläutert – vor allem mit den unterschiedlichen Eigenschaften der mit der jeweiligen Methode erhobenen Daten zusammenhängt. Die Triangulation der beiden Datenerhebungsmethoden führte folglich zu aufschlussreicheren Einsichten in die Strategien der Schüler/innen beim Lösen der Testaufgaben.

Durch den Einsatz der beiden introspektiven Verfahren konnten darüber hinaus die Schwachstellen der beiden Methoden ausgeglichen werden, was auch von Gass und Mackey in ihrem Buch zu *Stimulated Recall* als Datenerhebungsmethode empfohlen wird: „[t]he triangulation of stimulated recall data with other measures of noticing is useful in mitigating potential limitations of the methodology“ (2017, S. 101). Für die Deutung der unklaren Stellen in den Lautdenkprotokollen leisteten die Ausführungen der Schüler/innen in den *Stimulated-Recall*-Interviews in der vorliegenden Studie eine grosse Interpretationshilfe. Die Lautdenkprotokolle waren wiederum behilflich, um Erinnerungslücken oder unzuverlässige Erinnerungen der Schüler/innen zu entdecken und beim Codieren entsprechend zu berücksichtigen. Ähnliches stellte auch Suvorov (2018) in seiner Studie fest, in der *eye-tracking* mit Verbalprotokollen der Testteilnehmenden zu ihren Teststrategien trianguliert wurde: In einigen Fällen konnten Diskrepanzen zwischen den Erinnerungen der Teilnehmenden an ihre Vorgehensweise und der Evidenz aus den *eye-tracking*-Daten ermittelt werden. In solchen Fällen wurde die Evidenz aus den Blickbewegungsdaten den Interview-Aussagen vorgezogen. Die Erklärungen der Testteilnehmenden in den Interviews fand Suvorov wiederum für die Interpretation der *eye-tracking*-Daten nützlich: Es zeigte sich

beispielsweise, dass eine Teilnehmerin gemäss eigenen Aussagen dazu tendierte, beim Lösen der Testaufgaben ein beliebiges Wort anzustarren, ohne sich dabei kognitiv mit diesem Wort oder mit der Aufgabe zu beschäftigen.

Die Triangulation der Lautdenkprotokolle mit den *Stimulated-Recall*-Interviews war für die vorliegende Studie, wie bereits gesagt, äusserst gewinnbringend. Sie war aber auch mit einem grossen Zeit- und Ressourcenaufwand verbunden, der für die Erhebung, Aufbereitung, Codierung und Analyse der Daten aufgebracht werden musste. Auch wenn die Vorteile der Kombination dieser beiden Methoden mannigfaltig sind, muss letztlich in jeder Studie im Hinblick auf das Erkenntnisinteresse, die Forschungsfragen und die zur Verfügung stehenden Ressourcen erwogen werden, ob eine solche Methodentriangulation im Rahmen des Möglichen liegt. Die Erfahrung aus dem IFB-Vorgängerprojekt „Task Lab“ (vgl. Karges, Barras et al., 2021) zeigt, dass bei wenigen Ressourcen für die qualitative Pilotierung von neuen Testaufgaben *Stimulated-Recall*-Interviews trotz ihrer Einschränkungen hilfreich sind, um die wichtigsten Schwachstellen von neuen Aufgaben identifizieren zu können (vgl. Barras, 2018). Um die aufwändigen Transkriptionen der Gespräche zu vermeiden, wäre es auch denkbar, die wichtigsten Erkenntnisse aus den Interviews direkt auf Screenshots mit den Testaufgaben zu notieren und die Testitems auf dieser Grundlage zu überarbeiten, was im Projekt „Task Lab“ u.a. auch gemacht wurde (Karges, Barras et al., 2021). Dies könnte sich nicht nur für Forschungszwecke im eigentlichen Sinne, sondern auch für Lehrpersonen, die mehr über die von ihnen erstellten Beurteilungsinstrumente erfahren wollen, als sehr nützlich erweisen. Lehrpersonen könnten die *Stimulated-Recall*-Methode ausserdem auch im Sinne des *dynamic assessment* (vgl. u.a. Antón, 2012; Studer, 2016) einsetzen, um (Test-)Aufgaben aus ihrem Unterricht mit den Lernenden zu besprechen und so ein vollständigeres Bild über die Sprachkenntnisse ihrer Schüler/innen zu bekommen.

5.3.2.4 DIE INHALTLICH STRUKTURIERENDE UND EVALUATIVE QUALITATIVE INHALTSANALYSE ALS DATENAUSWERTUNGSMETHODEN¹³⁰

Die inhaltlich strukturierende und die darauf aufbauende evaluative qualitative Inhaltsanalyse (Kuckartz, 2018) bestätigten sich rückblickend als geeignete und zielführende Methoden für die Auswertung der vorliegenden, umfassenden Verbalprotokolle. Als besonders vorteilhaft zeigten sich dabei zwei Charakteristika der inhaltlich strukturierenden qualitativen Inhaltsanalyse. Erstens trug die Tatsache, dass bei dieser interpretativen Form der Inhaltsanalyse die Komplexität der Daten planmässig und kategorienbasiert reduziert wird,

¹³⁰ Auf die in dieser Studie unternommenen Bemühungen zur Einhaltung von Gütekriterien der qualitativen Inhaltsanalyse (vgl. Kuckartz, 2018, Kap. 9) wurde bereits in Kap. 3.7.4 ausführlich eingegangen.

erheblich zur systematischen und transparenten Darstellung der vorliegenden Resultate bei. Zweitens besteht das Ziel der Codierung in der inhaltlich strukturierenden qualitativen Inhaltsanalyse nicht primär darin, die vorgenommenen Codierungen zu quantifizieren. Stattdessen bleiben die codierten Aussagen auch nach der Codierphase wichtig und ihnen wird bei der Präsentation der Ergebnisse ein gebührender Platz gewährt. Wie in Kap. 4 deutlich geworden sein dürfte, wären in der vorliegenden Studie viele wichtige Erkenntnisse verborgen geblieben, wenn in dieser Studie nur die Quantifizierungen der codierten Segmente vorgenommen worden wären (man hätte dann beispielsweise zwar gesehen, wie oft bestimmte Strategien codiert wurden, es wäre aber verborgen geblieben, aus welchen Gründen die Schüler/innen eine bestimmte Strategie auswählten).

Die evaluative qualitative Inhaltsanalyse, die in der vorliegenden Studie aufbauend auf den Resultaten der inhaltlich strukturierenden qualitativen Inhaltsanalyse durchgeführt wurde, ermöglichte ihrerseits, die Erkenntnisse aus den gewonnenen Daten weiter zu nuancieren und zu differenzieren. Dies war besonders bei der Beantwortung der zweiten Forschungsfrage gewinnbringend: Die Codierungen der *test-wiseness*-Strategien, die bei der inhaltlich strukturierenden Inhaltsanalyse vorgenommen wurden, konnten in einem nächsten Schritt im Hinblick auf ihren Beitrag zur richtigen oder falschen Beantwortung der Testfragen eingeschätzt und bewertet werden (vgl. Kap. 4.3).

5.3.3 MÖGLICHE IMPLIKATIONEN DER STUDIE FÜR DEN FREMDSPRACHENUNTERRICHT

Die vorliegende Studie ist dem Bereich der Sprachtestforschung zuzuordnen und die gewonnenen Erkenntnisse sollten in erster Linie für Testentwickler/innen und Testforschende von Bedeutung sein. Dennoch können von den hier dargestellten Einblicken in die Testlösungsprozesse von Schüler/innen auch Fremdsprachenlehrpersonen, Fremdsprachendidaktiker/innen und Entwickler/innen von Unterrichtsmaterialien profitieren, um Lern- und Beurteilungsanlässe der Lernenden im Fremdsprachenunterricht optimaler zu gestalten und so die Schüler/innen beim Erlernen von Fremdsprachen und Bearbeiten von Sprachtestaufgaben besser zu fördern (vgl. Arras, 2013; Sasaki, 2014).

Wie in Kap. 4.5.3 dargelegt, zeigten die Gespräche mit einigen Schüler/innen, dass sie bisher kaum Erfahrung mit standardisierten Sprachtestformaten hatten. Sie waren beispielsweise überrascht, dass bei einer Multiple-Choice-Aufgabe im Inputtext nicht nur die richtige Antwortoption, sondern auch Elemente aus den Distraktoren vorkamen. Einige Lernende wählten deswegen bei der Bearbeitung einer Aufgabe sofort eine Antwortoption aus, zu der sie im Inputtext ein passendes Wort fanden und lasen auch die Inputtexte nicht ordentlich

durch. Dies führte oft dazu, dass die gewählte Antwort nicht korrekt war. Den mündlichen Zusammenfassungen der Inputtexte während der *Stimulated-Recall*-Interviews war zu entnehmen, dass gewisse Lernende wie z.B. Alison durchaus in der Lage gewesen wären, bei einer aufmerksameren Lektüre die korrekten Antwortoptionen zu wählen. Aus diesem Grund kann Fremdsprachenlehrpersonen empfohlen werden, ihre Schüler/innen darauf aufmerksam zu machen, dass sie beim Lösen von Testaufgaben die Inputtexte genau(er) zur Kenntnis nehmen und nicht etwa die erste Antwortoption wählen sollten, zu der sie im Inputtext ein passendes Wort identifizieren können (auch sollten Lehrpersonen ihre eigenen Tests möglichst so konzipieren, dass der Einsatz solcher Strategien nicht zum Erfolg führt). Mit der Thematisierung der Strategien beim Lösen von Testaufgaben im Fremdsprachenunterricht ist hier jedoch nicht gemeint, dass mit den Lernenden alle Strategien behandelt werden sollen, die bei einem Test zum Einsatz kommen könnten (vgl. auch Nikolov, 2006). Die Resultate der vorliegenden Studie veranschaulichen u.a., dass der Einsatz von *test-wiseness*-Strategien überwiegend zu falschen Antworten im Test führte. Auch zeigte sich hier, dass den Schüler/innen bestimmte konstruktrelevante Strategien wie z.B. die Suche nach Kognaten oder die Nutzung von grammatischen Informationen bei der Bearbeitung der Testaufgaben nicht behilflich sind, wenn ihre allgemeinen Sprachkompetenzen und vor allem ihre Wortschatzkenntnisse nicht ausreichend sind. Den Lehrpersonen kann daher empfohlen werden, die zu thematisierenden Strategien mit Blick besonders auch auf den Stand der Kompetenzentwicklung der Lernenden abzuwägen.

Die Interviews mit den Schüler/innen zeigten darüber hinaus, dass die IFB-Testaufgaben bei den meisten von ihnen sehr gut ankamen. Die Begeisterung vieler Schüler/innen für diese Aufgaben kann z.T. bestimmt auf ihren Neuigkeitseffekt und ihre computerbasierte und attraktive Aufmachung zurückgeführt werden. Sie wurden von den Lernenden aber nicht nur deswegen geschätzt: Einige Schüler/innen erwähnten, dass die Aufgaben gut erklärt, „modern“, nah an ihrem Alltag und auch thematisch sowie von den Textsorten her interessant waren. Howard meinte auch dazu, dass diese Aufgaben in seinen Augen mehr für die Schüler/innen als für die Lehrpersonen gemacht wurden. Für Testentwickler/innen heisst das, dass sie sich weiterhin darum bemühen sollten, ihre Testaufgaben so interessant und motivierend wie möglich zu gestalten, denn gemäss Wagner (2014, S. 59) kann durch *large-scale tests* (vor allem wenn es sich dabei um *high stakes tests* handelt) ein positiver Washback-Effekt auf die Curricula, den Fremdsprachenunterricht und sogar die ganze Gesellschaft erzielt werden. Was und wie getestet wird, muss nämlich zuerst unterrichtet werden.

Die vorliegenden Daten zeigen jedoch auch, dass sich der motivierende Effekt der Testaufgaben aufgrund der Testsituation in Grenzen hielt, denn die Schüler/innen waren gemäss ihrer eigenen Angaben vor allem daran interessiert, die Antworten im Test richtig auszuwählen und nicht daran, aus den Inputtexten etwas zu lernen (vgl. Barras et al., 2020). Es kann daher vermutet werden, dass Aufgaben wie die im Projekt IFB entwickelten ihre motivierende Wirkung im Fremdsprachenunterricht viel stärker als während eines Tests entfalten könnten (vgl. z.B. die in Kap. 4.5.2 zitierten Aussagen von Arya, die behauptete, dass sie beim Bearbeiten von alltagsbezogenen Themen und Aufgaben motivierter ist, oder von Omega, den die „coolen“ und „witzigen“ Inputtexte beim Lösen der IFB-Testaufgaben zuerst motivierten, bis er sich daran erinnerte, dass es Testaufgaben sind und dass er sich jetzt lieber doch auf das Finden von richtigen Lösungen konzentrieren soll). Es wäre auf jeden Fall interessant, dies zu untersuchen und bei der Gelegenheit auch mehr darüber zu erfahren, ob und welches Potenzial die szenariobasierten Aufgaben beim Einsatz im Fremdsprachenunterricht haben könnten: Es könnte sein, dass die Einbettung der einzelnen Aufgaben in ein übergreifendes Szenario mehr für den Fremdsprachenunterricht als für die Überprüfung der Sprachkompetenzen der Lernenden bringen könnte.

Schliesslich machten die in dieser Studie eingesetzten introspektiven Verfahren sehr deutlich, welche wichtige Erkenntnisse den Äusserungen der Lernenden entnommen werden können. Es wäre daher wünschenswert, dass nicht nur bei der Validierung der Testaufgaben, sondern auch bei der Konzipierung von Unterrichtsmaterialien und Lernaufgaben vermehrt auf die Stimmen der Lernenden gehört wird. Ohne Zweifel könnten auch Lehrpersonen ab und zu aus den (in einer vereinfachten Form durchgeführten) *Stimulated-Recall*-Interviews im Rahmen der Erforschung ihres eigenen Unterrichts nutzbringende Einblicke in das Lernen ihrer Schüler/innen gewinnen.

6 FAZIT UND AUSBLICK

Das Hauptziel der vorliegenden Studie war es, mittels Lautem Denken und *Stimulated-Recall*-Interviews die Strategien zu beforschen, die Testteilnehmende beim Bearbeiten von neu erstellten, prototypischen, szenariobasierten Testaufgaben einsetzten. Dadurch sollte wichtige und auf Basis quantitativer Methoden unzugängliche Evidenz zum Funktionieren und zur Validität der beforschten Aufgaben gewonnen werden, was vor allem bei der Entwicklung von neuen und wenig bekannten Item- und Aufgabentypen von grosser Bedeutung ist (vgl. Kenyon & MacGregor, 2012; Reed, 2014). Die Studie sollte Erkenntnisse dazu liefern, 1) welche Strategien bei Schüler/innen beim Lösen der Testaufgaben beobachtet werden können, 2) ob sich Zusammenhänge zwischen konstruktirrelevanten Strategien und erfolgreichen bzw. nicht erfolgreichen Aufgabenlösungen erkennen lassen und 3) ob sich Unterschiede beim Einsatz der Strategien feststellen lassen, wenn szenariobasierte und nicht szenariobasierte Testaufgaben gelöst werden, und, falls ja, worauf dies zurückgeführt werden könnte.

An der Untersuchung, die in Kap. 3 detailliert vorgestellt wurde, nahmen insgesamt 30 Französischlernende aus Klassen teil, die am quantitativ ausgerichteten Projekt IFB beteiligt waren. In individuellen Sitzungen mit der Forscherin lösten diese 30 Schüler/innen während 3 Lektionen à 45 Minuten entweder die Aufgaben aus dem Szenario „Stadttrip“, oder eine vergleichbare Anzahl thematisch losgelöster Aufgaben, die aus unterschiedlichen Szenarien stammten. Sie verbalisierten dabei ihre Gedanken und wurden im Nachhinein in den *Stimulated-Recall*-Interviews zu ihrer Vorgehensweise beim Lösen der Testaufgaben befragt. Im Rahmen der vorliegenden Untersuchung wurden die Transkriptausschnitte von 20 kriterienbasiert ausgewählten Schüler/innen zu ausgewählten Testaufgaben vertieft analysiert. Als Datenauswertungsverfahren wurde die inhaltlich strukturierende und evaluative Inhaltsanalyse nach Kuckartz (2018) gewählt. Die Analyse der Strategien orientierte sich am Vorschlag von A. D. Cohen (u.a. 2006, 2011b, 2014), der Strategien beim Lösen von Testaufgaben in konstruktrelevante Lernerstrategien und Testmanagementstrategien sowie konstruktirrelevante *test-wiseness*-Strategien unterteilt.

Die Ergebnisse der vorliegenden Untersuchung wurden in Kap. 4 detailliert vorgestellt und im anschliessenden Kap. 5 zusammengefasst und diskutiert. Im Hinblick auf die Beantwortung der ersten Forschungsfrage zeigte sich, dass die Lernenden beim Lösen der Testaufgaben auf verschiedene Strategien zurückgriffen, von denen die meisten konstruktrelevant waren. Dies kann als positives Zeichen gedeutet werden, weil es bedeutet, dass sich die Testteilnehmenden vor allem darum bemühten, die französischen Inputtexte zu verstehen und

die Testfragen richtig zu beantworten. Nicht alle konstruktrelevanten Strategien führten aber automatisch zum Erfolg: Wurden die Inputtexte nicht genügend verstanden, konnten die Aufgaben in den meisten Fällen auch unter Einbezug dieser Strategien nicht korrekt gelöst werden. Die konstruktirrelevanten *test-wiseness*-Strategien wurden von den Schüler/innen vor allem dann eingesetzt, wenn ihre Sprachkompetenzen nicht ausreichten, um die Inputtexte zu verstehen und die Testaufgaben mithilfe der konstruktrelevanten Strategien zu lösen, oder auch dann, wenn sie die Testaufgaben schnell lösen wollten und sich nicht gründlich genug mit ihnen auseinandersetzten. Dank der Untersuchung der Teststrategien konnten auch einige Elemente der Inputtexte und/oder der Aufgaben identifiziert werden, die vor einem eventuellen operativen Einsatz der IFB-Aufgaben im Rahmen eines Bildungsmonitorings überarbeitet und nochmals erprobt werden müssten.

Die Ergebnisse zur zweiten Forschungsfrage brachten an den Tag, dass die konstruktirrelevanten *test-wiseness*-Strategien in den meisten Fällen von den schwächsten Französischlernenden eingesetzt wurden, was dann meistens auch zu nicht erfolgreichen Lösungen der Testaufgaben führte. Der erfolgreiche Einsatz der *test-wiseness*-Strategien war vor allem dem Zufall, in einigen wenigen Fällen aber auch bestimmten Eigenschaften der Testaufgaben geschuldet, die aufgrund der vorliegenden Evidenz überarbeitet werden müssten. Dieses alles in allem ebenfalls positive Resultat kann als Evidenz dafür gedeutet werden, dass die IFB-Aufgaben in den meisten Fällen nicht für den Einsatz von konstruktirrelevanten Strategien anfällig waren.

Die Resultate zur dritten Forschungsfrage lassen erkennen, dass die Szenarioeinbettung der IFB-Aufgaben von den Szenario-Schüler/innen kaum bewusst wahrgenommen und von den Off-Szenario-Schüler/innen nicht vermisst wurde. Dies kann eventuell darauf zurückgeführt werden, dass in dieser Studie nur ein einziges Szenario und wenige Off-Szenario-Aufgaben vertieft untersucht werden konnten: Es wäre daher wichtig, ähnliche Untersuchungen zu einer breiteren Auswahl an Szenarien und Off-Szenario-Aufgaben durchzuführen. Allerdings konnten schon im hier eingeschränkten Rahmen Unterschiede beim Einsatz der Strategien beobachtet werden, wenn Szenario- und Off-Szenario-Aufgaben gelöst wurden. Die Gründe für diese Unterschiede lassen sich aufgrund der vorliegenden Evidenz nicht eindeutig feststellen, obwohl die Tatsache, dass Off-Szenario-Schüler/innen schwierigere Aufgaben als die Szenario-Gruppe zu lösen bekamen, dabei bestimmt eine Rolle gespielt haben dürfte. In weiteren, ähnlichen Untersuchungen wäre es daher wichtig, für den Vergleich der Szenario- und Off-Szenario-Aufgaben nicht nur eine höhere Anzahl Aufgaben vorzusehen, sondern auch 1) Items mit einem vergleichbaren Itemschwierigkeitswert und 2) Aufgaben mit den

gleichen Textsorten in den Inputtexten, mit der gleichen Anzahl Items und mit dem gleichen Itemtyp auszuwählen.

In dieser Studie konnten spannende Einblicke in die Testlösungsprozesse und Ansichten der Schüler/innen gewonnen werden, die die quantitativen Daten des Projekts IFB besser zu interpretieren und einzuordnen helfen (vgl. Barras et al., 2020; Karges, Lenz et al., 2021). Die hier erhobenen Verbalprotokolle fielen sehr umfassend aus und konnten im Rahmen der vorliegenden Dissertation nur teilweise und vor allem im Hinblick auf die eingangs formulierten Fragestellungen ausgewertet werden. Allerdings stehen alle 20 codierten und 30 uncodierten Transkripte interessierten Forschenden im Forschungsdatenarchiv des KFM zur Verfügung und können auch im Hinblick auf weitere Fragestellungen untersucht werden, auf die hier nicht eingegangen werden konnte. Folgende Fragestellungen wären zum Beispiel möglich:

- Welche Strategienketten bzw. Strategiencluster lassen sich in den Herangehensweisen der Schüler/innen an die Testaufgaben erkennen?
- Lassen sich aufgrund der vorliegenden Evidenz bestimmte Typen der Herangehensweisen an die Lösung der Testaufgaben bilden?
- Wie nehmen die Schüler/innen die Schwierigkeit der Aufgaben wahr?
- Was sagen die Schüler/innen zur Usability der computerbasierten Sprachtestaufgaben?
- Welche subjektiven Theorien haben die Schüler/innen in Bezug auf die Sprachentests, den Sprachunterricht oder Französisch als Fremdsprache?
- Wie gern lernen die Schüler/innen Französisch als Fremdsprache (im Gegensatz zu Englisch als Fremdsprache) und wie nehmen sie ihren gegenwärtigen Französischunterricht wahr?

Der vorliegenden Studie kann ausserdem entnommen werden, wie lohnenswert es ist, den Meinungen der Fremdsprachenlernenden zuzuhören. Es wäre daher sehr zu begrüßen, wenn es solche Studien vermehrt nicht nur im Kontext der Validierung von Sprachtests, sondern allgemein im Kontext der Bemühungen um die Optimierung des Fremdsprachenunterrichts geben würde. Die introspektiven Verfahren als Datenerhebungsmethoden und die qualitative Inhaltsanalyse als Datenanalysemethode haben sich in dieser Studie als geeignet erwiesen und können (natürlich auch in Kombination mit weiteren Methoden) für weitere ähnliche Untersuchungen empfohlen werden.

7 LITERATURVERZEICHNIS

- Adolf, S. M., Perfetti, C. A., & Catts, H. W. (2011). Developmental changes in reading comprehension: Implications for assessment and instruction. In S. J. Samuels & A. E. Farstrup (Hrsg.), *What research has to say about reading instruction* (S. 186-214). International Reading Association.
- Aguado, K. (2013). Die Qualitative Inhaltsanalyse in der empirischen Fremdsprachenforschung: Grenzen, Potentiale, Desiderata. In K. Aguado, L. Heine, & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (S. 119–135). Peter Lang.
- Aguado, K. (2018). Lautes Denken als Datenerhebungsverfahren in der empirischen Fremdsprachenforschung. In K. Aguado, C. Finkbeiner, & B. Tesch (Hrsg.), *Lautes Denken, „Stimulated Recall“ und Dokumentarische Methode: Rekonstruktive Verfahren in der Fremdsprachenlehr- und -lernforschung* (S. 9–25). Peter Lang.
- Aguado, K., Heine, L., & Schramm, K. (Hrsg.). (2013). *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung*. Peter Lang.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
- Alderson, J. C., Haapakangas, E.-L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language*. Routledge.
- Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9(2), 101–119.
- ALTE (Hrsg.). (1998). *Multilingual glossary of language testing terms*. Cambridge University Press.
- American Educational Research Association (Hrsg.). (2011). *Report and recommendations for the reauthorization of the institute of education sciences*. American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Hrsg.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460–472.
- Antón, M. (2012). Dynamic assessment. In G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing* (S. 106–119). Routledge.
- Arras, U. (2007). *Wie beurteilen wir Leistung in der Fremdsprache? Strategien und Prozesse bei der Beurteilung schriftlicher Leistungen in der Fremdsprache am Beispiel der Prüfung „Test Deutsch als Fremdsprache (TestDaF)“*. Narr Francke Attempto.

- Arras, U. (2013). Introspektive Verfahren in der Sprachtestforschung. In K. Aguado, L. Heine, & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (S. 74–91). Peter Lang.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Banerjee, J. (2004). *Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Section D: Qualitative analysis methods*. Council of Europe, Language Policy Division. <http://www.coe.int/t/dg4/linguistic/CEF-ref-supp-SectionD.pdf>
- Barras, M. (2018). „Stimulated Recall“ in der Sprachtestforschung. Ein praktisches Beispiel aus der Erprobung eines computerbasierten Leseverstehenstests. In K. Aguado, C. Finkbeiner, & B. Tesch (Hrsg.), *Lautes Denken, „Stimulated Recall“ und Dokumentarische Methode. Rekonstruktive Verfahren in der Fremdsprachenlehr- und -lehrforschung* (S. 69–86). Peter Lang.
- Barras, M., Karges, K., & Lenz, P. (2016). Leseverstehen überprüfen: Welche Sprache für die Fragen und Antworten in den Testitems? *Babylonia*, 16(2), 13–18.
- Barras, M., Karges, K., & Lenz, P. (2020). „Also es ist zuerst eine Motivation und dann irgendwie kommt dir in den Sinn, dass es ein Test ist“—Zur Wirkung digitaler Textsorten in einem computerbasierten Fremdsprachentest. In M. Eisenmann & J. Steinbock (Hrsg.), *Sprache, Kulturen, Identitäten: Umbrüche durch Digitalisierung Dokumentation zum 28. Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung Würzburg 2019*. (S. 177–188). Schneider Hohengehren.
- Bax, S. (2013). *Readers' cognitive processes during IELTS reading tests: Evidence from eye tracking* [ELT Research Papers 13]. British Council. <http://oro.open.ac.uk/47504/1/Bax%20ELTRA%20report.pdf>
- Biggs, J. B. (1984). Learning strategies, student motivation patterns, and subjectively perceived success. In J. R. Kirby (Hrsg.), *Cognitive strategies and educational performance* (S. 111–134). Academic Press.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical theories of mental test scores* (S. 397–479). Addison-Wesley.
- Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge.

- Brunfaut, T. (2016a). Assessing listening. In D. Tsagari & J. Banerjee (Hrsg.), *Handbook of second language assessment* (S. 97–112). De Gruyter Mouton.
- Brunfaut, T. (2016b). *Looking into reading II: A follow-up study on test-takers' cognitive processes while completing APTIS B1 reading task*. The British Council.
https://www.britishcouncil.org/sites/default/files/brunfaut_final_with_hyperlinks_3.pdf
- Brunfaut, T., & McCray, G. (2015). *Looking into test-takers' cognitive processes whilst completing reading tasks: A mixed-method eye-tracking and stimulated recall study*. The British Council.
http://eprints.lancs.ac.uk/72005/1/Brunfaut_and_McCray_final_report_FINAL.pdf
- Buck, G. (2001). *Assessing listening*. Cambridge University Press.
- Buck, G. (2018). Preface. In G. J. Ockey & E. Wagner (Hrsg.), *Assessing L2 listening: Moving towards authenticity* (S. XI–XVI). John Benjamins.
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15(2), 119–157.
- Burwitz-Melzer, E., & Steininger, I. (2016). Inhaltsanalyse. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 256–269). Narr.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Hrsg.), *Language and communication* (S. 2–14). Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, (1), 1–47.
- Caspari, D. (2016). Grundlagen fremdsprachendidaktischer Forschung. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 7–21). Narr.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254–272.
- Chapelle, C. A. (2012). Conceptions of validity. In G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing* (S. 21–33). Routledge.
- Chapelle, C. A., & Voss, E. (2014). Evaluation of language tests through validation research. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume III: Evaluation, methodology, and interdisciplinary themes* (S. 1081–1097). Wiley-Blackwell.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, 6(3), 271–315.

- Cohen, A. D. (1997). Towards enhancing verbal reports as a source of insights on test-taking strategies. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Hrsg.), *Current developments and alternatives in language assessment. Proceedings of LTRC 96* (S. 339–366). University of Jyväskylä.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. Longman.
- Cohen, A. D. (2006). The coming of age for research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307–331.
- Cohen, A. D. (2007a). Coming to terms with language learner strategies: Surveying the experts. In A. D. Cohen & E. Macaro (Hrsg.), *Language learner strategies: Thirty years of research and practice* (S. 29–45). Oxford University Press.
- Cohen, A. D. (2007b). The coming of age for research on test-taking strategies. In J. D. Fox, M. Welsche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Hrsg.), *Language testing reconsidered* (S. 89–111). University of Ottawa Press.
- Cohen, A. D. (2011a). Second language learner strategies. In E. Hinkel (Hrsg.), *Handbook of research in second language teaching and learning. Vol. 2* (S. 681–698). Routledge.
- Cohen, A. D. (2011b). *Strategies in learning and using a second language* (2nd edition). Longman.
- Cohen, A. D. (2012). Test-taking strategies and task design. In G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing* (S. 262–277). Routledge.
- Cohen, A. D. (2014). Using test-wiseness strategy research in task development. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume II: Approaches and development* (S. 893–905). Wiley-Blackwell.
- Cohen, A. D., & Upton, T. A. (2006). Strategies in responding to the new TOEFL reading tasks. *ETS Research Report Series, 2006* (1), i–162. <https://doi.org/10.1002/j.2333-8504.2006.tb02012.x>
- Cohen, A. D., & Upton, T. A. (2007). “I want to go back to the text”: Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209–250.
- Cohen, A. D., & Wang, I. K.-H. (2018). Fluctuation in the functions of language learner strategies. *System*, 74, 169–182.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th edition). Routledge.
- Council of Europe (Hrsg.). (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Creswell, J. W., & Plano Clark, V. L. (2006). *Designing and conducting mixed methods research*. Sage.

- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2007). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Hrsg.), *Handbook of mixed methods in social & behavioral research* (S. 209–240). Sage.
- Cumming, A. (2012). Validation of language assessments. In C. A. Chapelle (Hrsg.), *The encyclopedia of applied linguistics*. Blackwell Publishing Ltd.
<https://doi.org/10.1002/9781405198431.wbeal1242>
- D-EDK. (2016). *Lehrplan 21: Sprachen. Von der D-EDK Plenarversammlung am 1.10.2014 zur Einführung in den Kantonen freigegebene Vorlage. Bereinigte Fassung vom 29.02.2016*. https://v-ef.lehrplan.ch/container/V_EF_DE_Fachbereich_SPR.pdf
- Demirkaya, S. (2014). Analyse qualitativer Daten. In S. Demirkaya, A. Feldmeier, N. Gültekin-Karakoç, & C. Riemer (Hrsg.), *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache: Eine Einführung* (S. 213–227). UTB GmbH.
- Diekmann, A. (2008). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. Rowohlt Taschenbuch.
- DIPF, & Nagarro. (2018a, 6. November). *CBA Execution Environment*. Nagarro.
<https://tba.dipf.de/de/infrastruktur/softwareentwicklung/cba-execution-environment>
- DIPF, & Nagarro. (2018b, 6. November). *CBA ItemBuilder*. Nagarro.
<https://tba.dipf.de/de/infrastruktur/softwareentwicklung/cba-item-builder/cba-itembuilder>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge, Taylor & Francis Group.
- Eberharter, K., Kremmel, B., & Zehentner, M. (2018). Die Erstellung von Testaufgaben: Der Testentwicklungszyklus. In B. Hinger & W. Stadler (Hrsg.), *Testen und Bewerten fremdsprachlicher Kompetenzen* (S. 57–68). Narr Francke Attempto.
- EDK (Hrsg.). (2011). *Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards*.
http://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf
- EDK. (2019, 14. Oktober). *Schulmodell(e) Sekundarstufe I*.
<https://www.edk.ch/de/bildungssystem/kantonale-schulorganisation/kantonsumfrage/a-4-schulmodell-e-auf-der-sekundarstufe-i>
- EDK. (2021, 23. Mai). *ÜGK Schweiz: Informationen für Eltern, Schülerinnen und Schüler, Lehrpersonen, Schulleitende und Forschende*. <http://uegk-schweiz.ch/>
- educa. (2019, 14. Oktober). *Strukturmodelle auf Sekundarstufe I*. bildungssystem.educa.ch.
<https://bildungssystem.educa.ch/de/strukturmodelle-sekundarstufe-i>
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

- Ercikan, K., & Pellegrino, J. W. (Hrsg.). (2017a). *Validation of score meaning for the next generation of assessments: The use of response processes*. Routledge.
- Ercikan, K., & Pellegrino, J. W. (2017b). Validation of score meaning using examinee response processes for the next generation of assessments. In K. Ercikan & J. W. Pellegrino (Hrsg.), *Validation of score meaning for the next generation of assessments: The use of response processes* (S. 1–8). Routledge.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215–251.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. MIT Press.
- ETS. (2019, 8. August). ETS. <https://www.ets.org/>
- ETS Research. (2019, 8. August). *Reading for understanding*. https://www.ets.org/research/topics/reading_for_understanding/
- Europarat (Hrsg.). (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Langenscheidt.
- Færch, C., & Kasper, G. (Hrsg.). (1987). *Introspection in second language research*. Multilingual Matters.
- Feick, D. (2013). „Sehen Sie sich Ines an.“ Zur sozialen Situiertheit des Videobasierten Lauten Erinnerns. In K. Aguado, L. Heine, & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (S. 54–73). Peter Lang.
- Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Hrsg.), *Examining listening: Research and practice in assessing second language listening* (S. 77–151). Cambridge University Press.
- Field, J. (2019). *Rethinking the second language listening test: From theory to practice*. Equinox.
- Finkbeiner, C., & Schluer, J. (2018). „Stimulated Recall“ als Zugang zu mentalen Prozessen in der Fremdsprachenforschung. In K. Aguado, C. Finkbeiner, & B. Tesch (Hrsg.), *Lautes Denken, „Stimulated Recall“ und Dokumentarische Methode: Rekonstruktive Verfahren in der Fremdsprachenlehr- und -lernforschung* (S. 27–49). Peter Lang.
- Fox, J. (2014). Large-scale assessment. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume II: Approaches and development* (S. 533–541). Wiley-Blackwell.
- Freedle, R. O., & Kostin, I. W. (1993). *The prediction of TOEFL reading comprehension item difficulty for expository prose passages for three item types: Main idea, inference, and supporting idea items* (TOEFL Research Report RR-44). ETS. http://www.ets.org/research/policy_research_reports/rr-93-13

- Freedle, R. O., & Kostin, I. W. (1996). *The prediction of TOEFL listening comprehension item difficulty for Minitalk passages: Implications for construct validity* (TOEFL Research Report RR-56). ETS.
http://www.ets.org/research/policy_research_reports/rr-96-29
- Galaczi, E. D. (2014). Content analysis. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume III: Evaluation, methodology, and interdisciplinary themes* (S. 1325–1339). Wiley-Blackwell.
- Gass, S. M., & Mackey, A. (2011). *Data elicitation for second and foreign language research*. Routledge Taylor & Francis Group.
- Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research* (Second edition). Routledge.
- Geranpayeh, A., & Taylor, L. (2013). *Examining listening: Research and practice in assessing second language listening*. Cambridge University Press.
- Grabe, W. (2000). Reading research and its implications for reading assessment. In A. J. Kunnan (Hrsg.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (S. 226–262). Cambridge University Press.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Grabe, W., & Jiang, X. (2014). Assessing reading. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume I: Abilities, context and learners* (S. 185–200). Wiley-Blackwell.
- Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd edition). Routledge.
- Green, Alison (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge University Press.
- Green, Anthony (2014). *Exploring language assessment and testing: Language in action*. Routledge.
- Green, R. (2017). *Designing listening tests: A practical approach*. Palgrave Macmillan.
- Griffiths, C. (2018). *The strategy factor in successful language learning: The tornado effect* (2nd edition). Multilingual Matters.
- Grotjahn, R. (2012). Hörverstehen: Konstrukt und Messung. *Fremdsprachen Lehren und Lernen*, 41(1).
- Gu, L., & So, Y. (2017). Strategies used by young English learners in an assessment context. In M. K. Wolf & Y. G. Butler (Hrsg.), *English language proficiency assessments for young learners* (S. 118–134). Taylor & Francis.

- Gu, Y., Hu, G., & Zhang, L. J. (2005). Investigating language learner strategies among lower primary school pupils in Singapore. *Language and Education*, 19(4), 281–303.
- Gu, Y. (2012). Learning strategies: Prototypical core and dimensions of variation. *Studies in Self-Access Learning Journal*, 3(4), 330–356.
- Gu, Y. (2014). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System*, 43, 74–81.
- Harsch, C. (2016). Testen. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 205–218). Narr.
- Harsch, C., & Schröder, K. (2007). Textrekonstruktion: C-Test. In B. Beck & Klieme Eckhard (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 212–225). Beltz Verlag.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 143–171). Springer.
- Hasselgreen, A., & Caudwell, G. (2016). *Assessing the language of young learners*. Equinox.
- Heine, L. (2005). Lautes Denken als Forschungsinstrument in der Fremdsprachenforschung. *Zeitschrift für Fremdsprachenforschung*, 16(2), 163–185.
- Heine, L. (2014). Introspektion. In S. Demirkaya, A. Feldmeier, N. Gültekin-Karakoç, & C. Riemer (Hrsg.), *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache: Eine Einführung* (S. 123–135). UTB GmbH.
- Heine, L., & Schramm, K. (2007). Lautes Denken in der Fremdsprachenforschung: Eine Handreichung für die empirische Praxis. In H. J. Vollmer (Hrsg.), *Synergieeffekte in der Fremdsprachenforschung: Empirische Zugänge, Probleme, Ergebnisse* (S. 167–206). Peter Lang.
- Heine, L., & Schramm, K. (2016). Introspektion. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 173–181). Narr.
- Holliday, A. (2015). Qualitative research and analysis. In B. Paltridge & A. Phakiti (Hrsg.), *Research methods in applied linguistics: A practical resource* (S. 49–62). Bloomsbury Academic.
- Holzknicht, F. (2019). *Double play in listening assessment*. Lancaster University.
<https://doi.org/10.17635/LANCASTER/THESIS/812>
- Holzknicht, F., Eberharter, K., Kremmel, B., McCray, G., Konrad, E., & Spöttl, C. (2017). *Looking into listening: Using eye-tracking to establish the cognitive validity of the Aptis Listening Test* [ARAGs Research Reports Online]. British Council.
https://www.britishcouncil.org/sites/default/files/looking_into_listening.pdf

- Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Hughes, A. (2003). *Testing for language teachers* (2nd edition). Cambridge University Press.
- In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219–244.
- Jamieson, J., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Frameworks for a new TOEFL. In C. A. Chapelle, M. Enright, & J. Jamieson (Hrsg.), *Building a validity argument for the Test of English as a Foreign Language* (S. 55–95). Routledge.
- Karges, K., Barras, M., & Lenz, P. (i. V.). Assessing young language learners' receptive skills: Should we ask the questions in the language of schooling? In S. Frisch, E. Romeik, & J. Rymarczyk (Hrsg.), *Current research into young FL and EAL learners' literacy skills*. Peter Lang.
- Karges, K., Barras, M., & Lenz, P. (2021). *Leseverstehen in einer Fremdsprache testen. Task Lab – eine empirische Studie zu computerbasierten Aufgaben in Französisch*. Kompetenzzentrum für Mehrsprachigkeit.
- Karges, K., Lenz, P., Aeppli, T., & Barras, M. (2021). *Fremdsprachenkompetenzen testen nahe an der Realität. Szenariobasierte Testaufgaben für den Computer – eine Vertiefungsstudie*. Kompetenzzentrum für Mehrsprachigkeit.
- Kenyon, D. M., & MacGregor, D. (2012). Pre-operational testing. In G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing* (S. 295–306). Routledge.
- KFM. (2020a, 8. Dezember). *Innovative Formen der Beurteilung: Vertiefungsstudie zur kompetenzorientierten Leistungsbeurteilung im rezeptiven Bereich*. <http://www.institut-plurilinguisme.ch/de/content/innovative-formen-der-beurteilung>
- KFM. (2020b, 15. Dezember). *Task Lab: Untersuchungen zum besseren Verständnis und zur Erhöhung der Validität von kommunikativen Testaufgaben*. <http://www.institut-mehrsprachigkeit.ch/de/content/task-lab-untersuchungen-zum-besseren-verstaendnis-und-zur-erhoehung-der-validitaet-von>
- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Kirsch, I. S. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured* (RR-01-25). ETS. <https://www.ets.org/Media/Research/pdf/RR-01-25-Kirsch.pdf>

- Knorr, P. (2013). Zur Differenzierung retrospektiver verbaler Daten: Protokolle Lautes Erinnerns erheben, verstehen und analysieren. In K. Aguado, L. Heine, & K. Schramm (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung* (S. 31–53). Peter Lang.
- Knorr, P., & Schramm, K. (2012). Datenerhebung durch Lautes Denken und Lautes Erinnern in der fremdsprachendidaktischen Empirie. Grundlagenbeitrag. In S. Doff (Hrsg.), *Fremdsprachenunterricht empirisch erforschen. Grundlagen – Methoden – Anwendung* (S. 184–201). Narr Francke Attempto.
- Knorr, P., & Schramm, K. (2016). Triangulation. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 90–97). Narr.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.
- Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning*, 57(1), 1–44.
- Koda, K. (2012). Assessment of reading. In C. A. Chapelle (Hrsg.), *The encyclopedia of applied linguistics* (S. 226–233). Wiley Blackwell.
- Kong, J. (2019). *Investigating the role of test methods in testing reading comprehension: A process-focused perspective*. Springer.
- Kostin, I. W. (2004). *Exploring item characteristics that are related to the difficulty of TOEFL dialogue items* (Research Report RR-04-11). ETS.
<https://www.ets.org/Media/Research/pdf/RR-04-11.pdf>
- Kuckartz, U. (1999). *Computergestützte Analyse qualitativer Daten: Eine Einführung in Methoden und Arbeitstechniken*. Westdeutscher Verlag.
- Kuckartz, U. (2012). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Beltz Juventa.
- Kuckartz, U. (2014). *Mixed Methods*. Springer.
- Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (4. Auflage). Beltz Juventa.
- Kuckartz, U., & Rädiker, S. (2019). *Analyse qualitativer Daten mit MAXQDA: Text, Audio und Video*. Springer.
- Kunnan, A. J. (2008). Large scale language assessments. In E. Shohamy & N. H. Hornberger (Hrsg.), *Encyclopedia of language and education. Volume 7: Language testing and assessment* (2nd edition, S. 135–155). Springer.
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.

- Lazaraton, A. (2017). Qualitative methods of validation. In E. Shohamy, I. G. Or, & S. May (Hrsg.), *Language Testing and Assessment* (S. 211–224). Springer.
- Lazaraton, A., & Taylor, L. (2007). Qualitative research methods in language test development and validation. In J. D. Fox, M. Welsche, D. Bayliss, L. Cheng, C. E. Turner, & C. Doe (Hrsg.), *Language testing reconsidered* (S. 113–129). University of Ottawa Press.
- Lenz, P., Karges, K., & Barras, M. (2019). Investigating test method effects in French L2 reading items for young learners. In A. Huhta, G. Erickson, & N. Figueras (Hrsg.), *Developments in language education: A memorial volume in honour of Sauli Takala* (S. 182–202). University of Jyväskylä & EALTA.
- Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*, 26(1), 35–57.
- Leung, C. (2013). Qualitative research in language assessment. In C. A. Chapelle (Hrsg.), *The Encyclopedia of applied linguistics*. Blackwell Publishing Ltd.
<https://doi.org/10.1002/9781405198431.wbeal0979>
- Lumley, T., & Brown, A. (2005). Research methods in language testing. In E. Hinkel (Hrsg.), *Handbook of research in second language teaching and learning* (S. 833–855). Lawrence Erlbaum Associates Inc.
- Mayring, P. (1983). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz.
- Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12., aktualisierte Neuauflage). Beltz.
- Messick, S. (1989). Validity. In R. L. Linn (Hrsg.), *Educational measurement* (S. 13–104). American Council on Education and Macmillan.
- Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82.
- Nikolov, M. (2006). Test-taking strategies of 12- and 13-year-old hungarian learners of EFL: Why whales have migraines. *Language Learning*, 56(1), 1–51.
- Nold, G., & Rossa, H. (2007). Hörverstehen. In B. Beck & Klieme Eckhard (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)* (S. 178–196). Beltz Verlag.
- Nold, G., Rossa, H., & Hartig, J. (2008). Proficiency scaling in DESI listening and reading EFL tests: Task characteristics, item difficulty and cut-off points. In L. Taylor & C. J. Weir (Hrsg.), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity—proceedings of the ALTE Berlin conference may 2005* (S. 94–116). Cambridge University Press.
- OBS Studio. (2017). [Windows]. OBS Studio Contributors. <https://obsproject.com/>

- Ockey, G. J. (2012). Assessment of listening. In C. A. Chapelle (Hrsg.), *The encyclopedia of applied linguistics* (S. 212–218). Blackwell Publishing Ltd.
- Ockey, G. J., & Wagner, E. (Hrsg.). (2018a). *Assessing L2 listening: Moving towards authenticity*. John Benjamins.
- Ockey, G. J., & Wagner, E. (2018b). Introduction. In G. J. Ockey & E. Wagner (Hrsg.), *Assessing L2 listening: Moving towards authenticity* (S. 1–10). John Benjamins.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. PISA, OECD Publishing.
- O'Malley, J. M., Chamot, A. U., & Küpper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418–437.
- O'Malley, M. J., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.
- O'Reilly, T., & Sabatini, J. P. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (Research Report RR-13-31). ETS.
https://www.ets.org/research/policy_research_reports/publications/report/2013/jrmi
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Heinle.
- Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd revised edition). Routledge.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 67(3), 215–227.
- Passepartout: Fremdsprachen an der Volksschule. (2015). *Lehrplan Französisch und Englisch*. https://be.lehrplan.ch/passepartout/Lehrplan_Passepartout.pdf
- Perfetti, C. (1999). Comprehending written language: A blueprint of the reader. In C. M. Brown & P. Hagoort (Hrsg.), *The neurocognition of language* (S. 167–208). Oxford University Press.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11(4), 357–383.
- Peyer, E., Andexlinger, M., Kofler, K., & Lenz, P. (2016). *Projekt Fremdsprachenevaluation BKZ. Schlussbericht zu den Sprachkompetenztests*. Institut für Mehrsprachigkeit.
http://www.institut-mehrsprachigkeit.ch/sites/default/files/documents_projets/fremdsprachenevaluation_bkz_schlussbericht_sprachkompetenztests_2016-01-26_2.pdf
- Phakiti, A. (2007). *Strategic competence and EFL reading test performance: A structural equation modeling approach*. Peter Lang.

- Phakiti, A. (2016). Test takers' performance appraisals, appraisal calibration, and cognitive and metacognitive strategy use. *Language Assessment Quarterly*, 13(2), 75–108.
- Porsch, R., Grotjahn, R., & Tesch, B. (2010). Hörverstehen und Hör-Sehverstehen in der Fremdsprache – Unterschiedliche Konstrukte? In *Zeitschrift für Fremdsprachenforschung* (Bd. 21, Nr. 2, S. 143–189).
- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge University Press.
- Purpura, J. E. (2014). Cognition and language assessment. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume III: Evaluation, methodology, and interdisciplinary themes* (S. 1325–1339). Wiley-Blackwell.
- R Core Team. (2014). *R: A language and environment for statistical computing* (3.1.2) [En; Windows 7 64bit]. R Foundation for Statistical Computing. <http://www.R-project.org>
- Reed, D. J. (2014). Field testing of test items and tasks. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume II: Approaches and development* (S. 876–892). Wiley-Blackwell.
- Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35(1), 31–65.
- Robitzsch, A. (2015). *Essays zu methodischen Herausforderungen im Large-Scale Assessment*. Humboldt-Universität zu Berlin. <https://d-nb.info/1082283517/34>
- Rose, H. (2012). Reconceptualizing strategic learning in the face of self-regulation: Throwing language learning strategies out with the bathwater. *Applied Linguistics*, 33(1), 92–98.
- Rossa, H. (2012). *Mentale Prozesse beim Hörverstehen in der Fremdsprache: Eine Studie zur Validität der Messung sprachlicher Kompetenzen*. Peter Lang.
- Rost, M. (2011). *Teaching and researching listening*. Pearson/Longman.
- Rubin, J. (1975). What the “good language learner” can teach us. *TESOL Quarterly*, 9(1), 41–51.
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441–474.
- Sabatini, J. P., O'Reilly, T., & Deane, P. (2013). *Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design* (Research Report RR-13-30). ETS. https://www.ets.org/research/policy_research_reports/publications/report/2013/jrmh

- Sabatini, J. P., O'Reilly, T., Halderman, L., & Bruce, K. (2014a). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Année Psychologique*, 114(4), 693–723.
- Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014b). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice*, 29(1), 36–43.
- Sasaki, M. (2014). Introspective methods. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume III: Evaluation, methodology, and interdisciplinary themes* (S. 1340–1357). Wiley-Blackwell.
- Schmidt, T., Schütte, W., & Winterscheid, J. (2015). *CGAT. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2)*. <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/4616>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212–226.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95(1), 26–43.
- Schramm, K. (2001). *L2-Leser in Aktion: Der fremdsprachliche Leseprozess als mentales Handeln*. Waxmann.
- Schramm, K. (2016). Empirische Forschung. In D. Caspari, F. Klippel, M. K. Legutke, & K. Schramm (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch* (S. 49–58). Narr.
- Seale, C. (1999). *The quality of qualitative research*. Sage.
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günther, S., Hartung, M., Kern, F., Merzluff, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schütte, W., Stukenbrock, A., Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 10, 354–402.
- Shafer, N. (2018). *Varietäten und Varianten verstehen lernen: Zum Umgang mit Standardvariation in Deutsch als Fremdsprache*. Universitätsverlag Göttingen.
- Shiotsu, T. (2010). *Components of L2 reading: Linguistic and processing factors in the reading test performances of Japanese EFL learners*. Cambridge University Press.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147–170.

- Shore, J. R., Wolf, M. K., O'Reilly, T., & Sabatini, J. P. (2017). Measuring 21-st century reading comprehension through scenario-based assessments. In M. K. Wolf & Y. G. Butler (Hrsg.), *English language proficiency assessments for young learners* (S. 234–252). Taylor & Francis.
- Stadler, W. (2018). Rezeptive Fertigkeiten überprüfen und bewerten. In B. Hinger & W. Stadler (Hrsg.), *Testen und Bewerten fremdsprachlicher Kompetenzen* (S. 69–86). Narr Francke Attempto.
- Stadler, W., & Kremmel, B. (2018). Testprinzipien. In B. Hinger & W. Stadler (Hrsg.), *Testen und Bewerten fremdsprachlicher Kompetenzen* (S. 39–55). Narr Francke Attempto.
- Stern, H. H. (1975). What can we learn from the good language learner? *Canadian Modern Language Review*, 34(4), 304–319.
- Stevenson, M. (2015). Researching reading. In B. Paltridge & A. Phakiti (Hrsg.), *Research methods in applied linguistics: A practical resource* (S. 315–334). Bloomsbury Academic.
- Studer, T. (2010). Kompetenzmodelle und Bildungsstandards für Deutsch als Fremd- und Deutsch als Zweitsprache. In Krumm, Hans-Jürgen, Fandrych, Christian, Hufeisen, Britta, & Riemer, Claudia (Hrsg.), *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch* (Bd. 2, S. 1264–1271). De Gruyter Mouton.
- Studer, T. (2016). Nützliche(re) Sprachtests. Beispiele für die formative Beurteilungs-Praxis vor dem Hintergrund einer neueren Perspektive auf Test-Gütekriterien. *Babylonia*, 2, 36–43.
- Suvorov, R. (2018). *Investigating test-taking strategies during the completion of computer-delivered items from Michigan English Test (MET): Evidence from eye tracking and cued retrospective reporting*. *Cambridge Michigan Language Assessment (CaMLA) Working Papers 2018–02*. <https://michiganassessment.org/wp-content/uploads/2019/03/CWP-2018-02.pdf>
- Taylor, L. (2005). Using qualitative research methods in test development and validation. *Research Notes*, 21, 2–4.
- Taylor, L. (2013). Introduction. In A. Geranpayeh & L. Taylor (Hrsg.), *Examining listening: Research and practice in assessing second language listening* (S. 1–35). Cambridge University Press.
- Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product, and practice*. Routledge.
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(03), 191–210.

- Vandergrift, L. (2011). Second language listening: Presage, process, product, and pedagogy. In E. Hinkel (Hrsg.), *Handbook of research in second language teaching and learning*. Vol. 2 (S. 455–471). Routledge.
- Vandergrift, L. (2015). Researching listening. In B. Paltridge & A. Phakiti (Hrsg.), *Research methods in applied linguistics: A practical resource* (S. 299–313). Bloomsbury Academic.
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis: L2 listening comprehension. *Language Learning*, 65(2), 390–416.
- Vandergrift, L., & Baker, S. C. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *The Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes*, 74(1), 79–100.
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- VERBI Software. (2017). *MAXQDA 2018*. VERBI GmbH. <https://www.maxqda.de>
- VERBI Software. (2019). *MAXQDA 2020*. VERBI GmbH. <https://www.maxqda.de>
- Vollmer, H. J. (2010). Kompetenzforschung in den Fremdsprachendidaktiken: Ein Überblick. In K. Aguado, K. Schramm, & H. J. Vollmer (Hrsg.), *Fremdsprachliches Handeln beobachten, messen, evaluieren* (S. 29–64). Peter Lang.
- Wagner, E. (2014). Assessing listening. In A. J. Kunnan (Hrsg.), *The companion to language assessment. Volume I: Abilities, context and learners* (S. 47–63). Wiley-Blackwell.
- Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen—Eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Beltz.
- Wenden, A. (1991). *Learner strategies for learner autonomy: Planning and implementing learner training for language learners*. Prentice Hall.
- Wisniewski, K. (2016). Die strategische Kompetenz in der Fremdsprache – ein Forschungsüberblick. In S. Brunetti, J. Klingebiel-Schieke, C. M. Pedron, M.-C. Piotrowski, A. Ruggieri, & R. Schreiber (Hrsg.), *Versprachlichung von Welt – Il mondo in parole. Festschrift zum 60. Geburtstag von Maria Lieber* (S. 637–658). Stauffenburg.
- Xi, X., & Sawaki, Y. (2017). Methods of test validation. In E. Shohamy, I. G. Or, & S. May (Hrsg.), *Language Testing and Assessment* (S. 193–209). Springer.

Zum Anhang dieser Arbeit gehören:

- 1) Infobrief an die Erziehungsberechtigten der 30 Schüler/innen aus der qualitativen Stichprobe inkl. Antworttalon.
- 2) Transkripte und Screenshots der analysierten Szenario- und Off-Szenario-Aufgaben.
- 3) Transkriptionsrichtlinien für die Verbalprotokolle (GAT 2).
- 4) Kategorienleitfaden.

Im Forschungsdatenarchiv des KFM¹³¹ sind ausserdem folgende Dateien zu finden:

- 1) 20 codierte Transkripte
 - MAXQDA_codierte_20_Transkripte.mx20
 - Alison_Szenario.pdf
 - Aria_Szenario.pdf
 - Arya_Szenario.pdf
 - Boebu_Szenario.pdf
 - Bonnie_Off.pdf
 - Cersei_Szenario.pdf
 - Cyberman_Szenario.pdf
 - Dustin_Off.pdf
 - Elena_Szenario.pdf
 - Eleven_Off.pdf
 - Ezra_Off.pdf
 - Howard_Szenario.pdf
 - Loup_Off.pdf
 - Mogli_Off.pdf
 - Nancy_Off.pdf
 - Omega_Off.pdf
 - Picard_Off.pdf
 - Sheldon_Szenario.pdf
 - Steve_Off.pdf
 - Toby_Szenario.pdf
 - Uebersicht_Memos.docx

¹³¹ <https://portailplurilingue.unifr.ch/starweb/KFM/k.skca-catalog/servlet.starweb?path=KFM/k.skca-catalog/FastLink.web&inum=000012367>

- Uebersicht_Memos.pdf
- Zusammenstellung_der_codierten_Transkriptpassagen.docx
- Zusammenstellung_der_codierten_Transkriptpassagen.pdf

2) 30 uncodierte Transkripte (Projekt IFB)

- MAXQDA_alle_30_Transkripte_uncodiert.mx20
- Alison_Szenario.docx
- Amy_Szenario.docx
- Aria_Szenario.docx
- Arya_Szenario.docx
- Billy_Szenario.docx
- Boebu_Szenario.docx
- Bonnie_Off.docx
- Cersei_Szenario.docx
- Chiara_Off_Teil_1.docx
- Chiara_Off_Teil_2.docx
- Cybermen_Szenario.docx
- Dustin_Off.docx
- Elena_Szenario.docx
- Eleven_Off.docx
- Ezra_Off.docx
- Hector_Off.docx
- Howard_Szenario.docx
- Jon_Off.docx
- Loup_Off.docx
- Matthew_Szenario.docx
- Mogli_Off.docx
- Nancy_Off.docx
- Omega_Off.docx
- Picard_Off.docx
- Sheldon_Szenario.docx
- Sofia_Szenario.docx
- Spencer_Szenario.docx
- Steve_Off.docx
- Toby_Szenario.docx
- Tyler_Szenario.docx

- Willy_Szenario.docx
- 3) Infobrief an die Erziehungsberechtigten der 30 Schüler/innen
- Infobrief_IFB_Zusatzerhebung.pdf
- 4) Tasks, die in der qualitativen Studie zum Einsatz kamen
- Off_Szenario_Screenshots_alle_Tasks.pdf
 - Off-Szenario_analysierte_Tasks.pdf
 - Szenario_Stadttrip_analysierte_Tasks.pdf
 - Szenario_Stadttrip_Screenshots_alle_Tasks.pdf
- 5) Transkriptionsrichtlinien
- Transkriptionsrichtlinien_GAT2_Schnelluebersicht.pdf
- 6) Kategorienleitfaden
- Kategorienleitfaden.pdf
- 7) Tabellen (stellen die Häufigkeiten der Codierungen dar und wurden mithilfe von MAXQDA (VERBI Software, 2019) angefertigt)
- Auflistung_Strategien_mit_Gewichtungen.xlsx
 - Code-Relations-Browser_AlleCodes_mit_allen_Codes.xlsx
 - Haeufigkeiten_Strategien_Pro_SuS_mit_korrigierten_Gewichtungen.xlsx
 - Strategien_Szenario_vs_Off-Szenario.xlsx
- 8) Grafiken (grafische Darstellungen der codierten Daten, die aufgrund der Tabellen bzw. aufgrund der Daten aus der quantitativen IFB-Studie erstellt wurden)
- IFB_quali_2pl_wles.pdf
 - time_spent_pro_task.pdf
 - top7_L_strats_per_part.pdf
 - top7_strats_per_part.pdf
 - top7_strats_per_task_off.pdf
 - top7_strats_per_task_szenario.pdf
 - top7_TM_strats_per_part.pdf
 - top7_TW_strats_per_part.pdf

8.1 ANHANG 1: INFOBRIEF AN DIE ERZIEHUNGSBERECHTIGTEN

CSP **Center scientific da cumpetenza per la plurilinguitad** Cogniziun Società Formation Bildung Migration Furmaziun Gesellschaft
CSP **Centro scientifico di competenza per il plurilinguismo** Scuola Arbeit Politique Communitad School Travail Ecole Community
CSP **Centre scientifique de compétence sur le plurilinguisme** Migrazione Furmaziun Societad Cognition Society scola Migration
KFM **Wissenschaftliches Kompetenzzentrum für Mehrsprachigkeit** Societè Cognizione Migraziun Schule Communauté Kognition
RCM **Research Centre on Multilingualism** Formazione Lavoro Politics Comunità Work Politik Lavur Politica Formation Gemeinschaft

Liebe Eltern

Im Rahmen des Forschungsprojekts „Innovative Formen der Beurteilung“ führt das Kompetenzzentrum für Mehrsprachigkeit in Fribourg eine Studie zum Funktionieren und zur Qualität computerbasierter Testaufgaben in den Fremdsprachen durch. Dazu haben Sie bereits einen Elternbrief erhalten.

Neben der Erhebung mit der ganzen Klasse an Computern arbeiten wir mit zufällig ausgewählten Schülerinnen und Schülern parallel dazu auch einzeln zusammen. Wir tun dies, um zu bestimmten neuen Aufgaben noch genauere Informationen zu bekommen. Für die Schülerinnen und Schüler sollte dies nicht belastend sein: Sie lösen Aufgaben und geben dazu detailliert Auskunft über ihre Lösungswege. Sie geben uns auch eine Rückmeldung zur Qualität und Schwierigkeit der Aufgaben. **Dabei geht es nicht um eine Bewertung Ihres Kindes** – im Gegenteil: Sie geben *uns* wichtige Hinweise zum Funktionieren und zu Schwächen der Aufgaben.

Während der Einzelarbeit entstehen Video- und Tonaufnahmen, ohne die wir die Informationen aus der Zusammenarbeit mit den Kindern nicht festhalten könnten. Sie werden ausschliesslich für Forschungszwecke genutzt und jederzeit vertraulich behandelt. Sie werden der allgemeinen Öffentlichkeit nicht zugänglich gemacht. Die Anonymität der Teilnehmenden wird bei dieser Untersuchung generell gewahrt. Beispielsweise wird bereits während der Einzelarbeit nicht mit dem echten Namen, sondern mit einem selbst gewählten Spitznamen gearbeitet.

Die Zusammenarbeit mit den Schülerinnen und Schülern ist für den Erfolg des Projekts sehr wichtig. Wir wären Ihnen deshalb sehr dankbar, wenn Sie Ihrem Kind die Teilnahme erlauben würden. Aufgrund unserer bisherigen Erfahrungen wissen wir, dass Schülerinnen und Schüler gerne an solchen Projekten teilnehmen. Wir bitten Sie, den untenstehenden Talon auszufüllen und bis zum angegebenen Termin wieder mit in die Schule zu geben.

Vielen Dank für Ihre Unterstützung!

Mit freundlichen Grüssen



Malgorzata Barras



Peter Lenz (Projektleiter)

Falls Sie **Fragen** haben, wenden Sie sich bitte an:

Malgorzata Barras (malgorzata.barras@unifr.ch)

Katharina Karges (katharina.karges@unifr.ch)

Antworttalon

Bitte bis zum an die Schule zurück.

Vorname und Name des Kindes:

- ☐ Wir sind damit einverstanden, dass unser Kind bei der Einzelarbeit im Rahmen der Aufgabenerprobung mitmacht.
- ☐ Wir sind mit der Teilnahme unseres Kindes nicht einverstanden.

Ort, Datum:

Name des/der Erziehungsberechtigten:

Unterschrift des/der Erziehungsberechtigten:

8.2 ANHANG 2: ANALYSIERTE SZENARIO-AUFGABEN

Szenario „Stadttrip“ (Originalbezeichnung Projekt IFB: Szenario 6): Transkripte

Szenarioeinbettung

	Französisch
Situierung	Du bist während 3 Wochen für einen Sprachkurs in Frankreich. Nächstes Wochenende möchtest du die Stadt Avignon besuchen.

Task A

Handlungsziel: Einen schriftlichen Dialog mit Empfehlungen verstehen

		Französisch	
	Situierung	<p>Du hast Bekannte in Avignon: Marc, Evan und Anne.</p> <p>In einem Gruppenchat geben sie dir Tipps, was man in Avignon unternehmen kann.</p> <p>Deine Aufgabe: Du liest den Chat. Auf den nächsten Seiten sammelst du die Tipps deiner Bekannten.</p>	
Seite 1	Aufgabe	Lies den Anfang des Chats.	
	Frage/Stimulus	<p>Ziehe die Informationen in die passende Kategorie.</p> <p>Die Reihenfolge ist egal.</p>	
	Inputtext	<p>Original</p> <p>Toi: Salut! Notre prof nous a parlé du Château d'Avignon aujourd'hui. C'est cool?</p> <p>Evan: Le château est un classique et un des monuments les plus visités. Mais franchement, je n'y irais pas.</p> <p>Anne: Non, c'est trop ennuyeux! Par contre, Avignon a un nouveau stade de sport. Ne fais pas la visite guidée, ça coûte très cher, mais il y a un musée du rugby juste à côté.</p> <p>Marc: Oh oui ! Tu dois absolument visiter ça. C'est super.</p> <p>Toi: Magnifique. C'est long à visiter?</p>	<p>Deutsche Übersetzung</p> <p>Du: Hallo! Unser Lehrer hat uns heute vom Schloss Avignon erzählt. Ist das cool?</p> <p>Evan: Das Schloss ist ein Klassiker und eine der am häufigsten besuchten Sehenswürdigkeiten. Aber ehrlich gesagt würde ich dort nicht hingehen.</p> <p>Anne: Nein, es ist zu langweilig! Avignon hat aber ein neues Sportstadion. Du solltest die geführte Tour nicht machen, weil das sehr teuer ist, aber es gibt ein Rugby-Museum direkt daneben.</p> <p>Marc: Oh ja! Du musst das unbedingt besuchen. Es ist super.</p> <p>Du: Schön. Dauert der Besuch lange?</p>

		<p>Anne: Non, une heure ou deux. Après, tu peux prendre un super déjeuner au Coq Rouge, le meilleur resto d'Avignon.</p> <p>Evan: Oui! Tu dois absolument goûter leurs burgers!</p>	<p>Anne: Nein, eine oder zwei Stunden. Danach könnt ihr im Coq Rouge zu Mittag essen. Das ist das beste Restaurant der Stadt.</p> <p>Evan: Ja! Du musst unbedingt ihre Burger probieren!</p>
	Antwortoptionen	<p>A1w: Das Schloss anschauen (nein) A1x: Die Stadiontour machen (nein) A1y: Die Rugby-Ausstellung besuchen (ja) A1z: Im ‚Coq Rouge‘ essen (ja)</p>	
Seite 2	Aufgabe	Lies die Fortsetzung des Chats.	
	Frage/Stimulus	Ziehe die Informationen in die passende Kategorie. Die Reihenfolge ist egal.	
	Inputtext	<p>Original Marc: Pour avoir des entrées réduites dans les musées et le bus gratuit, il y a la Carte Avignon. Elle coûte 20€. Tu la trouves à l'office du tourisme.</p> <p>Anne: La carte n'est pas valable sur les bateaux, mais c'est pas grave. Les tours en bateau sont ennuyeux.</p> <p>Evan: Et il y a cette application pour smartphone qui te guide à travers la vieille ville. Je ne sais plus comment elle s'appelle...</p> <p>Marc: Oui, c'est l'appli Avignon. Et Evan a raison, ça vaut la peine de l'utiliser. L'appli contient plusieurs tours de la ville avec des infos historiques.</p> <p>Toi: D'accord, je vais faire ça. Autre chose?</p> <p>Marc: Sinon, ne monte surtout pas dans les bus pour touristes. Tu ne verras rien de plus qu'à pied.</p>	<p>Deutsche Übersetzung Marc: Für einen ermässigten Eintritt in die Museen und kostenloses Busfahren gibt es die Carte Avignon. Sie kostet 20€. Du findest sie im Tourismusbüro.</p> <p>Anne: Die Karte ist auf Booten nicht gültig, aber das ist nicht schlimm. Bootstouren sind langweilig.</p> <p>Evan: Und es gibt diese Smartphone-App, die dich durch die Altstadt führt. Ich weiss nicht mehr, wie sie heisst...</p> <p>Marc: Ja, die AppliAvignon. Und Evan hat Recht, es lohnt sich, sie zu nutzen. Die App enthält mehrere Stadtführungen mit historischen Informationen.</p> <p>Sie: Okay, das werde ich tun. Sonst noch etwas?</p> <p>Marc: Steige ansonsten auf keinen Fall in die Touristenbusse ein. Du wirst nicht mehr sehen als zu Fuss.</p>
	Antwortoptionen	<p>A2w: Den Sightseeing-Bus nehmen (nein) A2x: Die Carte Avignon kaufen (ja) A2y: Einen Stadtrundgang machen (ja) A2z: Eine Bootsfahrt machen. (nein)</p>	

Task B

Handlungsziel: Einen längeren narrativen Text in groben Zügen verstehen

		Französisch	
Situierung		<p>Du hast im Internet den Audioblog von Lou entdeckt. Lou ist eine Touristin aus Québec. Sie war in Avignon und hat in ihrem Blog davon erzählt.</p> <p>Deine Aufgabe: Du hörst ihren Blog und beantwortest dazu einige Fragen.</p>	
Seite 1	Aufgabe	Höre den ersten Teil von Lous Podcast. Beantworte dann die Frage.	
	Frage/Stimulus	Lou war für drei Tage nicht im Internet. Warum?	
	Inputtext	<p>Original Salut à tous! Je suis enfin de retour sur mon blog. Je n'ai pas eu de connexion internet pendant trois jours, quand j'étais à Avignon. Ce matin, je suis arrivée dans un petit village appelé Saintes-Maries-de-la-Mer. C'est au bord de la mer Méditerranée, au Sud d'Avignon. Je suis loin de la grande ville, mais ici, j'ai enfin une connexion internet rapide à l'hôtel ! A Avignon, à l'hôtel, il n'y avait rien: la réceptionniste a juste haussé les épaules, et m'a dit que leur connexion internet était cassée et qu'elle ne serait pas réparée avant la semaine suivante. Malheureusement, je n'avais pas non plus internet sur mon smartphone, parce que je n'ai pas acheté de carte SIM française quand je suis arrivée à Paris. Je ne peux pas utiliser internet avec ma carte SIM canadienne, ça coûterait beaucoup trop cher.</p>	<p>Deutsche Übersetzung Hallo zusammen! Endlich bin ich zurück auf meinem Blog. Ich hatte drei Tage lang keine Internetverbindung als ich in Avignon war. Heute Morgen bin ich in einem kleinen Dorf namens Saintes-Maries-de-la-Mer angekommen. Es liegt am Mittelmeer, südlich von Avignon. Ich bin weit weg von der Grossstadt, aber hier habe ich endlich eine schnelle Internetverbindung im Hotel! In Avignon, im Hotel, gab es nichts: Die Rezeptionistin hat nur mit den Schultern gezuckt und mir gesagt, dass ihre Internetverbindung unterbrochen ist und dass sie erst in der folgenden Woche repariert wird. Leider hatte ich auch kein Internet auf meinem Smartphone, weil ich bei meiner Ankunft in Paris keine französische SIM-Karte gekauft hatte. Mit meiner kanadischen SIM-Karte kann ich das Internet nicht nutzen, das würde viel zu viel kosten.</p>
	Antwortoptionen	<p>B1a: Sie in einem kleinen Dorf ohne Internetempfang. B1b: Ihre neue französische SIM-Karte war noch nicht aktiviert. B1c: In ihrem Hotel in Avignon funktionierte das Internet nicht.</p>	
Seite 2	Aufgabe	Höre den zweiten Teil von Lous Podcast. Beantworte dann die Frage.	
	Frage/Stimulus	Was hat Lou in Avignon besonders beeindruckt?	

	Inputtext	<p>Original</p> <p>Donc, quand je suis arrivée à Avignon, j'ai vu le château par la fenêtre du bus. Il est sur une colline, et je pense que la vue de là-haut doit être fantastique. J'ai aussi vu de jolies maisons anciennes au centre-ville. Mais je n'ai rien vu d'autre: mes trois jours à Avignon ont été trois jours d'enfer! Pendant les deux premiers jours, il a plu du matin au soir, sans arrêt. Je ne voyais même pas les bâtiments de l'autre côté de la rue. Je pense que je n'ai jamais vu une pluie pareille ! Le troisième jour, il a fait beau, mais je n'ai pas pu sortir parce que j'étais malade. J'avais mangé des fruits de mer le deuxième jour, et ils ne devaient pas être bons. Je pense que je ne vais plus jamais manger de fruits de mer ! C'est la troisième fois que ça me rend malade.</p>	<p>Deutsche Übersetzung</p> <p>Also, als ich in Avignon ankam, habe ich das Schloss durch das Fenster des Busses gesehen. Es liegt auf einem Hügel, und ich denke, die Aussicht von dort oben muss fantastisch sein. Ich habe auch einige schöne alte Häuser im Stadtzentrum gesehen. Aber ich habe nichts anderes gesehen: Meine drei Tage in Avignon waren drei Höllentage! Während der ersten beiden Tage hat es von morgens bis abends ununterbrochen geregnet. Ich konnte nicht einmal die Gebäude auf der anderen Strassenseite sehen. Ich glaube, so einen Regen habe ich noch nie gesehen! Am dritten Tag war es schön, aber ich konnte nicht rausgehen, weil ich krank war. Ich hatte am zweiten Tag Meeresfrüchte gegessen, und die müssen nicht gut gewesen sein. Ich glaube, dass ich nie wieder Meeresfrüchte essen werde! Das ist das dritte Mal, dass es mich krank macht.</p>
	Antwortoptionen	<p>B2a: Das gute Essen am zweiten Tag. B2b: Der starke Regen in der Stadt. B2c: Die schöne Aussicht vom Schloss.</p>	

Task C

Handlungsziel: Aus Suchresultaten/Kurzzusammenfassungen eines aufgrund von Kriterien auswählen

		Französisch	
Situierung		<p>Da du über das Wochenende in Avignon sein wirst, informierst du dich in einem Forum über verschiedene günstige Hotels.</p> <p>In dem Forum haben Gäste darüber geschrieben, wie ihnen die Hotels gefallen haben.</p> <p>Deine Aufgabe: Du suchst in den Bewertungen Informationen zu den Hotels.</p>	
Seite 1	Aufgabe	<p>Lies die Bewertungen über das Hôtel Sainte-Louise.</p> <p>Du hast dir für jedes Hotel drei Fragen gestellt. Beantworte die Frage mithilfe der Bewertungen. Klicke an! In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.</p> <p>Manchmal findest du in der gleichen Bewertung mehrere Antworten.</p>	
	Frage/Stimulus	<p>Gibt es im Zimmer gratis WLAN? Ist das Hotel im Stadtzentrum? Gibt es im Hotel Zimmer unter 35 € pro Nacht?</p>	
	Inputtext	<table><tr><td><p>Original Hôtel Sainte-Louise</p><p>1) Avis de Thérèse: Nous avons passé une semaine à l'Hôtel Avignon, et nous avons passé un bon séjour. Les chambres sont jolies, propres et bien décorées. Le Wi-Fi est gratuit à la réception, mais comme nous voulions une connexion dans notre chambre, nous avons payé 15€ de plus pour la semaine.</p><p>2) Avis de Miguel: J'ai passé une nuit ici avec un ami, après un concert au centre-ville d'Avignon, en novembre. Nous avons payé environ 30€ pour la nuit, ce qui n'est pas beaucoup. La chambre était spacieuse, calme et</p></td><td><p>Deutsche Übersetzung Hotel Sainte-Louise</p><p>1) Thérèses Meinung: Wir haben eine Woche im Hotel Avignon verbracht und der Aufenthalt war gut. Die Zimmer sind schön, sauber und gut dekoriert. Wi-Fi ist an der Rezeption kostenlos, aber da wir eine Verbindung in unserem Zimmer haben wollten, haben wir 15€ mehr für die Woche bezahlt.</p><p>2) Miguels Meinung: Ich habe hier mit einem Freund übernachtet, nach einem Konzert im Zentrum von Avignon, im November. Wir haben etwa 30€ für die Nacht bezahlt.</p></td></tr></table>	<p>Original Hôtel Sainte-Louise</p> <p>1) Avis de Thérèse: Nous avons passé une semaine à l'Hôtel Avignon, et nous avons passé un bon séjour. Les chambres sont jolies, propres et bien décorées. Le Wi-Fi est gratuit à la réception, mais comme nous voulions une connexion dans notre chambre, nous avons payé 15€ de plus pour la semaine.</p> <p>2) Avis de Miguel: J'ai passé une nuit ici avec un ami, après un concert au centre-ville d'Avignon, en novembre. Nous avons payé environ 30€ pour la nuit, ce qui n'est pas beaucoup. La chambre était spacieuse, calme et</p>
<p>Original Hôtel Sainte-Louise</p> <p>1) Avis de Thérèse: Nous avons passé une semaine à l'Hôtel Avignon, et nous avons passé un bon séjour. Les chambres sont jolies, propres et bien décorées. Le Wi-Fi est gratuit à la réception, mais comme nous voulions une connexion dans notre chambre, nous avons payé 15€ de plus pour la semaine.</p> <p>2) Avis de Miguel: J'ai passé une nuit ici avec un ami, après un concert au centre-ville d'Avignon, en novembre. Nous avons payé environ 30€ pour la nuit, ce qui n'est pas beaucoup. La chambre était spacieuse, calme et</p>	<p>Deutsche Übersetzung Hotel Sainte-Louise</p> <p>1) Thérèses Meinung: Wir haben eine Woche im Hotel Avignon verbracht und der Aufenthalt war gut. Die Zimmer sind schön, sauber und gut dekoriert. Wi-Fi ist an der Rezeption kostenlos, aber da wir eine Verbindung in unserem Zimmer haben wollten, haben wir 15€ mehr für die Woche bezahlt.</p> <p>2) Miguels Meinung: Ich habe hier mit einem Freund übernachtet, nach einem Konzert im Zentrum von Avignon, im November. Wir haben etwa 30€ für die Nacht bezahlt.</p>		

Seite 2		<p>propre. En résumé, de belles chambres pour un petit prix.</p> <p>3) Avis de Curt: L'hôtel offre une belle vue sur le Rhône et il se trouve en plein cœur d'Avignon. Mais quand je suis arrivé dans la chambre, la fenêtre était ouverte et il faisait vraiment froid ! En général, l'hôtel n'était pas mal, il a une connexion Internet, mais je n'y retournerais pas.</p>	<p>was nicht viel ist. Das Zimmer war geräumig, ruhig und sauber.</p> <p>Zusammengefasst: schöne Zimmer zu einem kleinen Preis.</p> <p>3) Curts Meinung: Das Hotel bietet einen schönen Blick auf die Rhone und liegt im Herzen von Avignon. Aber als ich in den Raum kam, war das Fenster offen, und es war wirklich kalt! Im Allgemeinen war das Hotel nicht schlecht, es hat einen Internetanschluss, aber ich würde nicht noch einmal dorthin zurückgehen.</p>
	Antwortoptionen	<p>WiFi: nein (1) Zentrale Lage: ja (3) Preiswert: ja (2)</p>	
	Aufgabe	<p>Lies die Bewertungen über das Hôtel du Sud.</p> <p>Du hast dir für jedes Hotel drei Fragen gestellt. Beantworte die Frage mithilfe der Bewertungen. Klicke an! In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.</p> <p>Manchmal findest du in der gleichen Bewertung mehrere Antworten.</p>	
	Frage/Stimulus	<p>Gibt es im Zimmer gratis WLAN? Ist das Hotel im Stadtzentrum? Gibt es im Hotel Zimmer unter 35 € pro Nacht?</p>	
	Inputtext	<p>Original Hôtel du Sud</p> <p>1) Avis de Lyle: Joli endroit avec des chambres grandes et très propres. Le personnel était très serviable et le petit déjeuner est l'un des meilleurs que j'aie jamais mangé dans un hôtel. Par contre, il n'est pas inclus dans le prix.</p> <p>2) Avis de Roberta: La réceptionniste était très</p>	<p>Deutsche Übersetzung Hôtel du Sud</p> <p>1) Lyles Meinung: Schöner Standort mit grossen und sehr sauberen Räumen. Das Personal war sehr hilfsbereit und das Frühstück ist eines der besten, das ich je in einem Hotel gegessen habe. Es ist jedoch nicht im Preis inbegriffen.</p> <p>2) Robertas Meinung: Die Rezeptionistin</p>

		<p>gentille et serviable. La chambre était assez grande, propre, et il y avait une connexion internet sans fil. Le seul problème a été le repas au restaurant de l'hôtel - des légumes trop cuits et des pommes de terre pas assez cuites, et un repas pour deux coûte plus de 35€.</p> <p>3) Avis de Kelly: L'année dernière, on était dans un hôtel à côté de la Gare Centrale ... qui n'est pas du tout centrale! L'hôtel du Sud est situé juste à côté de la mairie, on a fait tout le centre à pied. En plus, c'est un hôtel bon marché où les prix commencent à 25€ par nuit.</p>	<p>war sehr nett und hilfsbereit. Der Raum war ziemlich gross, sauber, und es gab eine drahtlose Internetverbindung. Das einzige Problem war das Essen im Hotelrestaurant - zu stark gekochtes Gemüse und zu wenig gekochte Kartoffeln, und ein Menü für zwei Personen kostet mehr als 35€.</p> <p>3) Kellys Meinung: Letztes Jahr waren wir in einem Hotel neben dem Hauptbahnhof... der überhaupt nicht zentral ist! Das Hôtel du Sud befindet sich direkt neben dem Rathaus, wir sind durch das ganze Zentrum zu Fuss gegangen. Ausserdem ist es ein günstiges Hotel, wo die Preise bei 25€ pro Nacht beginnen.</p>
	Antwortoptionen	<p>WiFi: ja (2) Zentrale Lage: ja (3) Preiswert: ja (3)</p>	

Screenshots Szenario „Stadttrip“

Task A

Einbettung



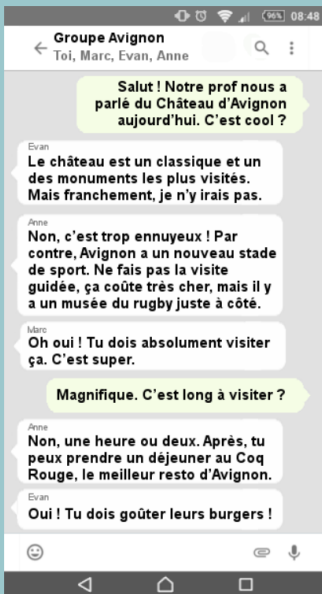
Du hast Bekannte in Avignon: Marc, Evan und Anne.

In einem Gruppenchat geben sie dir Tipps, was man in Avignon unternehmen kann.

Deine Aufgabe: Du liest den Chat. Auf den nächsten Seiten sammelst du die Tipps deiner Bekannten.

Weiter

Seite 1



Lies den Anfang des Chats.

Ziehe die Informationen in die passende Kategorie.
Die Reihenfolge ist egal.

✓ Das empfehlen die drei:	✗ Das empfehlen die drei <u>nicht</u> :
?	?
?	?

Das Schloss anschauen

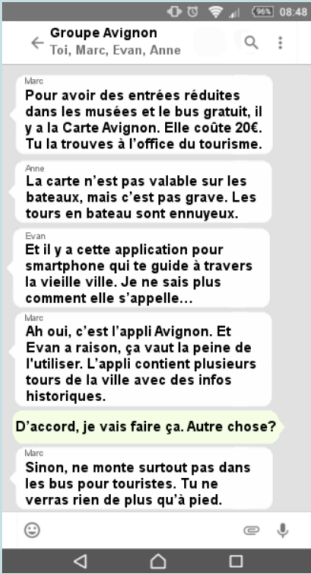
Die Stadiontour machen

Die Rugby-Ausstellung besuchen

Im ‚Coq Rouge‘ essen

Weiter

Seite 2



Lies die Fortsetzung des Chats.

Ziehe die Informationen in die passende Kategorie.
Die Reihenfolge ist egal.

✓ Das empfehlen die drei:	✗ Das empfehlen die drei <u>nicht</u> :
Die Rugby-Ausstellung besuchen	Die Stadiontour machen
Im 'Coq Rouge' essen	Das Schloss anschauen
?	?
?	?

Den Sightseeing-Bus nehmen

Die Carte Avignon kaufen

Einen Stadtrundgang machen

Eine Bootsfahrt machen.

Weiter

Task B

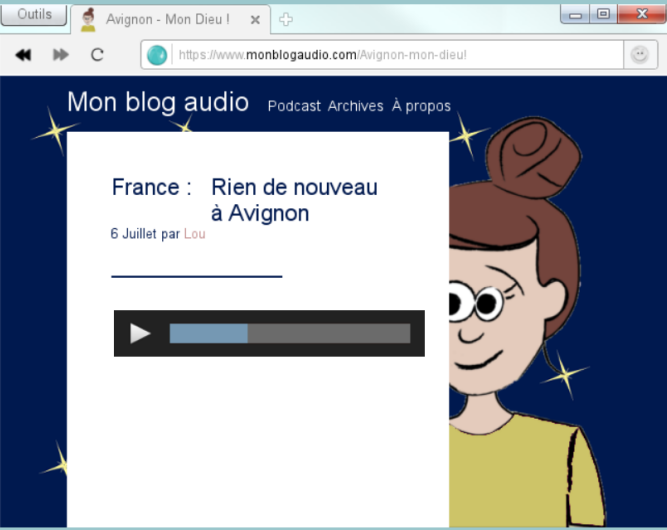
Einbettung

Du hast im Internet den Audioblog von Lou entdeckt. Lou ist eine Touristin aus Québec. Sie war in Avignon und hat in ihrem Blog davon erzählt.

Deine Aufgabe: Du hörst ihren Blog und beantwortest dazu einige Fragen.

Weiter

Seite 1



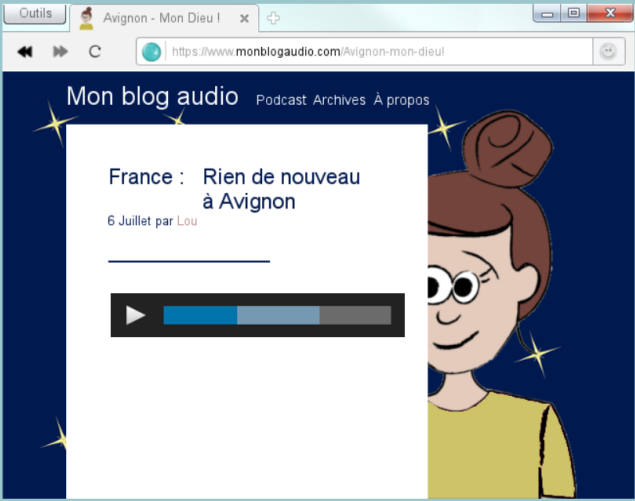
Höre den ersten Teil von Lous Podcast. Beantworte dann die Frage.

Lou war für drei Tage nicht im Internet. Warum?

- Sie war in einem kleinen Dorf ohne Internetempfang.
- Ihre neue französische SIM-Karte war noch nicht aktiviert.
- In ihrem Hotel in Avignon funktionierte das Internet nicht.

Weiter

Seite 2



Höre den zweiten Teil von Lous Podcast. Beantworte dann die Frage.

Was hat Lou in Avignon besonders beeindruckt?

- Das gute Essen am zweiten Tag.
- Der starke Regen in der Stadt.
- Die schöne Aussicht vom Schloss.

Weiter

Task C

Einbettung

Da du über das Wochenende in Avignon sein wirst, informierst du dich in einem Forum über verschiedene günstige Hotels.

In dem Forum haben Gäste darüber geschrieben, wie ihnen die Hotels gefallen haben.

Deine Aufgabe: Du suchst in den Bewertungen Informationen zu den Hotels.

Weiter

Seite 1

The screenshot shows a web browser window with the address bar displaying 'https://www.hôtels.fr/Avignon/Hôtel-Sainte-Louise'. The page title is 'H Hôtels à Avignon'. The main content area shows reviews for 'Hôtel Sainte-Louise'. There are three reviews visible:

- Avis de Thérèse:** Nous avons passé une semaine à l'Hôtel Avignon, et nous avons passé un bon séjour. Les chambres sont jolies, propres et bien décorées. Le Wi-Fi est gratuit à la réception, mais comme nous voulions une connexion dans notre chambre, nous avons payé 15€ de plus pour la semaine.
- Avis de Miguel:** J'ai passé une nuit ici avec un ami, après un concert au centre-ville d'Avignon, en Novembre. Nous avons payé environ 30€ pour la nuit, ce qui n'est pas beaucoup. La chambre était spacieuse, calme et propre. En résumé, de belles chambres pour un petit prix.
- Avis de Curt:** L'hôtel offre une belle vue sur le Rhône et il se trouve en plein cœur d'Avignon. Mais quand je suis arrivé dans la chambre, la fenêtre était ouverte et il faisait vraiment froid ! En général, l'hôtel n'était pas mal, il a une connexion Internet, mais je n'y retournerais pas.

Lies die Bewertungen über das **Hôtel Sainte-Louise**.

Du hast dir für jedes Hotel drei Fragen gestellt.

1. Beantworte die Frage mithilfe der Bewertungen. Klicke an!
2. In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du in der gleichen Bewertung mehrere Antworten.

1. Gibt es im Zimmer gratis WLAN?

☐ ja ☒ ja ☐ nein ☒ nein

Avis de ?

2. Ist das Hotel im Stadtzentrum?

☐ ja ☒ ja ☐ nein ☒ nein

Avis de ?

3. Gibt es im Hotel Zimmer unter 35€ pro Nacht?

☐ ja ☒ ja ☐ nein ☒ nein

Avis de ?

Weiter

The screenshot shows a web browser window with the URL <https://www.hotels.fr/Avignon/Hotel-du-Sud>. The page title is 'S Hôtel du Sud'. There are three reviews listed:

- Avis de Lyle:** Joli endroit avec des chambres grandes et très propres. Le personnel était très serviable et le petit déjeuner est l'un des meilleurs que j'aie jamais mangés dans un hôtel. Par contre, il n'est pas inclus dans le prix.
- Avis de Roberta:** La réceptionniste était très gentille et serviable. La chambre était assez grande, propre, et il y avait une connexion internet sans fil. Le seul problème a été le repas au restaurant de l'hôtel - des légumes trop cuits et des pommes de terre pas assez cuites, et un repas pour deux coûte plus de 35€.
- Avis de Kelly:** L'année dernière, on était dans un hôtel à côté de la Gare Centrale ... qui n'est pas du tout centrale! L'hôtel du Sud est situé juste à côté de la mairie, on a fait tout le centre à pied. En plus, c'est un hôtel bon marché où les prix commencent à 25€ par nuit.

Lies die Bewertungen über das **Hôtel du Sud**.

Du hast dir für jedes Hotel drei Fragen gestellt.

1. Beantworte die Frage mithilfe der Bewertungen. Klicke an!
2. In welcher Bewertung hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du in der gleichen Bewertung mehrere Antworten.

1. Gibt es im Zimmer gratis WLAN?

☐ ja ✓
 ☐ nein ✗

Avis de ?

2. Ist das Hotel im Stadtzentrum?

☐ ja ✓
 ☐ nein ✗

Avis de ?

3. Gibt es im Hotel Zimmer unter 35€ pro Nacht?

☐ ja ✓
 ☐ nein ✗

Avis de ?

Weiter

8.3 ANHANG 3: ANALYSIERTE OFF-SZENARIO-AUFGABEN

Task 1: Walkies (Originalbezeichnung Projekt IFB: Szenario 3, Task C): Transkripte

Handlungsziel: Einen schriftlichen Dialog mit Anweisungen verstehen

		Französisch	
Einleitung		In einem Gruppenchat diskutierst du mit Schulkollegen über eine Gruppenarbeit. Ihr müsst einen Prospekt erstellen und verteilt die Aufgaben dafür.	
Seite 1	Aufgabe	Beantworte die Frage. Ziehe die Nachricht, in der du die Antwort gefunden hast, in das Feld unten.	
	Frage/Stimulus	Was ist deine Aufgabe? In welcher Nachricht hast du die Antwort gefunden?	
	Inputtext	<p>Original</p> <p>Adrien (a): Salut! Je pense qu'il nous faut: un texte très court et une image pour la première page, un texte qui résume l'initiative, et peut-être un tableau qui montre les avantages et les inconvénients des Walkies.</p> <p>Estelle (b): Et il nous faut un slogan! Tous les prospectus ont des slogans.</p> <p>Adrien (c): C'est vrai. Des idées?</p> <p>Estelle (d): Non, pas vraiment.</p> <p>Noémie (e): Commençons par le contenu du prospectus. Avec Estelle, on a discuté des avantages et des inconvénients des Walkies. On peut s'occuper du tableau.</p> <p>Toi (f): Je pense qu'il faut aussi parler de l'inventrice des Walkies. Ça va donner une touche plus personnelle. Je peux</p>	<p>Deutsche Übersetzung</p> <p>Adrien (a): Hallo! Ich glaube, wir brauchen: einen sehr kurzen Text und ein Bild für die erste Seite, einen Text, der die Initiative zusammenfasst, und vielleicht eine Tabelle, die die Vor- und Nachteile von Walkies aufzeigt.</p> <p>Estelle (b): Und wir brauchen einen Slogan! Alle Prospekte haben Slogans.</p> <p>Adrien (c): Das stimmt. Irgendwelche Ideen?</p> <p>Estelle (d): Nein, nicht wirklich.</p> <p>Noémie (e): Beginnen wir mit dem Inhalt des Prospekts. Estelle und ich haben über die Vor- und Nachteile von Walkies diskutiert. Wir können uns um die Tabelle kümmern.</p> <p>Du (f): Ich denke, wir sollten auch über die Erfinderin der Walkies sprechen. Das gibt dem Ganzen einen persönlicheren Touch. Ich</p>

		le faire. J'ai vu un documentaire sur elle il n'y a pas longtemps.	kann das tun. Ich habe vor kurzem einen Dokumentarfilm über sie gesehen.
	Antwortoptionen	C1a: Eine Tabelle mit Vor- und Nachteilen zusammenstellen. C1b: Einen Slogan für den Prospekt erfinden. C1c: Einen Text über die Erfinderin der „Walkies“ schreiben. (f)	
Seite 2	Aufgabe	Beantworte die Frage. Ziehe die Nachricht, in der du die Antwort gefunden hast, in das Feld unten.	
	Frage/Stimulus	Was wird Esther machen? In welcher Nachricht hast du die Antwort gefunden?	
	Inputtext	<p>Original</p> <p>Adrien (a): Noémie, est-ce que tu peux aussi écrire le texte sur les Walkies? Je pense que c'est bien de faire le texte et le tableau en même temps.</p> <p>Noémie (b): Oui. J'écris le texte sur les Walkies et Estelle fait le tableau. Ensuite, on met tout ensemble. D'accord?</p> <p>Estelle (c): Ok, on trouve beaucoup d'infos sur internet, de toute façon.</p> <p>Adrien (d): En fait, Estelle, je pense que tu devrais faire ce que tu fais le mieux. Tu vois ce que je veux dire?</p> <p>Estelle (e): Bon... Ok. Comme tu le demandes si gentiment, je m'occupe du design. Noémie, ça va aller toute seule?</p> <p>Noémie (f): Bon, d'accord. Mais je compte sur toi pour la correction!</p>	<p>Deutsche Übersetzung</p> <p>Adrien (a): Noémie, kannst du auch den Text über Walkies schreiben? Ich denke, es ist gut, den Text und die Tabelle gleichzeitig zu machen.</p> <p>Noémie (b): Ja. Ich schreibe den Text über die Walkies und Estelle macht die Tabelle. Dann setzen wir alles zusammen. Einverstanden?</p> <p>Estelle (c): Okay, man findet auf jeden Fall eine Menge Informationen im Internet.</p> <p>Adrien (d): Eigentlich denke ich, dass du, Estelle, das tun solltest, was du am besten kannst. Weisst du, was ich meine?</p> <p>Estelle (e): Na ja... okay. Da du so nett fragst, kümmere ich mich um das Design. Noémie, kommst du allein zurecht?</p> <p>Noémie (f): Na gut, in Ordnung. Aber für die Korrektur zähle ich auf dich!</p>

		Adrien (g): Si tu veux, je peux relire tes textes. Estelle, tu as assez de travail avec le design.	Adrien (g): Wenn du willst, kann ich deine Texte noch einmal durchlesen. Estelle, du hast mit dem Design genug zu tun.
	Antwortoptionen	C2a: Den Prospekt grafisch gestalten. (e) C2b: Die Texte der anderen Korrektur lesen. C2c: Im Internet nach Informationen suchen.	

Task 2: Stadttrip (Originalbezeichnung Projekt IFB: Szenario 6, Task B)

Handlungsziel: Einen längeren narrativen Text in groben Zügen verstehen

		Französisch	
	Einleitung	Du hörst den ersten Teil eines Podcasts. Darin berichtet die Touristin Lou von ihrer Reise nach Avignon.	
Seite 1	Aufgabe	Beantworte die Frage.	
	Frage/Stimulus	Lou war für drei Tage nicht im Internet. Warum?	
	Inputtext	Original Salut à tous! Je suis enfin de retour sur mon blog. Je n'ai pas eu de connexion internet pendant trois jours, quand j'étais à Avignon. Ce matin, je suis arrivée dans un petit village appelé Saintes-Maries-de-la-Mer. C'est au bord de la mer Méditerranée, au Sud d'Avignon. Je suis loin de la grande ville, mais ici, j'ai enfin une connexion internet rapide à l'hôtel ! A Avignon, à l'hôtel, il n'y avait rien: la réceptionniste a juste haussé les épaules, et m'a dit que leur connexion internet était cassée et qu'elle ne serait pas réparée avant la semaine suivante. Malheureusement, je n'avais pas non plus internet sur mon smartphone, parce que je n'ai pas acheté de carte SIM française quand je suis arrivée à Paris. Je ne peux pas utiliser internet avec ma carte SIM	Deutsche Übersetzung Hallo zusammen! Endlich bin ich zurück auf meinem Blog. Ich hatte drei Tage lang keine Internetverbindung als ich in Avignon war. Heute Morgen bin ich in einem kleinen Dorf namens Saintes-Maries-de-la-Mer angekommen. Es liegt am Mittelmeer, südlich von Avignon. Ich bin weit weg von der Grossstadt, aber hier habe ich endlich eine schnelle Internetverbindung im Hotel! In Avignon, im Hotel, gab es nichts: Die Rezeptionistin hat nur mit den Schultern gezuckt und mir gesagt, dass ihre Internetverbindung unterbrochen ist und dass sie erst in der folgenden Woche repariert wird. Leider hatte ich auch kein Internet auf meinem Smartphone, weil ich bei meiner Ankunft in Paris keine französische SIM-Karte gekauft hatte. Mit meiner kanadischen

Seite 2		canadienne, ça coûterait beaucoup trop cher.	SIM-Karte kann ich das Internet nicht nutzen, das würde viel zu viel kosten.
	Antwortoptionen	B1a: Sie in einem kleinen Dorf ohne Internetempfang. B1b: Ihre neue französische SIM-Karte war noch nicht aktiviert. B1c: In ihrem Hotel in Avignon funktionierte das Internet nicht.	
	Aufgabe	Höre den zweiten Teil von Lous Podcast. Beantworte dann die Frage.	
	Frage/Stimulus	Was hat Lou in Avignon besonders beeindruckt?	
	Inputtext	Original Donc, quand je suis arrivée à Avignon, j'ai vu le château par la fenêtre du bus. Il est sur une colline, et je pense que la vue de là-haut doit être fantastique. J'ai aussi vu de jolies maisons anciennes au centre-ville. Mais je n'ai rien vu d'autre: mes trois jours à Avignon ont été trois jours d'enfer! Pendant les deux premiers jours, il a plu du matin au soir, sans arrêt. Je ne voyais même pas les bâtiments de l'autre côté de la rue. Je pense que je n'ai jamais vu une pluie pareille! Le troisième jour, il a fait beau, mais je n'ai pas pu sortir parce que j'étais malade. J'avais mangé des fruits de mer le deuxième jour, et ils ne devaient pas être bons. Je pense que je ne vais plus jamais manger de fruits de mer! C'est la troisième fois que ça me rend malade.	Deutsche Übersetzung Also, als ich in Avignon ankam, habe ich das Schloss durch das Fenster des Busses gesehen. Es liegt auf einem Hügel, und ich denke, die Aussicht von dort oben muss fantastisch sein. Ich habe auch einige schöne alte Häuser im Stadtzentrum gesehen. Aber ich habe nichts anderes gesehen: Meine drei Tage in Avignon waren drei Höllentage! Während der ersten beiden Tage hat es von morgens bis abends ununterbrochen geregnet. Ich konnte nicht einmal die Gebäude auf der anderen Strassenseite sehen. Ich glaube, so einen Regen habe ich noch nie gesehen! Am dritten Tag war es schön, aber ich konnte nicht rausgehen, weil ich krank war. Ich hatte am zweiten Tag Meeresfrüchte gegessen, und die müssen nicht gut gewesen sein. Ich glaube, dass ich nie wieder Meeresfrüchte essen werde! Das ist das dritte Mal, dass es mich krank macht.
	Antwortoptionen	B2a: Das gute Essen am zweiten Tag. B2b: Der starke Regen in der Stadt. B2c: Die schöne Aussicht vom Schloss.	

Task 3: Berufswahl (Originalbezeichnung Projekt IFB: Szenario 2, Task C)

Handlungsziel: Aus Suchresultaten/Kurzzusammenfassungen eines aufgrund von Kriterien auswählen

Einleitung		Du liest Beschreibungen von jungen Leuten über den Beruf „représentant commercial“. Zu diesem Beruf musst du drei Fragen beantworten.	
Seite 1	Aufgabe	1. Beantworte jede Frage mithilfe der Texte. Klicke an! 2. In welchem Text hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen. Manchmal findest du im gleichen Text mehrere Antworten.	
	Stimulus/Frage	Muss man im Beruf „représentant commercial“...	
	Inputtext	Original C1d: ...contrairement aux vendeurs en magasin, nous devons beaucoup voyager. Je vais régulièrement rencontrer les clients francophones de mon entreprise. Bien sûr, j'essaie aussi de trouver de nouveaux clients pour nos produits. Il y a d'autres représentants commerciaux dans mon équipe qui parlent allemand, italien ou anglais. Chacun cherche des clients qui parlent sa propre langue... C1e: Après que j'aie fini ma formation, j'ai voyagé autour du monde en vélo pendant trois ans. Après ça, j'ai travaillé comme guide touristique au sud de l'Angleterre. C'était très stimulant et j'ai appris beaucoup d'anglais. Maintenant, je suis de retour en France et je travaille comme représentant commercial au centre de Paris. C'est un travail intéressant et je suis bien payé, donc je suis déjà en train de réfléchir à une nouvelle aventure de voyage.	Deutsche Übersetzung C1d: ...im Gegensatz zu den Verkäufern im Geschäft müssen wir sehr viel reisen. Ich treffe mich regelmässig mit den französischsprachigen Kunden meiner Firma. Natürlich versuche ich auch, neue Kunden für unsere Produkte zu finden. In meinem Team gibt es andere Verkaufsvertreter , die Deutsch, Italienisch oder Englisch sprechen. Jeder sucht Kunden, die ihre eigene Sprache sprechen... C1e: Nachdem ich meine Ausbildung abgeschlossen hatte, bin ich drei Jahre lang mit dem Fahrrad um die Welt gereist. Danach habe ich als Reiseleiter im Süden von England gearbeitet. Es war sehr spannend und ich habe viel Englisch gelernt. Jetzt bin ich wieder in Frankreich und arbeite als Verkaufsvertreter im Zentrum von Paris. Es ist ein interessanter Job und ich werde gut bezahlt, so dass ich schon über ein neues Reiseabenteuer nachdenke.

Seite 2		C1f: Quand il y a un nouveau produit, les techniciens doivent le tester. On ne vend rien avant qu'on ne soit pas sûrs que ça marche ! Ensuite, on rencontre les représentants commerciaux pour décider de la stratégie de vente : qui sont les futurs acheteurs, quel langage utiliser pour le décrire. Mon rôle, c'est de...	C1f: Wenn es ein neues Produkt gibt, müssen die Techniker es testen. Wir verkaufen nichts, bevor wir nicht sicher sind, dass es funktioniert! Dann treffen wir uns mit den Verkaufsvertretern , um über die Verkaufsstrategie zu entscheiden: Wer sind die zukünftigen Käufer, in welcher Sprache (=in welchem Stil) soll das Produkt beschrieben werden. Meine Aufgabe ist es...
	Antwortoptionen	C1a: ... selbst neue Produkte testen? (nein, C1f) C1b: ... mehrere Sprachen sprechen? (nein, C1d) C1c: ... viel reisen? (ja, C1d)	
	Stimulus/Frage	Muss man im Beruf „responsable grands comptes“ ...	
	Inputtext	<p>Original</p> <p>C2d: Après l'université, j'ai trouvé un emploi dans une grande compagnie. J'y travaille en tant que responsable grands comptes : je dois trouver les meilleures offres pour nos clients les plus importants. Comme nous sommes une compagnie internationale, je dois parler la langue de mes clients. Je suis donc bien contente d'avoir appris trois langues à l'école. Parfois, j'écris des e-mails à destination de quatre pays différents dans une seule matinée!</p> <p>C2e: ... avant, je ne pensais pas que je deviendrais vendeuse. J'ai fait des études d'ingénieure parce que je voulais créer et tester de nouveaux produits. Finalement, je ne fais pas ça, mais c'est très bien comme ça. Quand nous avons créé notre entreprise, il fallait quelqu'un qui s'occupe des clients importants. Maintenant, je suis responsable grands comptes, et je dois beaucoup voyager en Europe pour mon métier. Cela signifie aussi que...</p>	<p>Deutsche Übersetzung</p> <p>C2d: Nach dem Studium fand ich eine Stelle in einem grossen Unternehmen. Ich arbeite dort als Grosskundenbetreuerin: Ich muss die besten Angebote für unsere wichtigsten Kunden finden. Da wir ein internationales Unternehmen sind, muss ich die Sprache meiner Kunden sprechen. Ich bin also sehr froh, in der Schule drei Sprachen gelernt zu haben. Manchmal schreibe ich an einem Vormittag E-Mails in vier verschiedene Länder!</p> <p>C2e: ... Ich hätte nie gedacht, dass ich einmal Verkäuferin werden würde. Ich habe Ingenieurwesen studiert, weil ich neue Produkte entwickeln und testen wollte. Letztendlich mache ich das nicht, aber das ist auch gut so. Als wir unser Unternehmen gründeten, brauchten wir jemanden, der sich um wichtige Kunden kümmert. Jetzt bin ich Grosskundenbetreuerin und muss für meinen Job viel in</p>

		<p>C2f: Moi, j'aime beaucoup mon travail. En tant que responsable grands comptes, il faut savoir gérer son temps. Le travail doit être terminé à temps, et il n'y a pas de place à l'erreur. Le succès de l'entreprise en dépend. Mais si tu planifie bien ton temps, ce travail laisse aussi beaucoup de temps libre. Moi, je l'utilise pour voyager. Le mois prochain, je vais en Espagne pour tester ma nouvelle planche de surf !</p>	<p>Europa reisen. Das bedeutet auch, dass ...</p> <p>C2f: Ich mag meine Arbeit sehr. Als Grosskundenbetreuer muss man die eigene Zeit gut managen können. Die Arbeit muss pünktlich fertig sein und es gibt keinen Raum für Fehler. Der Erfolg des Unternehmens hängt davon ab. Aber wenn man seine Zeit gut plant, lässt dieser Job auch viel freie Zeit. Ich benutze sie zum Reisen. Nächsten Monat fahre ich nach Spanien, um mein neues Surfbrett zu testen!</p>
	Antwortoptionen	<p>C2a: ... selbst neue Produkte testen? (nein, C2e)</p> <p>C2b: ... mehrere Sprachen sprechen? (ja, C2d)</p> <p>C2c: ... viel reisen? (ja, C2e)</p>	

Screenshots Off-Szenario-Aufgaben

Task 1: Walkies

Seite 1

In einem Gruppenchat diskutierst du mit Schulkollegen über eine Gruppenarbeit. Ihr müsst einen Prospekt erstellen und verteilt die Aufgaben dafür.

Beantworte die Frage. Ziehe die Nachricht, in der du die Antwort gefunden hast, in das Feld unten.

Was ist deine Aufgabe?

- ☐ Eine Tabelle mit Vor- und Nachteilen erstellen.
- ☐ Einen Slogan für den Prospekt erfinden.
- ☐ Einen Text über die Erfinderin der „Walkies“ schreiben.

In welcher Nachricht hast du die Antwort gefunden?

?

Weiter

Seite 2

In einem Gruppenchat diskutierst du mit Schulkollegen über eine Gruppenarbeit. Ihr müsst einen Prospekt erstellen und verteilt die Aufgaben dafür.

Beantworte die Frage. Ziehe die Nachricht, in der du die Antwort gefunden hast, in das Feld unten.

Was wird Estelle machen?

- ☐ Den Prospekt grafisch gestalten.
- ☐ Die Texte der anderen Korrektur lesen.
- ☐ Im Internet nach Informationen suchen.

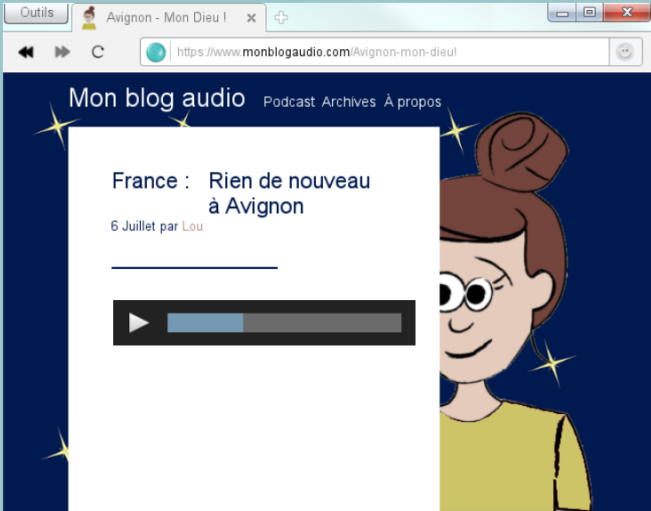
In welcher Nachricht hast du die Antwort gefunden?

?

Weiter

Task 2: Stadtrip

Seite 1



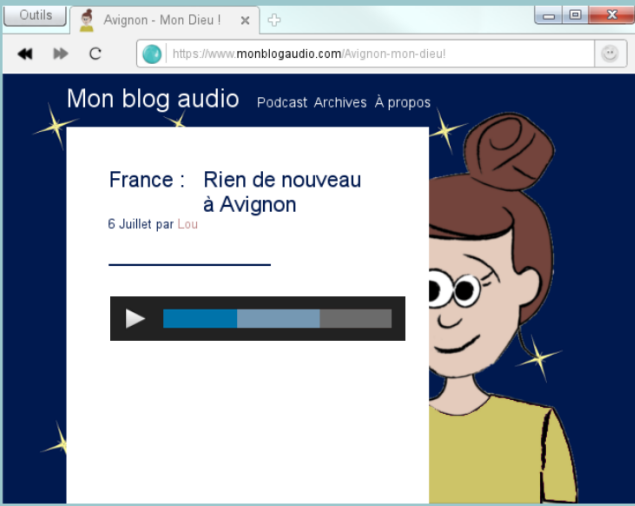
Du hörst den ersten Teil eines Podcasts. Darin berichtet die Touristin Lou von ihrer Reise nach Avignon. Beantworte die Frage.

Lou war für drei Tage nicht im Internet. Warum?

- ☐ Sie war in einem kleinen Dorf ohne Internetempfang.
- ☐ Ihre neue französische SIM-Karte war noch nicht aktiviert.
- ☐ In ihrem Hotel in Avignon funktionierte das Internet nicht.

Weiter

Seite 2



Höre den zweiten Teil von Lous Podcast. Beantworte dann die Frage.

Was hat Lou in Avignon besonders beeindruckt?

- ☐ Das gute Essen am zweiten Tag.
- ☐ Der starke Regen in der Stadt.
- ☐ Die schöne Aussicht vom Schloss.

Weiter

Task 3: Berufswahl

Seite 1

Tools Orientation professionnr: x

https://www.orientation.ch/tem-jeunes/recherche/?q=representant+commercial

Orientation professionnelle: Témoignages de jeunes

Résultat de la recherche:

représentant commercial

47 résultats

Témoignage de Mélanie

... contrairement aux vendeurs en magasin, nous devons beaucoup voyager. Je vais régulièrement rencontrer les clients francophones de mon entreprise. Bien sûr, j'essaie aussi de trouver de nouveaux clients pour nos produits. Il y a d'autres **représentants commerciaux** dans mon équipe qui parlent allemand, italien ou anglais. Chacun cherche des clients qui parlent sa propre langue...

Témoignage de Guillaume

Après que j'aie fini ma formation, j'ai voyagé autour du monde en vélo pendant trois ans. Après ça, j'ai travaillé comme guide touristique au sud de l'Angleterre. C'était très stimulant et j'ai appris beaucoup d'anglais. Maintenant, je suis de retour en France et je travaille comme **représentant commercial** au centre de Paris. C'est un travail intéressant et je suis bien payé, donc je suis déjà en train de réfléchir à une nouvelle aventure de voyage.

Témoignage de Étienne

Quand il y a un nouveau produit, les techniciens doivent le tester. On ne vend rien avant qu'on ne soit pas sûrs que ça marche ! Ensuite, on rencontre les **représentants commerciaux** pour décider de la stratégie de vente : qui sont les futurs acheteurs, quel langage utiliser pour le décrire. Mon rôle, c'est de...

Du liest Beschreibungen von jungen Leuten über den Beruf „représentant commercial“. Zu diesem Beruf musst du drei Fragen beantworten.

1. Beantworte jede Frage mithilfe der Texte. Klicke an!
2. In welchem Text hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du im gleichen Text mehrere Antworten.

Muss man im Beruf „représentant commercial“...

... selbst neue Produkte testen?

☒ ja ☒ nein

Témoignage de ?

... mehrere Sprachen sprechen?

☒ ja ☒ nein

Témoignage de ?

... viel reisen?

☒ ja ☒ nein

Témoignage de ?

Weiter

Seite 2

Tools Orientation professionnr: x

https://www.orientation.ch/tem-jeunes/recherche/?q=responsable+grands+comptes

Orientation professionnelle: Témoignages de jeunes

Résultat de la recherche:

responsable grands comptes

32 résultats

Témoignage de Diane

Après l'université, j'ai trouvé un emploi dans une grande compagnie. J'y travaille en tant que **responsable grands comptes** : je dois trouver les meilleures offres pour nos clients les plus importants. Comme nous sommes une compagnie internationale, je dois parler la langue de mes clients. Je suis donc bien contente d'avoir appris trois langues à l'école. Parfois, j'écris des e-mails à destination de quatre pays différents dans une seule matinée !

Témoignage de Camille

... avant, je ne pensais pas que je deviendrais vendeuse. J'ai fait des études d'ingénieure parce que je voulais créer et tester de nouveaux produits. Finalement, je ne fais pas ça, mais c'est très bien comme ça. Quand nous avons créé notre entreprise, il fallait quelqu'un qui s'occupe des clients importants. Maintenant, je suis **responsable grands comptes**, et je dois beaucoup voyager en Europe pour mon métier. Cela signifie aussi que...

Témoignage de Mattieu

Moi, j'aime beaucoup mon travail. En tant que **responsable grands comptes**, il faut savoir gérer son temps. Le travail doit être terminé à temps, et il n'y a pas de place à l'erreur. Le succès de l'entreprise en dépend. Mais si tu planifie bien ton temps, ce travail laisse aussi beaucoup de temps libre. Moi, je l'utilise pour voyager. Le mois prochain, je vais en Espagne pour tester ma nouvelle planche de surf !

Du liest Beschreibungen von jungen Leuten über den Beruf „responsable grands comptes“. Zu diesem Beruf musst du drei Fragen beantworten.

1. Beantworte jede Frage mithilfe der Texte. Klicke an!
2. In welchem Text hast du gelesen, ob die Antwort ja oder nein ist? Ziehe den Namen des Autors in das Kästchen.

Manchmal findest du im gleichen Text mehrere Antworten.

Muss man im Beruf „responsable grands comptes“ ...

... selbst neue Produkte testen?

☒ ja ☒ nein

Témoignage de ?

... mehrere Sprachen sprechen?

☒ ja ☒ nein

Témoignage de ?

... viel reisen?

☒ ja ☒ nein

Témoignage de ?

Weiter

8.4 ANHANG 4: TRANSKRIPTIONSRICHTLINIEN

Zusammenstellung der wichtigsten Transkriptionskonventionen für das Minimaltranskript und das darauf aufbauende Basistranskript (GAT 2), die für die Transkription der Verbalprotokolle angewandt wurden.

Quelle der Screenshots: Selting et al., 2009, S. 391–392.

Minimaltranskript

Sequenzielle Struktur/Verlaufsstruktur

[] Überlappungen und Simultansprechen
[]

Ein- und Ausatmen

°h / h° Ein- bzw. Ausatmen von ca. 0.2-0.5 Sek. Dauer
°hh / hh° Ein- bzw. Ausatmen von ca. 0.5-0.8 Sek. Dauer
°hhh / hhh° Ein- bzw. Ausatmen von ca. 0.8-1.0 Sek. Dauer

Pausen

(.) Mikropause, geschätzt, bis ca. 0.2 Sek. Dauer
(-) kurze geschätzte Pause von ca. 0.2-0.5 Sek. Dauer
(--) mittlere geschätzte Pause v. ca. 0.5-0.8 Sek. Dauer
(---) längere geschätzte Pause von ca. 0.8-1.0 Sek. Dauer
(0.5) gemessene Pausen von ca. 0.5 bzw. 2.0 Sek. Dauer
(2.0) (Angabe mit einer Stelle hinter dem Punkt)

Sonstige segmentale Konventionen

und_äh Verschleifungen innerhalb von Einheiten
äh öh äm Verzögerungssignale, sog. "gefüllte Pausen"

Lachen und Weinen

haha hehe hihi silbisches Lachen
((lacht)) ((weint)) Beschreibung des Lachens
<<lachend> > Lachpartikeln in der Rede, mit Reichweite
<<:-)> soo> "smile voice"

Rezeptionssignale

hm ja nein nee	einsilbige Signale
hm_hm ja_a nei_ein nee_e	zweisilbige Signale
ʔhmʔhm,	mit Glottalverschlüssen, meistens verneinend

Sonstige Konventionen

((hustet))	para- und außersprachliche Handlungen u. Ereignisse
<<hustend> >	sprachbegleitende para- und außersprachliche Handlungen und Ereignisse mit Reichweite
()	unverständliche Passage ohne weitere Angaben
(xxx), (xxx xxx)	ein bzw. zwei unverständliche Silben
(solche)	vermuteter Wortlaut
(also/alo)	mögliche Alternativen
(solche/welche)	
((unverständlich, ca. 3 Sek))	unverständliche Passage mit Angabe der Dauer
((...))	Auslassung im Transkript
→	Verweis auf im Text behandelte Transkriptzeile

Basistranskript

Sequenzielle Struktur/Verlaufsstruktur

=	schneller, unmittelbarer Anschluss neuer Sprecherbeiträge oder Segmente (<i>latching</i>)
---	---

Sonstige segmentale Konventionen

:	Dehnung, Längung, um ca. 0.2-0.5 Sek.
::	Dehnung, Längung, um ca. 0.5-0.8 Sek.
:::	Dehnung, Längung, um ca. 0.8-1.0 Sek.
ʔ	Abbruch durch Glottalverschluss

Akzentuierung

akZENT	Fokusakzent
ak!ZENT!	extra starker Akzent

Tonhöhenbewegung am Ende von Intonationsphrasen

?	hoch steigend
,	mittel steigend
—	gleichbleibend
;	mittel fallend
.	tief fallend

Sonstige Konvention

<<erstaunt> > interpretierende Kommentare mit Reichweite

Zusatzregel:

Zwischen zwei Sternchen *...* wird auf Standarddeutsch transkribiert, was die SuS im Dialekt gesagt haben, (z.B. *das war cool*, und nicht etwa *das isch kul gsi*)

Anmerkung:

Wenn in der Videodatei eine unverständliche Äusserung vorkam, wurde es an manchen Stellen auch so gekennzeichnet: (unv.). Die Dauer der unverständlichen Passage wurde dabei nicht eingegeben.

Allgemeine Bemerkungen

- In der Spalte „Beispiele“ wird vor dem jeweiligen Beispiel angegeben, in welchem Untersuchungsteil und auf welche Aufgabe(n) sich die Äusserungen der Lernenden beziehen (z.B. *Szenario, Lautes Denken, Task A, (Bildschirm-)Seite 2*). Da bei der Codierung stets die Aufgaben und Inputtexte berücksichtigt werden mussten, wird auch für das Nachvollziehen der Beispiele in manchen Fällen die Konsultation der Aufgabe und/oder des Inputtextes wohl nötig sein (vgl. Anhang 2 und 3).
- Mit „Text“ sind im Transkript der Inputtext bzw. die Inputtexte gemeint, die in der Zielsprache Französisch verfasst wurden. „Aufgabenstellung“ umfasst jeweils die Erklärungen zur Aufgabe und, ausser auf den Szenario-Einführungsseiten, auch das Item (Fragestellung und Antwortoptionen). Alle Einführungsseiten, Aufgabestellungen und Items waren in der Schulsprache Deutsch verfasst.
- Beim Codieren aufgrund der Videos (vgl. Spalte *Hinweise für Codierende*) wurden bei Bedarf die Transkripte im Nachhinein so verfeinert, dass die Codierungen nun lediglich aufgrund der Transkripte nachvollzogen werden können.

Übersicht über die Gewichtungen der Codierungen¹³²

In der vorliegenden Studie kamen folgende Gewichtungen zum Einsatz:

A. Gewichtung bei allen Items (pro besprochene Bildschirmseite bzw. besprochenes Item):

- **0:** Standardgewichtung; wenn sich eine bestimmte Äusserung auf das gerade besprochene Item bezieht.
- **2:** Wenn der/die Schüler/in (S) beim Besprechen einer Bildschirmseite eine **wiederholte Aussage** machte, die bei dieser Bildschirmseite bereits codiert wurde. Diese Gewichtung wird auch dann vergeben, wenn die erste Codierung mit einer unter B-D (siehe unten) spezifizierten Gewichtung versehen wurde.
- **5:** Wenn aus dem Kontext hervorgeht, dass der/die S eine **allgemeine(re)** Aussage zu seinem/ihrer Verhalten beim Lösen der Testaufgaben macht, die item- oder textstellenübergreifend ist und nicht nur das gerade besprochene Item oder die betroffene Textstelle betrifft. Diese Gewichtung wird auch dann vergeben, wenn es sich um eine wiederholte Erwähnung dieser Strategie handelt.

B. LD-L15 und SR-L15 („Kognaten identifizieren und nutzen“):

- Standardgewichtung (0) wenn das Wort richtig als Kognat erkannt wurde.
- **10:** Der/die S hat das Wort fälschlicherweise als Kognaten eingestuft bzw. hat die Bedeutung des Wortes im vorliegenden Kontext fälschlicherweise ermittelt; oder er/sie vermutet, dass es sich bei einem Wort um einen Kognaten handelt, er/sie weiss aber nicht, was das Wort auf Deutsch bedeutet.

¹³² Vgl. Kap 3.7.2.

C. LD-TM08 und SR-TM08 („Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)“):

- Standardgewichtung (0): Wenn die mit TM08 codierte Passage richtig bzw. ausreichend für die korrekte Beantwortung der Frage (z.B. Howard Pos. 22-27) verstanden wurde.
- **15**: Wenn die mit TM08 codierte Passage falsch oder nicht genügend verstanden wurde, um die richtige Antwort zu wählen.

D. LD-TM12 und SR-TM12 („Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen“):

- Standardgewichtung (0): Wenn die mit TM12 codierte Passage plausible Schlussfolgerungen beinhaltet, die aufgrund des Textinhaltes gemacht werden können.
- **20**: Wenn die mit TM12 codierte Passage Schlussfolgerungen beinhaltet, die nach dem Vergleich mit der entsprechenden Textstelle als unplausibel eingestuft wurden, weil es für diese Deutung im Inputtext keine bzw. nicht genügend Evidenz gab.

E. SR-L10 („Unbekannte Wörter, Satz- bzw. Textteile identifizieren“)

- **25**: Codierungen, die bei den mündlichen Zusammenfassungen der Inputtexte vorgenommen wurden.

F. SR-SC_2 („Szenario nach einem Hinweis wahrgenommen“)

- Standardgewichtung (0): Wenn das Szenario (bzw. Avignon, Reisen etc.) von der/dem S erwähnt wird, nachdem er/sie gefragt wurde, ob ihm/ihr thematisch etwas an den Aufgaben aufgefallen ist.
- **30**: Nachdem die Forscherin verraten hat, dass die Aufgaben thematisch verbunden sind und der/die S im Nachhinein zugibt, dass es ihm/ihr jetzt aufgefallen ist bzw. eigentlich schon früher „unbewusst“ aufgefallen ist.

Wenn in der Spalte „Beispiele“ bei der Angabe der Stelle im Transkript keine spezielle Gewichtung (vgl. Kap. 3.7.2) angemerkt wird, wie z.B. hier: *Howard_Szenario, Pos. 35*, handelte es sich um die Codierung mit der Standardgewichtung „0“. Ansonsten wird die Gewichtung der Codierung hinter der Angabe zur Position des Zitats im Transkript angegeben, z.B. *Elena_Szenario, Pos. 18-19; Gewichtung 10*.

8.5.1 CODIERUNG DER ALLGEMEINEN, NICHT STRATEGIEBEZOGENEN INFORMATIONEN

Code	Memo ¹³³
SC-LD¹³⁴	Szenario, Untersuchungsteil Lautes Denken
Intro	Bildschirmseite mit der Szenarioeinführung
A-S1	Szenario, Lautes Denken, Task A Seite 1, inkl. Intro Task A
A-S2	Szenario, Lautes Denken, Task A Seite 2
A-S3	Szenario, Lautes Denken, Task A Seite 3
B-S1	Szenario, Lautes Denken, Task B Seite 1, inkl. Intro Task B
B-S2	Szenario, Lautes Denken, Task B Seite 2
B-S3	Szenario, Lautes Denken, Task B Seite 3
C-S1	Szenario, Lautes Denken, Task C Seite 1, inkl. Intro Task C
C-S2	Szenario, Lautes Denken, Task C Seite 2
C-S3	Szenario, Lautes Denken, Task C Seite 3
D-S1	Szenario, Lautes Denken, Task D Seite 1, inkl. Intro Task D
D-S2	Szenario, Lautes Denken, Task D Seite 2
D-S3	Szenario, Lautes Denken, Task D Seite 3
SC-SR	Untersuchungsteil <i>Stimulated-Recall</i>-Interview
Intro	<i>Stimulated Recall</i> , Bildschirm mit der Einführung inkl. Diskussion am Anfang
A-S1	Szenario, <i>Stimulated Recall</i> , Task A Seite 1, inkl. Intro Task A
A-S2	Szenario, <i>Stimulated Recall</i> , Task A Seite 2
A-S3	Szenario, <i>Stimulated Recall</i> , Task A Seite 3
B-S1	Szenario, <i>Stimulated Recall</i> , Task B Seite 1, inkl. Intro Task B
B-S2	Szenario, <i>Stimulated Recall</i> , Task B Seite 2
B-S3	Szenario, <i>Stimulated Recall</i> , Task B Seite 3
C-S1	Szenario, <i>Stimulated Recall</i> , Task C Seite 1, inkl. Intro Task C
C-S2	Szenario, <i>Stimulated Recall</i> , Task C Seite 2
C-S3	Szenario, <i>Stimulated Recall</i> , Task C Seite 3
D-S1	Szenario, <i>Stimulated Recall</i> , Task D Seite 1, inkl. Intro Task D
D-S2	Szenario, <i>Stimulated Recall</i> , Task D Seite 2
D-S3	Szenario, <i>Stimulated Recall</i> , Task D Seite 3
SC-Abschluss	Allgemeine, abschliessende Diskussion mit den Szenario-Schüler/innen nach <i>Stimulated Recall</i>
Off-LD	Off-Szenario, Untersuchungsteil Lautes Denken
Walkies-1	Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 1

¹³³ Für vertiefende Informationen zur Codierung allgemeiner Kategorien vgl. Kap. 3.7.3.1.

¹³⁴ Oberkategorien wurden hier fett markiert.

Code	Memo ¹³³
Walkies-2	Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 2
Stadttrip-1	Off-Szenario, Lautes Denken, Szenario „Stadttrip“, Seite 1
Stadttrip-2	Off-Szenario, Lautes Denken, Szenario „Stadttrip“, Seite 2
Berufswahl-1	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1
Berufswahl-2	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 2
Waitaha-1	Off-Szenario, Lautes Denken, Szenario „Waitaha“, Seite 1
Waitaha-2	Off-Szenario, Lautes Denken, Szenario „Waitaha“, Seite 2
Schulfest-1	Off-Szenario, Lautes Denken, Szenario „Schulfest“, Seite 1
Off-SR	Off-Szenario, Untersuchungsteil <i>Stimulated-Recall</i>-Interview
Walkies-1	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1, inkl. Diskussion am Anfang
Walkies-2	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 2
Stadttrip-1	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Stadttrip“, Seite 1
Stadttrip-2	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Stadttrip“, Seite 2
Berufswahl-1	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 1
Berufswahl-2	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 2
Waitaha-1	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Waitaha“, Seite 1
Waitaha-2	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Waitaha“, Seite 2
Schulfest-1	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Schulfest“, Seite 1
Off-Abschluss	Allgemeine, abschliessende Diskussion mit den Off-Szenario-Schüler/innen nach <i>Stimulated Recall</i>
R	Richtige Antwort (beim LD, Endresultat)
F	Falsche Antwort (beim LD, Endresultat)

8.5.2 CODIERUNG DER STRATEGIEN BEIM LAUTEN DENKEN (LD)

Mit diesen Codes wurden Phänomene und Äusserungen festgehalten, die in den Lautdenkprotokollen der Schülerin oder des Schülers (S) beobachtet werden konnten.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
	Lautes Denken: <i>Lernerstrategien</i>	Gedanken und Handlungen, die von den Sprachlernenden bewusst gewählt und eingesetzt werden, um den fremdsprachigen Text zu verstehen . Diese Strategien sind für eine Testsituation nicht spezifisch, d.h. sie werden auch in anderen Sprachlern- und Sprachgebrauchssituationen eingesetzt.		
LD-L01	Sorgfältiges Lesen	S liest den gesamten Inputtext sorgfältig durch.	Wird aufgrund des Videos codiert, wenn der/die S den gesamten Text sorgfältig liest, und zwar mind. etwa eine Minute lang bzw. mind. 20 Sekunden lang, falls der/die S beim Task C (Internetreviews bei Szenario „Stadttrip“ und „Berufswahl“) einen der drei Texte separat liest. Wenn der/die S während des Lesens zusätzlich z.B. die Aufgabenstellung liest, wird die ganze Passage im Transkript	Szenario, Lautes Denken, Task A Seite 2: ((liest den text)) (1 min) (Howard_Szenario ¹³⁵ , Pos. 35) Szenario, Lautes Denken, Task C Seite 1: ((liest den zweiten text von miguel)) (38.9) (Howard_Szenario, Pos. 94)

¹³⁵ Bei Schüler/innen, die Szenario „Stadttrip“ bearbeiteten, wird dies nach ihrem Pseudonym und dem Unterstrich aufgeführt. Bei den Schüler/innen, die Off-Szenario-Aufgaben bearbeiteten, steht nach dem Unterstrich „Off“.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
			trotzdem nur einmal mit L01 codiert (und zusätzlich mit TM06).	
LD-L02	Schnelles Lesen	S überfliegt den gesamten Inputtext schnell. Er/Sie sucht dabei evtl. nach Schlüsselwörtern.	Wird aufgrund des Videos codiert, und zwar wenn die/der S den Text kürzer als etwa eine Minute lang liest (bei Task C kürzer als 20 Sek. lang pro Text). Wenn der/die S während des Lesens zusätzlich z.B. die Aufgabenstellung liest, wird die ganze Passage im Transkript trotzdem nur einmal mit L02 codiert (und zusätzlich mit TM06).	Szenario, Lautes Denken, Task A Seite 1: ((S blickt auf den bildschirm und liest den text)) (41.8) (Alison_Szenario, Pos. 14) Szenario, Lautes Denken, Task C Seite 1: ((S liest text 3, cursor auf text 3)) (14.7) (Arya_Szenario, Pos. 148)
LD-L03	Einen Teil des Inputtextes (erneut bzw. genauer) lesen	S liest einen Teil des Inputtextes erneut bzw. genauer durch.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S oder auch aufgrund des Videos codiert.	Szenario, Lautes Denken, Task C Seite 1: ((S liest einen teil des inputtextes)) (22.1) (Alison_Szenario, Pos. 110)
LD-L04	Weiterlesen, um Nichtverstandenes zu verstehen ¹³⁶	S verwendet andere Teile des Inputtextes, um das Verständnis eines bestimmten Teils zu erleichtern: Er/Sie sagt, dass er/sie weiterlesen will, um nach Informationen zu suchen, die zum Verständnis des bereits Gelesenen beitragen können.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	In den Lautdenkprotokollen konnte diese Strategie nicht beobachtet werden.
LD-L05	Bisheriges Verständnis hinterfragen	Nachdem er/sie mehr gelesen oder gehört hat, fragt sich der/die S, ob das bisherige Verständnis in Anbetracht der neuen Informationen noch zutrifft.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task C Seite 1: ((schaut auf den text)) *jetzt bin ich bei nummer zwei nicht sicher (.) (noch einmal lesen) - (Arya_Szenario, Pos. 160)

¹³⁶ Strategien, die nur in einem Untersuchungsteil (Lautes Denken oder *Stimulated Recall*) vorkamen, werden im Kategorienleitfaden übersichtshalber in beiden Teilen aufgeführt. In dem Untersuchungsteil, in dem sie nicht beobachtet werden konnten, wurden sie grau markiert und es wurde in der Spalte „Beispiele“ eine entsprechende Anmerkung gemacht.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
LD-L06	Bisheriges Verständnis korrigieren	S korrigiert das Verständnis des Textes oder eines Teils davon, nachdem er/sie den Text erneut gelesen bzw. weitergelesen/weitergehört hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 2: er hat gesagt em ((cursor bewegt sich über text 2)) (3.0) also (Und) (.) tEstet neue produkte- das hEisst/ also ich denke dass es Eher ein jA ist- ((S liest den zweiten text)) em [...] (7.4) Aber (.) im unteren satz steht "je ne fais PAS ça" also (.) ich mache NICHT- hm: [...] aber ich denke es ist eher nEin (Ezra_Off, Pos. 248-254)
LD-L07	Bisheriges Verständnis bestätigen	S bestätigt das aktuelle Verständnis des Inputtextes oder eines Teils davon, nachdem er/sie sich damit weiter und/oder vertiefter auseinandergesetzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 2: *also ganz am anfang haben sie etwas gesagt von (-)* "carte (.) avignon" °hh (2.0) *und (--) die kostet zwanzig euro <<tuschelnd> (unv.)> (2.2) glaube ich-* h°° (--) nein (.) *die karte ist NICHT (--) ah sie ist nicht (.) zum haben;* °h (--) ehm (([...] liest den text)) (4.5) * gut sie empfehlen es SCHON aber (-) man kann sie ja nicht haben- *(Boebu_Szenario, Pos. 79-84)
LD-L08	Intensive Auseinandersetzung mit Teilen des Inputtextes	S liest vor, wiederholt, paraphrasiert bzw. übersetzt Wörter, Satzteile oder Sätze, oder fasst Teile des Inputtextes zusammen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 1: château (---) ((schaut nach unten)) ah es ist schloss (---)((liest den inputtext weiter halblaut)) irgendetwas kreierte (---) das schloss ist ein klassisch_es monument immer (unv.) oder so ne faire aber *irgendetwas was nicht ist (--) ist nie (---) ((verzieht das

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				gesicht)) es interessiert mich nicht so?* (Aria_Szenario, Pos. 23-27)
LD-L09	Informationen verknüpfen	S setzt Teile von gelesenen bzw. gehörten Informationen miteinander in Beziehung.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Stadttrip“, Seite 2: *gut das essen- aber war das NICHT am dritten tag; oder so-* (---) (Omega_Off, Pos. 76)
LD-L10	Unbekannte Wörter, Satz- bzw. Textteile identifizieren	S identifiziert ein unbekanntes Wort, einen unbekannten Satz- bzw. Textteil oder er/sie zweifelt an seinem/ihrem Verständnis dieses Wortes oder dieser Satz- bzw. Textteile.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S oder des Videos codiert.	Szenario, Lautes Denken, Task C Seite 2: ("SAnS fil") (--) weiss ich jetzt nicht genAu (Alison_Szenario, Pos. 127) Szenario, Lautes Denken, Task C Seite 2: ((liest den zweiten text vor)) ((verzieht bei "sans fil" das gesicht)) (1.0) ((liest weiter vor)) (31.0) (Aria_Szenario, Pos. 198) Szenario, Lautes Denken, Task A Seite 2: * irgendetwas kostet (.) zwanzig EUro; man bekommt es im (--) touristen(-)office* (Arya_Szenario, Pos. 49-50)
LD-L11	Unbekannte Wörter ausblenden	S macht eine Äusserung dazu, dass er/sie sich nicht durch unbekannte Wörter ablenken lässt und diese ausblendet.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task C Seite 2: ("SAnS fil") (--) weiss ich jetzt nicht genAu aber dann (.) denke ich es ist gRAtis- (Alison_Szenario, Pos. 127)
LD-L12	Wortbedeutung identifizieren	S identifiziert die Bedeutung eines Wortes bzw. erinnert sich daran, z.B. aufgrund vom Kontext.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task C Seite 2: das personal ist (2.0) ((hebt die hände hoch)) serviable ((verzieht

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>das gesicht)) (1.5) serviert gut? (Aria_Szenario, Pos. 191-192)</p> <p>Szenario, Lautes Denken, Task A Seite 1: ah dieses (-) "château"` sch- ein schloss- (-) oder? ((verbirgt das gesicht hinter den händen)) (2.0) oder eine burg oder so` (-) (Cybermen_Szenario, Pos. 16-17)</p>
LD-L13	Morphosyntaktische Regeln anwenden	S nutzt explizit seine/ihre Kenntnisse der morphosyntaktischen Regeln.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 2: "je voul/" A:H; (2.0) also (---) "vouler" w_wollen; (-) wollte; (-) *vergangenheit; (Omega_Off, Pos. 140-141)
LD-L14	Interpunktionskenntnisse nutzen	S nutzt seine/ihre Kenntnisse von Interpunktionszeichen	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: hat es neue produkt.*testen (.) ((schaut auf den text)) man sieht nichts mit fragezeichen* (Eleven_Off, Pos. 112-113)
LD-L15	Kognaten identifizieren und nutzen	S identifiziert und nutzt (vermeintliche) Kognaten im Text.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: "LE tester " denke ich das hat eine verbindung mit (-) mit dem dEU/ also ein parallelwort mit tEsten- (Ezra_Off, Pos. 173-175) <p>Szenario, Lautes Denken, Task A Seite 1: das schloss anschauen ((zieht 1 nach A1w)) (2.0) *weil ich denke dass (-) hier STEHT ja (-) burg ((zeigt mit der maus auf "leurs burgers")) (-) das ist ja</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>schloss.* (3.0) (Elena_Szenario, Pos. 18-19; Gewichtung 10)</p> <p>Szenario, Lautes Denken, Task A Seite 3: *manchmal suche ich auch so parallelwörter (--) zum schauen (.) was es sein könnte;* (Elena_Szenario, Pos. 36-37; Gewichtung 5)</p>
LD-L16	Prosodie beim Hören nutzen	S achtet beim Hören auf die Prosodie und nutzt diese Informationen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	In den Lautdenkprotokollen konnte diese Strategie nicht beobachtet werden.
LD-L17	Relevantes Hintergrundwissen einbeziehen	S identifiziert und nutzt relevantes Hintergrundwissen in geeigneter Weise, d.h. er/sie greift <i>zusätzlich</i> zu den Informationen aus dem Text auf sein/ihr Weltwissen zurück, um die Textbedeutung zu verstehen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 2: also es [d.h. «Carte Avignon»] kostet halt VORher aber nachher kann man mit allen gratis fahren (.) so vielleicht für einen tag oder so ; (Sheldon_Szenario, Pos. 45-46)
LD-L18	Selbsteinschätzung	S macht beim Lauten Denken eine selbsteinschätzende (positive oder negative) Aussage in Bezug auf die eigenen Französischkenntnisse bzw. das Textverständnis.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 2: *okay ich habe KEINE ahnung; (2) ich habe keine ahnung;* ((schüttelt den kopf)) (Omega_Off, Pos. 41)</p> <p>Szenario, Lautes Denken, Task C Seite 2: bei dem ersten tex:t (-) verstehe ich nicht wirklich viel- (Cybermen_Szenario, Pos. 133)</p>
	Lautes Denken: <i>Testmanagementstrategien</i>	Strategien für eine sinnvolle Reaktion auf die Testaufgabe(n) . Bewusst und zielorientiert gewählt, um eine richtige Antwort zu		

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		produzieren. Sie sind typisch für eine Testsituation.		
LD-TM01	Instruktionen und/oder Antwortoptionen (erneut) lesen	S liest die Instruktionen und/oder die Antwortoptionen (erneut) durch.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert. Wird bei jeder Erwähnung im Transkript codiert, aus der hervorgeht, dass die Instruktionen/Antwortoptionen rezipiert werden (aber nicht dann, wenn der/die S. gerade die ausgewählte Antwort anklickt und sein/ihr Blick zwangsweise auf die Antwortoptionen richtet). Wird auch codiert, wenn der/die S die Einführungsseiten zum Szenario liest.	<p>Szenario, Lautes Denken, Task A Seite 1: ((S liest die instruktionen)) (5.6) (Alison_Szenario, Pos. 15)</p> <p>Szenario, Lautes Denken, Task A Seite 1: ((S liest die antwortmöglichkeiten)) (6.5) (Alison_Szenario, Pos. 21)</p> <p>Szenario, Lautes Denken, Task A Seite 1: ((S blickt abwechselnd auf den text und die aufgabenstellung)) (10.4) (Alison_Szenario, Pos. 26)</p>
LD-TM02	Paraphrasieren und/oder Übersetzen der Aufgabe ¹³⁷	S paraphrasiert die Aufgabe oder übersetzt für sich die Frage, die Antwortoptionen oder Teile davon.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, Lautes Denken, Szenario „Stadttrip“, Seite 1: ((liest aufgabenstellung und murmelt dazu mehr oder weniger laut)) (12.7) *sie war in einem kleinen dorf OHNe internetempfang, (--) neue französische sim karte noch nicht aktiviert; (-)im hotel in aviGon* ((statt avignon)) funktioniert das INternet nicht;* (Bonnie_Off, Pos. 66-69)</p> <p>Szenario, Lautes Denken, Task A Seite 3: also ich (--) proBIere (-) IMMer (-) dIE (-) möglichst auf</p>

¹³⁷ Die Aufgabenstellung, Fragen und Antwortoptionen wurden in der Schulsprache Deutsch formuliert und einige Schüler/innen versuchten, die Antwortoptionen auf Französisch zu übersetzen. Bei der Off-Szenario-Aufgabe „Berufswahl“ kamen in der Aufgabenstellung auch französische Berufsbezeichnungen vor.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				französisch (-) was das HEISSen könnte (-) (Cersei_Szenario, Pos. 65-66; Gewichtung 5)
LD-TM03	Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren ¹³⁸	S liest die Aufgabenstellung bzw. die Antwortoptionen und identifiziert das ihm/ihr darin unbekannte Vokabular.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert, wenn S verbalisiert, dass er/sie 1) in der Aufgabenstellung ein deutsches bzw. 2) bei Off-Szenario in der Aufgabe „Berufswahl“ ein französisches Wort nicht versteht, oder 3) nicht weiss, wie das Vokabular aus der Aufgabenstellung auf Französisch heissen würde.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: ((S liest die aufgabenstellung und spricht leise mit)) (2.2) <<verzieht das gesicht> *eh was IST das; "représentant (commercial)" (Omega_Off, Pos. 82-83) Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 2: *texte der anderen korrekturlesen- (---) KEINE ahnung was das ist-* (Omega_Off, Pos. 32-33)
LD-TM04	Vorhersage machen	S macht eine Vorhersage: Er/Sie überlegt sich vor der Rezeption des Inputtextes, welche Antwortoption(en) evtl. stimmen und welche ausgeschlossen werden könnten. Aus den weiteren Teilen des Transkripts geht aber hervor, dass er/sie diese Antwortoption(en) vor der Textrezeption noch nicht ausgeschlossen hat.	Wird aufgrund der Bemerkung(en) der/des S codiert. Wenn aus dem Transkript hervorgeht, dass der/die S eine Antwort von Anfang an (d.h. vor dem Lesen/Hören) ausgeschlossen hat, wird nicht TM04, sondern TW05 codiert.	Szenario, Lautes Denken, Task B Seite 2: ((liest die antwortoptionen vor)) *also (...) der starke regen in der stadt ist sicher nicht weil das ist ja schLECHT* ((setzt kopfhörer an)) (Aria_Szenario, Pos. 122-123)
LD-TM05	Frage und Antwortoptionen beim Hören anschauen	S schaut sich die Frage und/oder die Antwortoptionen beim Hören an.	Wird aufgrund des Videos codiert.	Off-Szenario, Lautes Denken, Szenario „Stadttrip“, Seite 1:

¹³⁸Die Aufgabenstellung, Fragen und Antwortoptionen wurden in der Schulsprache Deutsch formuliert. Bei der Off-Szenario-Aufgabe „Berufswahl“ kamen in der Aufgabenstellung auch französische Berufsbezeichnungen vor.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				((S hört sich den audiobeitrag an und schaut auf die aufgabenstellung)) (50.7) (Omega_Off, Pos. 56)
LD-TM06	Fokus auf relevante Textteile	S liest die Frage und/oder die Antwortoption(en). Er/Sie konzentriert sich beim Hören/Lesen auf die Teile des Inputtextes, die für die Beantwortung der Frage von Bedeutung sind und sucht, wo im Inputtext die entsprechenden Informationen (vermutlich) vorkommen. Evtl. ignoriert er/sie dabei die als nicht relevant eingestuft Teile des Inputtextes.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird auch codiert, wenn der/die S das Lesen des Inputtextes immer wieder unterbricht, um die Antwortoptionen zu lesen (weil es impliziert, dass der/die S. die Relevanz der gelesenen Stelle im Hinblick auf die Beantwortung der Testfragen überprüft).	Szenario, Lautes Denken, Task A Seite 2: eine boo:tOUr (--) könnte auch sein` (---) einen stAdtrundgang (---) mit einem: smArtphone ((S bewegt die augen zwischen dem text und den antwortmöglichkeiten hin und her)) (15.7) (Alison_Szenario, Pos. 47-50) Off-Szenario, Lautes Denken, Szenario „Stadttip“, Seite 1: ((liest die aufgabenstellung (...))) (16.1) *(unv.) ich mal etwas von französische sim karte hören* (unv.) (--) *okay entweder NICHT (-) DORF (-) "HOTEL" (-) sim karte und dann sollte man etwas von dem hören* (---) (Omega_Off, Pos. 52-55) Szenario, Lautes Denken, Task A Seite 3: also ich (--) proBIERE (-) IMMER (-) DIE (-) möglichst auf französisch (-)was das HEISSen könnte (-) und dann kann ich DIE hier im text (-) SUCHen? (---) und dann schauen ob sie das (.) emPFEHlen oder NICHT; (Cersei_Szenario, Pos. 65-68; Gewichtung 5)

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
LD-TM07	Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären	S. klärt oder bestätigt die Bedeutung von bestimmten Teilen des französischen Inputtextes mithilfe der auf Deutsch vorhandenen Frage und/oder der Antwortoptionen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S und/oder des Videos codiert.	Szenario, Lautes Denken, Task A Seite 1: das "cOq rouge" ist ein_e res:to ((blickt zu den Antwortmöglichkeiten und geht zurück zum text)) (.) ja. (Alison_Szenario, Pos. 28)
LD-TM08	Einsatz von Verstandenen bei Auswahl/Ausschluss der Antwortoption(en)	S wählt oder schliesst eine Antwortoption bzw. -optionen aufgrund von etwas aus, was er/sie vom Inputtext verstanden hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird auch codiert, wenn der/die S zwar etwas nicht richtig verstanden hat, aber auf das (falsch) Verstandene bei der Auswahl oder beim Ausschluss einer Antwortoption zurückgreift.	Szenario, Lautes Denken, Task A Seite 1: aber das (-) STAdion (-) da gibt es keine TOUR oder so etwas (-) also wollen sie die rugby ((zieht 3 nach Alx)) (2.0) ausstellung; (Howard_Szenario, Pos. 22-23) Szenario, Lautes Denken, Task B Seite 2: also wi/ ich bin mir ziemlich sicher das SCHLOSS (-) die aussicht (-) dass die fantastisch war hat sie gesagt. ((wählt B2c und klickt "weiter")) (Howard_Szenario, Pos. 73-74; Gewichtung 15)
LD-TM09	Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden	S liest bzw. hört den Inputtext nicht weiter, nachdem er/sie die (vermeintlich) für die Beantwortung der Frage benötigten Informationen im Text gefunden hat.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task C Seite 1: ((wählt bei C2_3 a)) ((schaut auf den inputtext)) ehm (--) ce n'est pas beaucoup (.) es ist nicht viel (unv.) (4.0) *keine ahnung (--) sauber °hh (-) jetzt muss ich nicht WEiterlesen (.) ja* hh° (2.0) *jetzt kann ich [...] weiterlesen) ((meint text 3))* (Aria_Szenario, Pos. 167-170)
LD-TM10	Richtigkeit der Antwortoption überprüfen	S geht zur Frage und/oder zum Inputtext zurück, um zu prüfen, ob die in Erwägung gezogene Antwortoption korrekt ist oder nicht. Er/Sie macht das bevor	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 1: ((S zieht 2 in Alw)) ((S blickt abwechselnd auf die rechte und linke seite des bildschirms)) (10.4) mh_mh °h (2.0) ((S nickt

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		oder nachdem er/sie eine Antwortoption wählt oder gewählt hat.		<p>mit dem kopf und liest den text weiter)) (Alison_Szenario, Pos. 25-27)</p> <p>Szenario, Lautes Denken, Task A Seite 2: ((S bewegt die augen zwischen dem text und den antwortmöglichkeiten hin und her)) (15.7) ((S zieht 3 in A2x)) (Alison_Szenario, Pos. 50-51)</p>
LD-TM11	Antwortoptionen miteinander vergleichen	S vergleicht die möglichen Antwortoptionen miteinander.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: vielleicht ist doch das EHer mehrere sprachen sprechen (-) ((zeigt mit dem cursor auf den text von guillaume)) weil der hat ja et(.)was (---) ah nein das da OBen ist (.) mehrere sprachen sprechen?* ((zeigt auf den text von melanie, zieht 1 nach C1_2rev und wählt C1_2a)) (3.0) (Bonnie_Off, Pos. 136-138)</p> <p>Szenario, Lautes Denken, Task B Seite 1: ah sie hat (.) gesagt dass in dem hotel das INternet nicht funktionierte und dass sie es erst (.) später reparieren würden. (--) und in dem kleinen DORF gab es (-) GUTen internetempfang, (--) und ihre (-) sim karte (-) wollte sie glaub ich gar nicht kaufen oder so (-) (Howard_Szenario, Pos. 63-67)</p>
LD-TM12	Auswahl/Ausschluss der Antwortoption(en) aufgrund von Inferenzen	S wählt oder schliesst eine Antwortoption aufgrund von Inferenzen aus. Kommt oft bei	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 2: also anne sagt äh ein eine BOOTsfahrt wäre (...) LANGweilig

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		jenen Items vor, bei denen die Antwort nicht wortwörtlich im Text steht.		<p>(-) also ist das sicher schonmal NICHT gut; ((zieht 4 nach A2y)) (2.0) (Howard_Szenario, Pos. 36-38)</p> <p>Szenario, Lautes Denken, Task A Seite 2: und sie sagen auch es gibt eine app (-) also muss er die karte NICHT kaufen; ((zieht 2 nach A2z)) (Howard_Szenario, Pos. 42-43; Gewichtung 20)</p>
LD-TM13	Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war	S wählt eine Antwortoption aus, die ein Wort bzw. Wörter enthält, weil er/sie im Inputtext korrekterweise ein passendes Wort oder passende Wörter zu den Schlüsselwörtern aus dieser Antwortoption identifiziert hat. Kommt oft bei Items vor, die auf selektives Lesen (<i>search reading</i>) abzielen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Szenario, Lautes Denken, Task C Seite 1: drEissig euro pro nAcht (--) °h dann (--) jA. ((S klickt C1_3a an)) (Alison_Szenario, Pos. 107)</p> <p>Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: (3.0) viele REIsen; [...] *beaucoup voyA:;;" ((schaut auf den text)) ((S zieht 1 nach C1_3rev)) (6.1) viel reisen` (1.0) "beaucoup voyager" (-) *viel reisen JA nehme ich in dem fall;*((wählt C1_3a an)) (6.1) (Omega_Off, Pos. 96-100)</p>
LD-TM14	Auswahl einer Antwortoption, weil andere nicht plausibel sind ¹³⁹	S wählt eine Antwortoption bzw. -optionen aus, nachdem er/sie andere Antwortoptionen ausgeschlossen hat, die ihm/ihr basierend auf seinem/ihrer Verständnis des	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task B Seite 2: (3.0) am ersten tag war sie (-) beim sCHLOss- beim zWeiten (1.0) war sie glaube ich KRAnk (und) vom rEgen (3.0) (hat) sie glaube

¹³⁹ Man könnte hier auch argumentieren, dass die Strategie konstruktirrelevant ist, weil die Entscheidung der Schülerin oder des Schülers nicht für eine Antwortoption fiel, sondern gegen die restlichen Antwortoptionen. Nach langem Überlegen wurde entschieden, diese Strategie doch als konstruktrelevant einzustufen, weil sie impliziert, dass die/der Lernende hier nicht einfach den Test „austricksen“ wollte, sondern die Frage aufgrund dessen zu beantworten versuchte, was sie/er vom Text verstand.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		Inputtext oder eines Teils davon als nicht plausibel erscheinen.		ich nichts gesagt- (--)((S klickt B2c an)) (Alison_Szenario, Pos. 88-90)
LD-TM15	Kenntnisse über das Itemformat nutzen ¹⁴⁰	S nutzt bei Szenario „Stadttrip“, Task A die Kenntnisse über das Itemformat, um eine oder mehrere Antwortoptionen zu wählen. Wenn in Szenario „Stadttrip“, Task A z.B. zwei positive Antworten gefunden worden waren, dann wusste der/die S, dass die restlichen möglichen Antworten negativ sein müssen.	Wird aufgrund des Videos (wenn z.B. die negativen Punkte ohne weitere Überlegung reingezogen werden, nachdem zwei positive gefunden wurden) oder der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task A Seite 1: da gibt es nur zwei ((cursor auf Alw)) das heisst (5.0) ((blickt auf den text)) ((schaut auf die antwortoptionen)) das heisst das ist negativ; ((zieht 1 von Alw in Aly)) (aber) das ist ganz sicher (.) positiv- (--)((zieht 4 in Alw)) und in diesem fall muss das schloss negativ sein; (Arya_Szenario, Pos. 39-42)
LD-TM16	Zeitmanagement	S macht eine Aussage über ihr/sein Zeitmanagement.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	In den Lautdenkprotokollen konnte diese Strategie nicht beobachtet werden.
LD-TM17	Selbstkorrektur F-R	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere Antwort war falsch, die neu gewählte ist richtig.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	Szenario, Lautes Denken, Task C Seite 2: hier heisst es (-) ((schaut auf die antwortoptionen)) für die NACHT (.) ((zeigt mit der maus auf die letzte zeile bei kelly)) dann denke ich doch nicht? [...] ((zieht 3 nach C2_3rev und ersetzt damit 2)) (2.0) SO (Cersei_Szenario, Pos. 164-167)
LD-TM18	Selbstkorrektur R-F	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“	Szenario, Lautes Denken, Task C Seite 1: ist glaub ich (-) ((schaut auf die aufgabenstellung)) NICHT (-) im STADT(-)zentrum. ((wechselt die antwort von C1_2a auf C1_2b))

¹⁴⁰ Die Strategie wurde als konstruktrelevant definiert (d.h. als Testmanagement- und nicht als *test-wiseness*-Strategie), weil in der Aufgabe, bei der diese Strategie vorkam, ein Notizzettel mit vier Lücken mit zwei empfohlenen und zwei nicht empfohlenen Sachen vervollständigt werden sollte. Es ist also wahrscheinlich dem Format geschuldet, dass manche Schüler/innen zu lesen aufhörten, nachdem sie zwei positive oder zwei negative Punkte gefunden hatten. Auch im realen Leben würde man u.U. so vorgehen. Es ist aber ein Grenzfall: Man könnte diese Strategie auch als konstruktirrelevant einstufen.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		Antwort war richtig, die neu gewählte ist falsch.	bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	(--) sondern an der RHOne (Cersei_Szenario, Pos. 151-152)
LD-TM19	Selbstkorrektur F-F	S wechselt die Antwortoption noch einmal, nachdem er/sie bereits eine Antwortoption ausgewählt hat. Die frühere Antwort war falsch, die neugewählte ist ebenfalls falsch.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 2: ((klickt zuerst option C2b an und wechselt anschliessend zu C2c)) (Loup_Off, Pos. 41)
	Lautes Denken: <i>test-wiseness-Strategien</i>	Strategien zur Nutzung von Wissen über die Testformate und andere Randinformationen. Bewusst eingesetzt, um die Testaufgaben zu lösen, ohne die erwarteten linguistischen und kognitiven Prozesse durchlaufen zu müssen.		
LD-TW01	Raten	S rät (intelligent oder zufällig). Er/Sie muss aufgrund des Testformats eine Antwort wählen, ist sich aber nicht sicher, ob die gewählte Antwort auch stimmt. Entweder hat er/sie den Text nicht bzw. nicht aufmerksam genug rezipiert; oder aber nicht (genug) verstanden, um eine Antwort aufgrund von Verstandenem zu wählen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario "Berufswahl", Seite 1: *okey also irgendwie habe ich da nichts gefunden von selber testen aber HIER (-) ((zeigt mit dem cursor auf die dritte nachricht)) [...] (-) ich glaube ich tue mal DAS hier,* ((zieht 3 nach C1_1rev)) (4.2) *ich tippe jetzt einfach mal auf nein;* ((wählt C1_1b, klickt "weiter")) (Dustin_Off, Pos. 89-93) Szenario, Lautes Denken, Task C Seite 2: *manchmal rate ich einfach weil ich keine lust habe das alles zu

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				lesen.* (5.0) (Elena_Szenario, Pos. 88, Gewichtung 5)
LD-TW02	Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war	S wählt eine Antwort aus, weil ein Wort oder mehrere Wörter aus den Antwortoptionen auch im Inputtext in der gleichen oder ähnlichen Form vorkommen. Dabei wird aber die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigt bzw. verstanden. ¹⁴¹	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird sowohl bei den falschen wie auch bei den richtigen Antworten codiert, wenn die Antwort aufgrund eines unpassenden Wortes gewählt wurde, welches in Antwortoption und Inputtext vorkommt; oder wenn es klare Indizien dafür gibt, dass der/die S den Text um das gefundene Wort herum nicht versteht.	<p>Szenario, Lautes Denken, Task C Seite 1: also das Wifi: (4.0) "e:st grAtuit" also ist es GRAtis- ((S klickt C1_1a an)) (Alison_Szenario, Pos. 103-104)</p> <p>Szenario, Lautes Denken, Task C Seite 1: ist das hotel I:m (1.5) zEntrum ((schaut auf den text)) (3.5) das WEIss ich n:ICHt (.) aber ich denke "en general" hOtEl heisst vielleicht (--) das generAlhotel und das generalhotel ist vielleicht ((blickt nach oben)) in der mitte der sTADT also denke ich ((schaut auf die aufgabenstellung)) könnte es sEin; (--) ((S klickt C1_2a an und drückt auf "weiter")) (6.6) (Sheldon_Szenario, Pos. 140-143)</p> <p>Off-Szenario, Lautes Denken, Szenario „Walkies“, Seite 1: dann nehme ich mal an ist es das da weil es ist das einzige wo AUCH mit walkies etwas steht (Dustin_Off, Pos. 18-19)</p>
LD-TW03	Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer	S schliesst oder wählt eine oder mehrere Antwortoption(en) aus bzw. wählt bei „ja/nein-Fragen“ die	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert, wenn	Szenario, Lautes Denken, Task B Seite 2: (3.0) am ersten tag war sie (-) beim sCHLOss- beim zwEiten (1.0) war sie glaube ich KRAnk (und)

¹⁴¹ Die Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchläuft, die durch das Item vorgesehenen sind.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
	negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern	negative Option aus, weil er/sie die entsprechenden Informationen nicht verstanden hat oder weil sie nach ihrem/seinem Verständnis im Inputtext nicht vorkamen. ¹⁴²	1) der/die S eine oder mehrere Antwortoption(en) ausschliesst, weil er/sie im Inputtext die entsprechenden Informationen nicht gefunden oder nicht verstanden hat; 2) das Itemformat das Anklicken von „ja“ oder „nein“ verlangte und der/die S die Option „nein“ wählte, weil er/sie im Text die passende Information nicht gefunden bzw. nicht verstanden hat.	vom rEgen (3.0) (hat) sie glaube ich nichts gesagt- (--) ((S klickt B2c)) (Alison_Szenario, Pos. 88-90) Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 1: "sElber produkte testen"- (---) da habe ich nichts im text gefunden- ((zieht 1 in C1_1rev)) (--) also em (--) (klicke) ich nEin an; (--) ((S klickt C1_1ban)) ja- (Picard_Off, Pos. 82-84)
LD-TW04	Auswahl einer Antwortoption mit bekannten Wörtern	S wählt eine Antwortoption aus, die nur Wörter enthält, die er/sie im Inputtext verstehen konnte.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	In den Lautdenkprotokollen konnte diese Strategie nicht beobachtet werden.
LD-TW05	Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen	S schliesst nach der Rezeption des Inputtextes eine Antwortoption oder mehrere Antwortoptionen aus bzw. wählt eine Antwortoption, weil er/sie diese Antwortoption(en) aufgrund von eigenem Wissen, von eigener Erfahrung oder von eigenen Vorlieben (und nicht aufgrund des Inputtextes) für falsch oder für richtig hält.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, Lautes Denken, Szenario „Stadtripp“, Seite 2: und (.) starker regen (-) ob das so etwas beEindruckendes ist (-) ((verzieht das gesicht)) glaub ich NICHT und (.) in dem fall nehme ich jetzt einfach das ESSen;* ((wählt B2a, klickt "weiter" 2x)) (Dustin_Off, Pos. 64-65) Szenario, Lautes Denken, Task A Seite 1: also ich denke mir (-) wenn ich SELber tipps geben würde was ich für tipps geben würde (--) oder was ich (-) empfehlen würde (Elena_Szenario, Pos. 20-22)

¹⁴² Die Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchläuft, die durch das Item vorgesehen sind. Ausserdem waren die eingesetzten Aufgaben so konstruiert, dass in den Inputtexten jeweils Elemente aller Antwortoptionen vorkamen.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				Szenario, Lautes Denken, Task A Seite 2: *es ist immer soviel TEXT dann muss ich schauen was ich (-) machen möchte.* (Elena_Szenario, Pos. 31; Gewichtung 5)
LD-TW06	Auswahl einer anders aussehenden Antwortoption	S wählt eine Antwortoption, weil sie anders als die restlichen Antwortoptionen aussieht.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	In den Lautdenkprotokollen konnte diese Strategie nicht beobachtet werden.
LD-TW07	Auswahl oder Ausschluss einer Antwortoption aufgrund des Aufgabenformats	S wählt oder schliesst eine Antwortoption aus, weil er/sie bestimmte Vorstellungen davon hat, wie die (Sprach-)Tests normalerweise aufgebaut sein sollten bzw. weil er/sie die Aufgabeninstruktionen nicht aufmerksam genug gelesen hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird z.B. codiert 1) wenn der/die S denkt, dass eine Antwort negativ oder positiv sein muss, wenn die früheren Antworten negativ oder positiv waren; 2) wenn der/die S bei Task C die Aufgabenstellung nicht richtig gelesen hat und dachte, dass alle Rezensionen einer Frage zugeordnet werden müssen.	Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 2: *ich sage in dem fall ne:in- (2.0) dann haben wir das gleiche wie vorhin- (--) vorher ((meint die frühere Bildschirmseite)) hatte ich auch nein-* ((wählt C2_1b an)) (---) ((schaut auf den text und die aufgabenstellung hin und her)) *kann ja nicht zweimal das gleiche sein (Omega_Off, Pos. 148-152) Off-Szenario, Lautes Denken, Szenario „Berufswahl“, Seite 2: hhh (2.5) *also muss dAs [S meint hier die noch nicht zugeordnete Rezension] automatisch dAs sein-* ((S zieht 3 in C2_3rev)) (Mogli_Off, Pos. 137)

8.5.3 CODIERUNG DER STRATEGIEN BEIM *STIMULATED RECALL* (SR)

Mit den untenstehenden (und zum Laut-Denk-Teil parallelen) Codes wurden die Aussagen der Schüler/innen festgehalten, die von ihnen während der anschliessenden *Stimulated-Recall*-Interviews gemacht wurden.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
	Lautes Denken: <i>Lernerstrategien</i>	Gedanken und Handlungen, die von den Sprachlernenden bewusst gewählt und eingesetzt werden, um den fremdsprachigen Text zu verstehen . Diese Strategien sind für eine Testsituation nicht spezifisch, d.h. sie werden auch in anderen Sprachlern- und Sprachgebrauchssituationen eingesetzt.		
SR-L01	Sorgfältiges Lesen	Es geht aus dem Interview hervor, dass der/die S den gesamten Inputtext sorgfältig durchgelesen hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: den ERSTen text habe ich RIChtig durchgelesen (...) dass ich so weiss (-) was er überhaupt FRAGT? (--) (Howard_Szenario, Pos. 272-273) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 1: *ich habe eigentlich immer ABSchnitt für abschnitt (-) zwischendurch habe ich eine pause gemacht (-) halt überlegt wo das HINkommen könnte oder so.* (Bonnie_Off, Pos. 834-836, Gewichtung 5)
SR-L02	Schnelles Lesen	Es geht aus dem Interview hervor, dass der/die S den gesamten Inputtext schnell überflogen hat. Er/Sie hat	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2:

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		dabei evtl. nach Schlüsselwörtern gesucht.		<p>S: zuerst [habe ich] den text nach den schlüsselwörtern (-) überflöGen-</p> <p>MB: okay und das heisst den tExt hast/ also Wie hast du den text gelesen- nUR um zu schauen wo die schlüsselwörter sind` oder (--) hast du Alles gelesen?</p> <p>S: eigentlich ein: wenig bEides; (...) Aber MEHR auf die schlüsselwörter (--) konzentriERT- (Alison_Szenario, Pos. 477-491)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 2: S: so ICH habe französisch nicht so gern (-) und überfliege dann einfach den text (-)</p> <p>MB: okey (-) also (-) du hast die texte NICHT ganz genau gelesen;</p> <p>S: *nein (-) also nur überfliegen* (Elena_Szenario, Pos. 362-373)</p>
SR-L03	Einen Teil des Inputtextes (erneut bzw. genauer) lesen	Es geht aus dem Interview hervor, dass der/die S einen Teil des Inputtextes erneut bzw. genauer durchgelesen hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird auch (in Kombination mit SR-L02) codiert, wenn der/die S sagt, er/sie habe den Text zuerst schnell gelesen und nach Schlüsselwörtern gesucht, um dann die entsprechende Stelle genauer zu lesen.	<p>Szenario, <i>Stimulated Recall</i>, Task C Seite 1: zuerst habe ich (--) einfach das schlüsselwo:rt (-) AH da ist das schlüsselwort dann den tExt ((S bewegt die hand nach oben und unten)) so (--) eigentlich (-) gelesen; (Alison_Szenario, Pos. 1106-1108)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 2: ich HABe ja wirklich schon ein bisschen überflogen und dann erst NACHher noch die (-) speZIfischen</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>textstellen ein bisschen (-) BESSer durchgelesen. (Howard_Szenario, Pos. 512-513, Gewichtung 5)</p> <p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Berufswahl“, Seite 1: S: [(...)ich habe es] überflogen; MB: und wenn du ein schlüsselwort gefunden hast` S: dann habe ich d/ eh (-) die sätze drum herum (---) genau gelesen; ja- (Omega_Off, Pos. 930-932, Gewichtung 5)</p>
SR-L04	Weiterlesen, um Nichtverstandenes zu verstehen	Es geht aus dem Interview hervor, dass der/die S Teile des Inputtextes verwendet hat, um das Verständnis eines bestimmten Teils zu erleichtern: Er/Sie sagt, dass er/sie weitergelesen hat, um nach Informationen zu suchen, die zum Verständnis des bereits Gelesenen beitragen können.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Walkies“, Seite 1: S: also JA ich war mir eben nicht sicher (.) "tableau" da (-) ob das auch: ta/ wirklich tabelle heisst`[...] und (.) deshalb habe ich noch weitergelesen` (Omega_Off, Pos. 403-409)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 2: MB: mh_mh hast du dann Erst am ENde alles eingefügt` oder hast du (das) schon WÄHrend des lesens gemacht-</p> <p>S: manchmal während dem lesen und wenn ich mir nicht sicher wa:r (-) habe ich es dAnn (--) zuerst noch den rEst gelesen` und dAnn erst eingefügt; (Alison_Szenario, Pos. 515-519, Gewichtung 5)</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
SR-L05	Bisheriges Verständnis hinterfragen ¹⁴³	Es geht aus dem Interview hervor, dass der/die S sich fragte, nachdem er/sie mehr gelesen oder gehört hat, ob das bisherige Verständnis in Anbetracht der neuen Informationen noch zuträfe.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Diese Strategie wurde von den SuS in den Interviews nicht erwähnt.
SR-L06	Bisheriges Verständnis korrigieren	Es geht aus dem Interview hervor, dass der/die S das Verständnis des Textes oder eines Teils davon korrigierte, nachdem er/sie den Text erneut gelesen bzw. weitergelesen/weitergehört hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: das SCHLOSS habe ich zuerst gedacht dann sagt ja je n y irais PAS (-) also (-) ne pas heisst ja (-) NICHT; (--) also (-) hab ich jetzt mal ANgenommen dass es NICHT gut ist (Howard_Szenario, Pos. 329-333)
SR-L07	Bisheriges Verständnis bestätigen	Es geht aus dem Interview hervor, dass der/die S das endgültige Verständnis des Inputtextes oder eines Teils davon bestätigte, nachdem er/sie sich damit weiter und/oder vertiefter auseinandergesetzt hatte.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Diese Strategie wurde von den SuS in den Interviews nicht erwähnt.
SR-L08	Intensive Auseinandersetzung mit Teilen des Inputtextes	Es geht aus dem Interview hervor, dass der/die S den Inputtext vorgelesen, wiederholt, paraphrasiert, übersetzt oder zusammengefasst hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: MB: und dann kamst du auf DIESE seite` (-) und was hast du hier gemacht? (-) um diese aufgabe zu lösen? S: ich habe versucht (.) den TEXT zu übersetzen` ((lächelt)) (Cybermen_Szenario, Pos. 312-314) Szenario, <i>Stimulated Recall</i> , Task B Seite 3:

¹⁴³ Strategien, die nur in einem Untersuchungsteil (Lautes Denken oder *Stimulated Recall*) vorkamen, werden im Kategorienleitfaden übersichtshalber in beiden Teilen aufgeführt. In dem Untersuchungsteil, in dem sie nicht beobachtet werden konnten, wurden sie grau markiert und es wurde in der Spalte „Beispiele“ eine entsprechende Anmerkung gemacht.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				S: also ich habe eben wieder dAs gelesen auf: was ich achten mUss eigentlich` und dann wieder während ich den tExt gehört habe- immer wieder wie im kOpf ein wenig übersEtzt was das hEissen könnte- (Alison_Szenario, Pos. 957-960, Gewichtung 5)
SR-L09	Informationen verknüpfen	Es geht aus dem Interview hervor, dass der/die S Teile von gelesenen bzw. gehörten Informationen miteinander in Beziehung gesetzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Stadttrip“, Seite 1: *also habe ich versucht (-) halt das was ich verstehe herauszulesen und (-) zusammenzusetzen zu einem TEXT den ich eigentlich MEHR oder weniger verSTEHe (-) und irgendwie die antwort finde für (-) die fragen.* (Bonnie_Off, Pos. 561-565) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1: S: *also ich habe vor allem nach wörtern gesucht die ich verSTEHe (-) und dann halt den rest je nach dem zusammengereimt (Bonnie_Off, Pos. 349-350; Gewichtung 5)
SR-L10	Unbekannte Wörter, Satz- bzw. Textteile identifizieren	Es geht aus dem Interview hervor, dass der/die S ein unbekanntes Wort, einen unbekannten Satz- bzw. Textteil identifiziert hat oder dass er/sie an seinem/ihrem Verständnis dieses Wortes oder dieser Satz- bzw. Textteile gezweifelt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird auch bei der mündlichen Zusammenfassung codiert, wenn sich dabei herausstellt, dass der/die S ein Schlüsselwort oder eine wichtige Stelle im Inputtext nicht verstanden hat.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2: S: also eigentlich dieses wort "EnnuyEUx" ich weiss nicht das versteHe ich eben nicht- und da war ich mir dann nicht genau sIcher- (Alison_Szenario, Pos. 533-536) Szenario, <i>Stimulated Recall</i> , Task C Seite 1: das wort "cHAMbre" wurde eigentlich in jEdem text benUTzt aber ich weiss leider nicht was es HEISSt- wenn ich es verstEhen würde *würde ich wahrscheinlich ein bisschen

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>mehr verstehen* (Sheldon_Szenario, Pos. 927-929)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: S: JA bei gewissen habe ich schon eher gezög/ [...] ein paar sachen habe ich nicht verSTANDen (-) (Howard_Szenario, Pos. 324-327, Gewichtung 5)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: dann sagt er (.) "je n IRAI pas-" (-) und ich weiss jetzt nicht was "irai" heisst (Cybermen_Szenario, Pos. 413-414, Gewichtung 25)</p>
SR-L11	Unbekannte Wörter ausblenden	Es geht aus dem Interview hervor, dass der/die S sich nicht durch unbekannte Wörter ablenken liess und diese ausgeblendet hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Szenario, <i>Stimulated Recall</i>, Task C Seite 2: hier steht eben (1.5) "il y avait une connexion internet sans fil"; ich habe einfach wirklich NICHT verstanden was jetzt "sans fil" heisst (.) heisst OHne nönönö und em ich habe jetzt trotzdem mal auf JA geklickt weil ich dachte (1.5) e:: "il y avait une connexion internet" und das sagt eigentlich es hat eine (-) verbindung zum internet ohne etwas (.) aber es HAT eine- (Arya_Szenario, Pos. 1453-1459)</p> <p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Berufswahl“, Seite 1: und dann versuche ich es (-) [...] herauszufinden (-) und wenn ich es WIRklich nicht herausfinde dann (-)</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				RATe ich einfach und (-) oder LASSe es einfach sein und ja. (Loup_Off, Pos. 1041-1043, Gewichtung 5)
SR-L12	Wortbedeutung identifizieren	Es geht aus dem Interview hervor, dass der/die S die Bedeutung eines Wortes z.B. aufgrund des Kontextes identifiziert hat bzw. dass er/sie sich an die Wortbedeutung erinnert hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1: MB: und was MACHST du dann wenn du ein WORT zum beispiel (-) NICHT verstehst oder einen SATZ mitten im text? S: *also dann versuche ich halt einfach mit den anderen wörtern die ich (-) verstehe versuche ich dann irgendwie darauf zu KOMMen?* (--) (Eleven_Off, Pos. 449-452, Gewichtung 5)
SR-L13	Morphosyntaktische Regeln anwenden	Es geht aus dem Interview hervor, dass der/die S explizit seine/ihre Kenntnisse der morphosyntaktischen Regeln genutzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: MB: woher WUSStest du dass sie das empfehlen? S: ah dass es keine verNEINung hat (-) also in dem satz; (Elena_Szenario, Pos. 239-240) Szenario, <i>Stimulated Recall</i> , Task C Seite 2: MB: hat es dich gestÖrt dass das die gleichen fragen sind? die noch einmal gekommen sind- S: vielleicht wenn man (aus der) vOrherigen aufgabe gesehen dass man/ dass es noch so: (--) TRICKS dabei hat wie zum beispiel das "ne PAS" mit vernEinung (-) dass man nAchher weiss dass mAn (-) es bEsser machen kann- (Sheldon_Szenario, Pos. 973-980, Gewichtung 5)

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
SR-L14	Interpunktionskenntnisse nutzen	Es geht aus dem Interview hervor, dass der/die S seine/ihre Kenntnisse von Interpunktionszeichen genutzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: zum beispiel "OUI" dann (mit einem) ausrufezeichen (-) kann man sich schon dEnken dass es dann (ein) (--) pOsitiv ist` also dass sie das empfEHLEn- (Alison_Szenario, Pos. 331-333)
SR-L15	Kognaten identifizieren und nutzen	Es geht aus dem Interview hervor, dass der/die S im Text (vermeintliche) Kognaten identifiziert und genutzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: MB: und wie BIST du drauf gekommen? S: *keine ah/ parallelwörter ; MB: hm_hm (-) WELChE, S: das; ((zeigt mit maus auf "burgers")) MB: was IST das für ein parallelwort, S: schloss (-) burg ; (Elena_Szenario, Pos. 222-227, Gewichtung 10) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1: ich versuche ALLes zu verstehen und ähm (-) so halt auch nach paraLLELwörtern zu suchen (--) (Nancy_Off, Pos. 331-332, Gewichtung 5)
SR-L16	Prosodie beim Hören nutzen	Es geht aus dem Interview hervor, dass der/die S beim Hören auf die Prosodie geachtet und diese Informationen genutzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 2: und dann wusste ich okay REgen; (---) ich weiss nicht was regen IST aber für die meisten menschen ist es ja etwas NEGAtives` (...) dann wusste ich okay (---) es ist etwas NEGAtives und vielleicht (-) hört man das dann auch an ihrer stimme also oder an ihrem AUSdruck- auch wenn ich jetzt das wort nicht verstehe- (Arya_Szenario, Pos. 1103-1110)

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
SR-L17	Relevantes Hintergrundwissen einbeziehen	Es geht aus dem Interview hervor, dass der/die S relevantes Hintergrundwissen identifiziert und es dann in geeigneter Weise genutzt hat, d.h. er/sie hat <i>zusätzlich</i> zu den Informationen aus dem Text auf sein/ihr Weltwissen zurückgegriffen, um die Textbedeutung zu verstehen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2: am ANfang wurde halt gerade die: (--) cArte Avignon (-) erwähnt` und ich dEnke das ist so eine halt (-) wie eine TAGESkarte wo man mit allem fahren kann- (Sheldon_Szenario, Pos. 371-373)
SR-L18	Selbsteinschätzung	Es geht aus dem Interview hervor, dass der/die S beim Lauten Denken eine selbsteinschätzende (positive oder negative) Aussage in Bezug auf die eigenen Französischkenntnisse bzw. das Textverständnis gemacht hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert	Diese Strategie wurde von den SuS in den Interviews nicht erwähnt.
	Lautes Denken: <i>Testmanagementstrategien</i>	Strategien für eine sinnvolle Reaktion auf die Testaufgabe(n) . Bewusst und zielorientiert gewählt, um eine richtige Antwort zu produzieren. Sie sind typisch für eine Testsituation.		
SR-TM01	Instruktionen und/oder Antwortoptionen (erneut) lesen	Es geht aus dem Interview hervor, dass der/die S die Instruktionen und/oder Antwortoptionen (erneut) durchgelesen hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 1: MB: was hast du HIER zuerst gemacht? S: also zuerst einma:l dIese durchgelesen und dann gehört- MB: die fragen` S: jA. (Alison_Szenario, Pos. 729-732) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Stadttripp“, Seite 1:

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				*also ich habe wieder oben gelesen (-) ich mache eigentlich immer das gleiche system (-) dann halt die FRAGE und dann (-) die ANTworten (-) (Bonnie_Off, Pos. 541-543, Gewichtung 5)
SR-TM02	Paraphrasieren und/oder Übersetzen der Aufgabe	S sagt oder es geht aus dem Interview hervor, dass er/sie die Aufgabe paraphrasiert oder die Frage, die Antwortoptionen oder ihre Teile für sich übersetzt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Stadttripp“, Seite 2: *und dann habe ich (-) die anderen fragen gelesen und dann habe ich mir je nach dem auch überlegt (-) was das vielleicht auf französisch oder so heissen könnte? * (Bonnie_Off, Pos. 723-725) Szenario, <i>Stimulated Recall</i> , Task B Seite 1: S: es ist vor allem beim lesen (...) extREM wichtig dass man (-) (...) überlegt (--) so ein bisschen was könnte es so das so:: HEISSen auf französisch kurz, MB: HAST du das gemacht; S: ja (-) einfach kurz überlegt (-) gelesen überlegt was es sein könnte (Howard_Szenario, Pos. 901-907, Gewichtung 5)
SR-TM03	Unbekanntes Vokabular in der Aufgabenstellung oder in den Antwortoptionen identifizieren	Es geht aus dem Interview hervor, dass der/die S die Aufgabenstellung bzw. die Antwortoptionen gelesen und das ihm/ihr darin unbekannte Vokabular identifiziert hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert, wenn S darüber berichtet, dass er/sie 1) in der Aufgabenstellung ein deutsches bzw. 2) bei Off-Szenario in der Aufgabe „Berufswahl“ ein französisches Wort nicht verstanden hat, oder 3) nicht wusste, wie das Vokabular aus der Aufgabenstellung auf Französisch heissen würde.	Szenario, <i>Stimulated Recall</i> , Task B Seite 2: MB: okey (-) als du dir die optIOnen angeschaut hast bevor du das gehört hast (-) hattest du das gefühl dass du diese [...] drei themen erKENNen würdest wenn das im TEXT vorkommt? S: also (...) der starke REgen in der stadt (-) ich kenne das wort für REgen nämlich (-) nicht (--) glaube ich also (-) JA es ist mir nicht

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>EINGefallen (Howard_Szenario, Pos. 1046-1054)</p> <p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Walkies“, Seite 2: S: diesen ((zeigt mit der hand auf den bildschirm)) punkt habe ich nicht genau verstanden (...)</p> <p>MB: die texte der anderen korrektURlesen;</p> <p>S: ja- (...) [ich habe] es schon DA ((zeigt mit der hand auf die antwortmöglichkeiten)) nicht verstand/ (...) nicht genau gewusst was dort gemeint ist- (Omega_Off, Pos. 542-549)</p>
SR-TM04	Vorhersage machen	Es geht aus dem Interview hervor, dass der/die S eine Vorhersage gemacht hat. Er/Sie hat sich vor der Rezeption des Inputtextes überlegt, welche Antwortoption(en) evtl. stimmen und welche ausgeschlossen werden könnten, er/sie hat aber diese Antwortoption(en) vor der Textrezeption noch nicht ausgeschlossen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wenn S berichtet, dass er/sie eine Antwort von Anfang an (d.h. vor dem Lesen/Hören) ausgeschlossen hat, wird nicht TM04, sondern TW05 codiert.	<p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Walkies“, Seite 1: ich habe zuerst (-) die fragen gelesen (...) dann überlegt (--) ob ich eine eigentlich schon ausschliessen könnte mit irgendwelchen informationen oder keine ahnung die (-) wenn es wissensfragen auch ein bisschen gewesen wären` aber waren es nicht` deshalb habe ich nachher (--) mal gelesen (Omega_Off, Pos. 356-362)</p> <p>Szenario, <i>Stimulated Recall</i>, Task B Seite 3: S: erst die fragen gelesen (-) und dann (-) habe ich mir vorge/ also (-) gefragt was ich da (-) erzählen würde (--) oder auch nicht (--) ja</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>(--) dann habe ich den TEXT angehört;</p> <p>MB: das hast du jetzt schon zum DRITTen mal (-) gemacht (-) dass du dir dann vorgestellt hast was DU an der stelle machen würdest (-)</p> <p>machst du es IMMer so?</p> <p>S: *ja (-) und dann ist es schon einfacher (Elena_Szenario, Pos. 913-920, Gewichtung 5)</p>
SR-TM05	Frage und Antwortoptionen beim Hören anschauen	Es geht aus dem Interview hervor, dass der/die S sich die Frage und/oder Antwortoptionen beim Hören angeschaut hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Szenario, <i>Stimulated Recall</i>, Task B Seite 2: S: ich habe auch wieder dIe geLesen` und dann (-) während dem höRen (--) wieder verglichen ob das jetzt stImmt- (Alison_Szenario, Pos. 848-849)</p> <p>Szenario, <i>Stimulated Recall</i>, Task B Seite 1: während ich es ANgehört habe (...) habe ich immer die optionen wieder kurz angeschaut (Howard_Szenario, Pos. 908-910, Gewichtung 5)</p>
SR-TM06	Fokus auf relevante Textteile	Es geht aus dem Interview hervor, dass der/die S die Frage und/oder Antwortoption(en) gelesen und sich beim Hören/Lesen auf die Teile des Inputtextes konzentriert hat, die für die Beantwortung der Frage von Bedeutung sind. Er/Sie hat gesucht, wo im Inputtext die entsprechenden Informationen (vermutlich) vorkommen und hat evtl. die als nicht relevant eingestuften Teile des Inputtextes ignoriert.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: S: em: (3.0) ich habe einfach die informatIOnEN die ich brauche einfach lAut überlegt ob das sein könnte- (...) und die frA:gen (--) die schlÜSSelwörter probiert herauszufinden- (Alison_Szenario, Pos. 204-207)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: habe ich eigentlich kurz (-) schnell [...] einfach SCHNELL durchgelesen so; (--) so die</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				SCHLÜSSELwörter so. (Howard_Szenario, Pos. 302-305, Gewichtung 5)
SR-TM07	Bedeutung eines Textteils mithilfe der Frage oder der Antwortoptionen klären	Es geht aus dem Interview hervor, dass der/die S die Bedeutung von bestimmten Teilen des französischen Inputtextes mithilfe der auf Deutsch vorhandenen Frage bzw. der Antwortoptionen geklärt oder bestätigt hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Berufswahl“, Seite 1: und dann viel REIsen eben voyage (-) wo ich also (-) dank DANK dem DEUTSCHen (-) reisen habe ich mich erinnert dass voyage (-) reisen heisst. (Loup_Off, Pos. 907-908)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: manchmal sind die tExte ein wenig schwierig zu verstehen manche WÖRter- dann habe ich zuerst ein wenig DIE da hier (-) gelesEn` [= die antwortoptionen]- und dAnn probiert das im text herauszulEsen; (Alison_Szenario, Pos. 259-267, Gewichtung 5)</p>
SR-TM08	Einsatz von Verstandenem bei Auswahl/Ausschluss der Antwortoption(en)	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption bzw. -optionen aufgrund von etwas ausgewählt oder ausgeschlossen hat, was er/sie vom Inputtext verstanden hat.	<p>Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.</p> <p>Wird auch codiert, wenn der/die S zwar etwas nicht richtig verstanden hat, aber auf das (falsch) Verstandene bei der Auswahl oder beim Ausschluss einer Antwortoption zurückgegriffen hat.</p>	<p>Szenario, <i>Stimulated Recall</i>, Task C Seite 1: beim LETZten abschnitt eigentlich` habe ich wie verstanden dass es im: (--) hERzen der stadt ist` also im: habe ich dann Angenommen dass es im zEntrum ist- (Alison_Szenario, Pos. 1123-1126)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 2: *und karte in avignon kaufen °hh glaube das haben sie einfach VO:Rgeschlagen aber es hatte glaube (.) keine mehr;* (Boebu_Szenario, Pos. 816-817, Gewichtung 15)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: habe es dann so ein bisschen (aus) (--)sortiert also wAs ich zum</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				beispiel sicher weiss (Aria_Szenario, Pos. 386-387, Gewichtung 5)
SR-TM09	Aufhören, den Inputtext zu rezipieren, nachdem relevante Informationen gefunden wurden	Es geht aus dem Interview hervor, dass der/die S den Inputtext nicht weiterlas bzw. -hörte, nachdem er/sie die (vermeintlich) für die Beantwortung der Frage benötigten Informationen im Text gefunden hatte.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Stadttrip“, Seite 2: S: dann dachte ich ja (--) DAS, MB: muss DAS sein; S: ja (-) und dann habe ich halt eigentlich AUFgehört (-) zu hören und (-) das ist nicht so (-) gut (-) ich hätte lieber FERTIG hören sollen (-) ja. (Loup_Off, Pos. 748-751)</p> <p>Szenario, <i>Stimulated Recall</i>, Task C Seite 1: MB: Und nachdem du das gefunden hast (-) wonach du gesucht hast (-) hast du dann NICHT weiter gelesen- oder doch-</p> <p>S: m: nEin eigentlich meistens nicht. (Alison_Szenario, Pos. 1101-1103, Gewichtung 5)</p>
SR-TM10	Richtigkeit der Antwortoption überprüfen	Es geht aus dem Interview hervor, dass der/die S zur Frage und/oder zum Inputtext zurückging, um zu prüfen, ob die in Erwägung gezogene Antwortoption korrekt wäre oder nicht. Er/Sie machte das, bevor oder nachdem er/sie eine Antwortoption wählte gewählt hatte.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	<p>Szenario, <i>Stimulated Recall</i>, Task C Seite 2: dann habe ich (-) perSONen eingeordnet dann (-) TEXTe noch einmal durchgelesen um sicher zu werden; (Elena_Szenario, Pos. 1100-1101)</p> <p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Berufswahl“, Seite 1: S: [*ich habe es zuerst schonmal] REINGetan und dann (-) habe ich es am SCHLUSS nochmal angeschaut weil wenn ich vielleicht noch etwas gefunden hätte was vielleicht falsch sein könnte hätte ich es</p>

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				noch so ändern können.* (Eleven_Off, Pos. 829-831, Gewichtung 5)
SR-TM11	Antwortoptionen miteinander vergleichen	Es geht aus dem Interview hervor, dass der/die S die möglichen Antwortoptionen miteinander verglichen hat.	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 2: ich glaube da habe ich so ein bisschen ABgewogen was: sinnvoller ist- also ich habe dann gewusst dass es DAS sicher nicht ist` da war ich ein bisschen UNsicher weil sie war noch (irgendwie) krA:nk` (Aria_Szenario, Pos. 979-982)
SR-TM12	Auswahl oder Ausschluss der Antwortoption(en) aufgrund von Inferenzen	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption auf der Grundlage von Inferenzen vom Inputtext gewählt hat. Kommt oft bei jenen Items vor, bei denen die Antwort nicht wortwörtlich im Text steht.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2: MB: und woher wUsstest du dass sie das empfehlen? ((S blickt auf den bildschirm)) (10.6) S: also er hat eigentlich schon geSAGt wo man sie [d.h. «Carte Avignon»] fIndet also denke ich (-) er empfIEHlt es auch` (Alison_Szenario, Pos. 548-551) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 2: und dann habe ich gedacht eben das proDUKT ist das surfbrett und das will er so gerade testen. (Eleven_Off, Pos. 1042-1043, Gewichtung 20)
SR-TM13	Korrektes Matching zwischen Antwortoption und Inputtext, das durch das Itemformat vorgesehen war	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, weil er/sie im Inputtext korrekterweise ein passendes Wort oder passende Wörter zu den Schlüsselwörtern aus dieser Antwortoption identifiziert hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task C Seite 2: S: äh also jetzt °h beim PREIS sehe ich ja (-) direkt fünfunddreissig fünfundzwanzig euro also es ist (-) habe ich KELLY genommen ja (-) also weil es (-) das PASSende Hotel weil es unter fünfunddreissig ist (-) (Howard_Szenario, Pos. 1502-1505)

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		Kommt oft bei Items vor, die auf selektives Lesen (<i>search reading</i>) abzielen.		
SR-TM14	Auswahl einer Antwortoption, weil andere nicht plausibel sind	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption bzw. -optionen ausgewählt, nachdem er/sie andere Antwortoptionen ausgeschlossen hatte, die ihm/ihr basierend auf seinem/ihrem Verständnis vom Inputtext oder eines Teils davon als nicht plausibel erschienen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 2: S: also so wie ich es verstAnden habe war sie am zWEiten tag krank- Und vom REGen habe ich eigentlich (--) nIchts gehört- Und am ersten tag war sie beim schloss- (Alison_Szenario, Pos. 853-855)
SR-TM15	Kenntnisse über das Itemformat nutzen	Es geht aus dem Interview hervor, dass der/die S beim Szenario „Stadttrip“, Task A die Kenntnisse über das Itemformat nutzte, um eine oder mehrere Antwortoptionen zu wählen. Wenn in Szenario „Stadttrip“, Task A z.B. zwei positive Antworten gefunden worden waren, dann wusste der/die S., dass die restlichen Antworten negativ sein müssen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2: *und stadtrundgang hatte ich eben auch nichts gelesen; von stadtrundgang-* MB: dann (.) hast du es einfach eingefügt weil du dachtest dass das andere (.) dort passt- war das so? S: <<nickend> ja > (Boebu_Szenario, Pos. 833-838) Szenario, <i>Stimulated Recall</i> , Task A Seite 1: *also ja wenn ich weiss dann was stimmt tue ich einfach das richtige hin und [...] ja das was übrig bleibt ist dann falsch.* (Elena_Szenario, Pos. 265-267, Gewichtung 5)
SR-TM16	Zeitmanagement	S macht eine Aussage über ihr/sein Zeitmanagement während der Testbewältigung.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 1: *also weil ich nicht wusste ob ich (--) noch zEit habe um etwas herAUSzuhören` oder ob es so FERTig ist und dann habe me:hr über dAs

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>nachgedacht anstatt zuzuhören;* (Aria_Szenario, Pos. 832-835)</p> <p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: vielleicht wenn man (---) jetzt eine zEITvorgabe hat und weiss wie VIEL (--) () Aufgaben DASS es hAt (---) könnte man auch denken wie viel zEIT dass man pro aufgabe hAt- und dann hätte ich mir auch (---) die vOrgegebene zEit genommen (Alison_Szenario, Pos. 404-407, Gewichtung 5)</p>
SR-TM17	Selbstkorrektur F-R	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass die frühere Antwort falsch und die neu gewählte richtig war.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	<p>Szenario, <i>Stimulated Recall</i>, Task A Seite 1: MB: [...] hast du dann soFORT die sachen eingefügt (-) in die tabelle oder erst nachdem du ALLes (-) einmal gelesen hast; S: äh das war der FEHler ich <<lachend>habe> ich habe es äh soFORT eingefügt und NICHT fertig gelesen eben und dann eins zwei mal war es wieder dann ANders im zweiten satz? MB: dann hast du dich sofort korriGIERT; S: und dann ja: (-) DAS das ich gemerkt habe (Toby_Szenario, Pos. 457-463)</p>
SR-TM18	Selbstkorrektur R-F	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass	Wird aufgrund des Videos und/oder der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	Diese Strategie wurde im Interviewteil nicht codiert.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		die frühere Antwort richtig und die neu gewählte falsch war.		
SR-TM19	Selbstkorrektur F-F	Es geht aus dem Interview hervor, dass der/die S die Antwortoption noch einmal wechselte, nachdem er/sie bereits eine Antwortoption ausgewählt hatte. Der Vergleich mit dem Lautdenkprotokoll zeigt, dass die frühere Antwort falsch und die neu gewählte ebenfalls falsch war.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	Diese Strategie wurde im Interviewteil nicht codiert.
SR-TM20	Selbstkorrektur beim SR	S merkt beim <i>Stimulated-Recall</i> -Interview, dass er/sie beim nochmaligen bzw. genaueren Lesen oder Hören nun mehr oder etwas anderes versteht. Er/Sie würde seine/ihre Antwort evtl. nachträglich korrigieren bzw. ist sich nicht mehr sicher, ob die früher gewählte Antwort tatsächlich stimmt. Es wird hier nicht festgehalten, ob die korrigierte Antwort richtig oder falsch wäre.	Wird nur bei den Teilen codiert, die als „richtig“ oder „falsch“ bewertet wurden (siehe Erklärungen zur Codierung von „richtig/falsch“ in Kap. 3.7.3.1).	Szenario, <i>Stimulated Recall</i> , Task A Seite 1: ich glaube das ist eigentlich nicht empfinden- ((S lacht)) sehe ich gerade- (...) dann denke ich habe ich es eigentlich falsch (-) angekreuzt hier- (Alison_Szenario, Pos. 345-348) Szenario, <i>Stimulated Recall</i> , Task B Seite 2: ((S hört weiter)) äh AH ((fast sich mit der hand an die stirn)) ich habe das CHateau durch das fenster vom BUS gesehen (-) hat sie gesagt (-) ah nein shit ((lacht)) das habe ich gar nicht richtig (-) ja. (Howard_Szenario, Pos. 1089-1092)
	Lautes Denken: <i>test-wiseness-Strategien</i>	Strategien zur Nutzung von Wissen über die Testformate und andere Randinformationen. Bewusst eingesetzt, um die Testaufgaben zu lösen, ohne die erwarteten linguistischen und kognitiven Prozesse durchlaufen zu müssen.		

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
SR-TW01	Raten	Es geht aus dem Interview hervor, dass der/die S geraten hat (intelligent oder zufällig). Er/Sie musste aufgrund des Testformats eine Antwort wählen, war sich aber nicht sicher, ob die gewählte Antwort auch stimmt. Entweder hat er/sie den Text nicht bzw. nicht aufmerksam genug rezipiert; oder aber nicht (genug) verstanden, um eine Antwort aufgrund von Verstandemem zu wählen.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task A Seite 2: da habe ich einfach ein wenig gerATEN- (Alison_Szenario, Pos. 560) Szenario, <i>Stimulated Recall</i> , Task A Seite 1: *MA:NCHmal habe ich geraten; * (Boebu_Szenario, Pos. 713, Gewichtung 5)
SR-TW02	Korrektes oder unkorrektes Matching zwischen Inputtext und Antwortoption(en), das durch das Itemformat nicht vorgesehen war	Es geht aus dem Interview hervor, dass der/die S eine Antwort ausgewählt hat, weil ein Wort oder mehrere Wörter aus den Antwortoptionen auch im Inputtext in der gleichen oder ähnlichen Form vorkamen. Dabei wurde aber die Bedeutung des ganzen Inputtextes oder seiner Teile nicht mitberücksichtigt bzw. verstanden. ¹⁴⁴	Wird aufgrund der entsprechenden Bemerkung(en) der/des S. codiert. Wird sowohl bei den falschen wie auch bei den richtigen Antworten codiert, wenn die Antwort aufgrund eines unpassenden Wortes gewählt wurde, welches in Antwortoption und Inputtext vorkommt; oder wenn es klare Indizien dafür gibt, dass der/die S den Text um das gefundene Wort herum nicht verstanden hat.	Szenario, <i>Stimulated Recall</i> , Task C Seite 1: habe ich zum beispiel speziell das wOrt Wlan- eigentlich- und dann habe ich hier Wifi (-) ist gratis (-) an der rezeptION- (...) ja; jetzt hätte ich noch WEiterlesen sollen wEil ja unten steht dass es im zIMmer fünfzehn euro mEHr kostet (-) in der woche` (Alison_Szenario, Pos. 1069-1074) Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 2: *also er ((meint mattieu)) hat da unten in der letzten (-) zeile hat er wieder etwas von (-) TESTer (-) also geredet (-) [...] und dann habe ich gedacht das könnte es noch sein.* (Bonnie_Off, Pos. 985-990)

¹⁴⁴ Die Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchlief, die durch das Item vorgesehenen sind.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				<p>Szenario, <i>Stimulated Recall</i>, Task A Seite 2: ICH habe französisch nicht so gern (-) und überfliege dann einfach den text (-) und dann wenn ich ein WORT finde das auch in der antwort steht (-) dann tue ich es einfach hin;* (Elena_Szenario, Pos. 362-364, Gewichtung 5)</p>
SR-TW03	Ausschluss einer Antwortoption bzw. mehrerer Antwortoptionen oder Auswahl einer negativen Antwortoption mit nicht rezipierten oder nicht verstandenen Wörtern	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption oder mehrere Antwortoptionen ausgeschlossen, ausgewählt bzw. bei „ja/nein-Fragen“ die negative Option angeklickt hat, weil er/sie die entsprechenden Informationen nicht verstanden hat oder weil sie nach ihrem/seinem Verständnis im Inputtext nicht vorkam(en). ¹⁴⁵	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert, wenn 1) der/die S sagte, dass er/sie eine oder mehrere Antwortoption(en) ausgeschlossen hat, weil er/sie im Inputtext die entsprechenden Informationen nicht gefunden oder nicht verstanden hat; 2) das Itemformat das Anklicken von „ja“ oder „nein“ verlangte und der/die S die Option „nein“ wählte, weil er/sie im Text die passende Information nicht gefunden bzw. nicht verstanden hat.	<p>Szenario, <i>Stimulated Recall</i>, Task B Seite 2: Und vom REGen habe ich eigentlich (--) nIchts gehört- (Alison_Szenario, Pos. 854)</p> <p>Off-Szenario, <i>Stimulated Recall</i>, Szenario „Walkies“, Seite 2 S: ich habe nicht genau gewusst was dort gemeint ist- (...)</p> <p>MB: mhm` (--) hast du es dann auch im text nicht gefunden-</p> <p>S: aha nein- (--) also (--) ich habe mich auch nicht sonderlich konzentriert auf DAS- [meint damit die Antwortoption mit nicht verstandenen Wörtern] (-) ich habe mich vor allem auf die anderen beiden konzentriert; (Omega_Off, Pos. 549-554)</p> <p>Szenario, <i>Stimulated Recall</i>, Task C Seite 3: MB: ach so wenn du (-) NICHTS gefunden hast (-) dann hast du NEIN angekreuzt; S: ja;</p>

¹⁴⁵ Die Strategie ist konstruktirrelevant, weil der/die S bei ihrem Einsatz nicht die kognitiven Prozesse durchläuft, die durch das Item vorgesehenen sind. Ausserdem waren die eingesetzten Aufgaben so konstruiert, dass in den Inputtexten jeweils Elemente aller Antwortoptionen vorkamen.

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
				MB: okey (-) hast du es IMMer so gemacht? S: ja ; (Elena_Szenario, Pos. 1255-1259, Gewichtung 5)
SR-TW04	Auswahl einer Antwortoption mit bekannten Wörtern	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, die nur Wörter enthält, die er/sie im Inputtext verstehen konnte.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 2: *aber da (-) die frage habe ich nicht ganz verstanden? * (--) ((S meint die frage "noemie, ça va aller tout seule?" in e , wo auch die richtige antwort war)) *und dann habe ich gedacht ich nehme das oben;* ((S meint c))(Bonnie_Off, Pos. 478-480)
SR-TW05	Ausschluss oder Auswahl von Antwortoption(en) aufgrund von Hintergrundwissen	Es geht aus dem Interview hervor, dass der/die S nach der Rezeption des Inputtextes eine Antwortoption oder mehrere Antwortoptionen ausgeschlossen bzw. eine Antwortoption ausgewählt hat, weil er/sie diese Antwortoption(en) aufgrund von eigenem Wissen, von eigenen Erfahrungen oder von eigenen Vorlieben (und nicht aufgrund des Inputtextes) für falsch oder für richtig hielt.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 2: dann hat sie gesagt dass es halt am ZWEIten tag geregnet hat (--) und das habe ich (-) dann habe ich so gedacht JA: aber das kann sie ja nicht beeindrucken (-) habe ich gedacht? (Nancy_Off, Pos. 624-625) Off-Szenario, <i>Stimulated Recall</i> , Szenario "Walkies", Seite 1: wenn ich finde dass em (-) dass ich etwas nicht wEIss` dann [...] schAue ich welche am meisten sinn gibt- oder welche am: (2.0) also am (-) idealsten für die situtation ist; ja; (Picard_Off, Pos. 227-232, Gewichtung 5)
SR-TW06	Auswahl einer anders aussehenden Antwortoption	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt hat, weil sie anders als die	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Kommt bei Off-Szenario Task „Walkies“ vor, wo bestimmte Nachrichten in	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1: ausserdem ist das (...) wie (...) meine sprechblase da ist et/

Code	Kurzbezeichnung der Strategie	Definition der Strategie	Hinweise für Codierende	Beispiele
		restlichen Antwortoptionen aussah.	graphisch anders aussehenden Sprechblasen dargestellt waren.	die meinung da steht deine aufgabe und so dann- (Omega_Off, Pos. 417-419)
SR-TW07	Auswahl/Ausschluss einer Antwortoption aufgrund des Aufgabenformats	Es geht aus dem Interview hervor, dass der/die S eine Antwortoption ausgewählt oder ausgeschlossen hat, weil er/sie bestimmte Vorstellungen davon hatte, wie die (Sprach-)Tests normalerweise aufgebaut sein sollten bzw. weil er/sie die Aufgabeninstruktionen nicht aufmerksam genug gelesen hatte.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert. Wird z.B. codiert 1) wenn der/die S dachte, dass eine Antwort negativ oder positiv sein muss, wenn die früheren Antworten schon negativ oder positiv waren; 2) wenn der/die S bei Task C die Aufgabenstellung nicht richtig gelesen hat und dachte, dass alle Rezensionen einer Frage zugeordnet werden müssen.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Berufswahl“, Seite 1: und dann habe ich einfach NEIN angekreuzt (-) und weil es schon zweimal JA ist (-) habe ich noch nein; * ((lacht)) (Dustin_Off, Pos. 623-624) Szenario, <i>Stimulated Recall</i> , Task C Seite 2: MB: hm_hm (-) weisst du noch wo du deine sachen geFUNDen hast? S: ja also (-) HIER die nacht HIER [fünfundzwanzig für die nacht,] MB: [bei kelly?] S: ja (...) MB: das ist aber bei kelly (-) du hast LYLE [gewählt]; (-) hast du bei LYLE noch etwas ANderes gehabt oder [hast du das falsche reingezogen;] (...) S: ich dachte (-) es muss in jedem TEXT sein MB: und du hast (-) trotzdem LYLE genommen; S: ja weil (.) der noch Übrig war. (Cersei_Szenario, Pos. 1154-1186)

8.5.4 CODIERUNG DER WEITEREN THEMATISCHEN KATEGORIEN

Codierung im Teil „Lautes Denken“				
LD-SC_Entd	Entdeckung des Szenarios	S macht eine spontane Anmerkung, die vermuten lässt, dass er/sie beim LD das Szenario entdeckt bzw. wahrgenommen hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Intro Task D: ((liest laut vor)) am wochenende in dem du in avignon bist findet= *es geht ja wieder um das das gleiche* (Aria_Szenario, Pos. 243)
LD-Sonstiges	Sonstiges	Sonstige interessante Stellen in den Lautdenkprotokollen, z.B. Äusserungen zur Testsituation, zu ihrer Motivation etc.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Lautes Denken, Task C Seite 1: *wenn es manchmal so viel text hat denkt man so (---) oh NEIN das muss ich jetzt lesen.* (--)*und dann noch auf französisch.* (Elena_Szenario, Pos. 68-70)
Codierung im Teil „Stimulated Recall“				
SR-SC_1	Szenario beim Test wahrgenommen	S erwähnt spontan die thematische Einbindung der Aufgaben in ein gemeinsames Szenario.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, <i>Stimulated Recall</i> , Task B Seite 1: *aber (sonst) glaube ich das geht jetzt alles mehr um EIN thema` Avignon;* MB: (das ist) dir AUfgefallen` dass das (--) alles um das gleiche thema geht- S: ja; (Aria_Szenario, Pos. 746-749, spontane Äusserung)
SR-SC_2	Szenario nach einem Hinweis wahrgenommen	Dieser Code wird vergeben, wenn: 1) das Szenario (bzw. Avignon, Reisen etc.) von der/dem S erwähnt wird, nachdem der/die S gefragt wurde, ob ihm/ihr <i>thematisch</i> etwas an den Aufgaben aufgefallen ist; oder 2) nachdem die Forscherin verraten hat, dass die	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	SC-Abschluss, <i>Stimulated Recall</i> : MB: okey (-) und vom THEma her ist dir etwas AUfgefallen hier, S: SO von (-) von REIsen und so (Howard_Szenario, Pos. 1999-2000) SC-Abschluss, <i>Stimulated Recall</i> : MB: Und vom thEma her (-) ist dir etwas speziElles aufgefallen vielleicht`

		Aufgaben thematisch verbunden sind und der/die S im Nachhinein zugibt, dass es ihm/ihr jetzt aufgefallen ist bzw. eigentlich schon früher „unbewusst“ aufgefallen ist.		<p>S: Em um rEisen geht es und veranstaltungen (Alison_Szenario, Pos. 1608-1609)</p> <p>SC-Abschluss, <i>Stimulated Recall</i>:</p> <p>MB: ist dir etwas speZIELLes aufgefallen (-) bei diesen aufgaben?</p> <p>S: nein (-) eigentlich nicht [...]</p> <p>MB: ist dir (-) vom THEma her etwas aufgefallen?</p> <p>S: ja einfach LEsen (-) und eben mit den personen [...]</p> <p>ja-</p> <p>MB: okey (-) ähm (-) ich FRAge (-) weil eigentlich geht es in ALlen diesen aufgaben um avi/ um avignon.</p> <p>S: ja avignon [ja geNAU;]</p> <p>MB: [es kommt immer wieder] (-) es kommt immer wieder vor so als (-) oberTHEma (-) ist das dir AUFGefallen oder-</p> <p>S: ja (-) eigentlich schon dass immer (-) avignon (-) aber ich es hat mir ist mal nicht in den sinn gekommen dass das das gleiche (-) war;</p> <p>MB: also du hast es nicht (-) beWUSST;</p> <p>S: nein (-) aber ich habe schon immer ah avignon avignon (-) ja; (Cersei_Szenario, Pos. 1355-1372, Gewichtung 30)</p>
SR-Off-SC	Verschiedene Themen stören bei Off-Szenario	Nachdem die Forscherin fragte, ob die S lieber Aufgaben zu einem übergreifenden Thema lösen	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	<p>Off-Abschluss, <i>Stimulated Recall</i>:</p> <p>MB: wie WAR das für dich dass du (-) mehrere themen hattest in diesem test?</p>

	nicht bzw. werden begrüsst ¹⁴⁶	würden, hat dies der/die S verneint.		S: es war (-) gute abwechslungs ´ und (.) weil es wenn es jetzt IMMER die gleichen themen gewesen wären hätte man vielleicht die einzelnen durcheinander gebracht (Omega_Off, Pos. 1505-1508)
SR-LD_1	Lautes Denken hat (eher) gestört	S sagt beim <i>Stimulated-Recall</i> -Interview, dass das LD ihn/sie eher gestört hat und/oder sein/ihr Testverhalten negativ beeinflusst hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Intro, <i>Stimulated Recall</i> : MB: wie war das für dich dass du (.) lAut denken musstest` S: ein wenig (.) UNgewohnt- ((S lächelt)) [...] MB: hat es dich (-) stArk gestört` S: manchmal ein wenig ja; (Alison_Szenario, Pos. 197-202)
SR-LD_2	Lautes Denken hatte einen positiven Einfluss auf das Testverhalten	S sagt beim <i>Stimulated-Recall</i> -Interview, dass das LD sein/ihr Testverhalten positiv beeinflusst hat, z.B. weil er/sie sich besser konzentriert hat.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Intro, <i>Stimulated Recall</i> : MB: mhm- °hh ä:h:m (.) hat es dich geSTÖRT dass du beim Lösen dieser aufgaben laut denken musstest? S: *äh:m es GEHT eigentlich nicht so [...] ja es war noch gut also (-) so habe ich es wohl besser aufgenommen für mich-* MB: okay du hast- [...] das gefühl dass du das BESSER jetzt (-) gelöst hast als wenn du NICHT laut denken würdest; S: °hh *j:a so in einer art (.) schon ein bisschen; also ich habe es einfach besser aufgenommen also (-) weil (ich mal gesehen habe) ich habe es gelesen (-) dann habe ich es nochmal wie

¹⁴⁶ Da die verschiedenen Themen bei Off-Szenario keine Schülerin und keinen Schüler gestört haben, wurde der Code „Verschiedene Themen stören bei Off-Szenario“ aus dem Kategorienleitfaden nachträglich gelöscht.

				wiederholt; * (Boebu_Szenario, Pos. 480-494)
SR-LD_3	Lautes Denken war (eher) egal	S sagt beim <i>Stimulated-Recall</i> -Interview, dass das LD ihn/sie weder positiv beeinflusst noch gestört hat bzw. eher keinen Einfluss auf sein/ihr Testverhalten hatte.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Szenario, Intro, <i>Stimulated Recall</i> : MB: und das laute DENKEN hat es dich geSTÖRT dass du (-) LAUT denken musstest, S: n::ein nicht wirklich (-) ich mache es ja manchmal auch selber (-) vor allem auch wenn ich zum beispiel mathematik (-) (Howard_Szenario, Pos. 189-192)
SR-Sonstiges	Sonstiges	Sonstige interessante Stellen in den <i>Stimulated-Recall</i> -Interviews, z.B. Äusserungen zur Testsituation, zum Fremdsprachenunterricht etc.	Wird aufgrund der entsprechenden Bemerkung(en) der/des S codiert.	Off-Szenario, <i>Stimulated Recall</i> , Szenario „Walkies“, Seite 1: MB: wie IST es für dich? wenn du so was (-) im test (-) was du dann eher witzig oder cool findest` bist du dann motivierter? [das zu machen] S: [ja als/ ja] eh nein` es macht (-) also es ist zuerst eine motivation` und dann (irgendwie) <<:-)> kommt dir in den sinn dass es ein test ist`> ((lacht)) dann denkst du JA ist doch witzig aber jetzt muss ich mich wieder konzentrieren. (Omega_Off, Pos. 471-478)