

Subjektive und objektive Messgrößen zur Erforschung des fremdsprachlichen Leseverstehens: Selbsteinschätzung und Testergebnisse im Vergleich

Elisabeth Peyer¹ & Irmtraud Kaiser²

The present study on foreign language reading examines two different types of self-assessment data: on the one hand it relates global self-assessed reading competence of francophone and italophone readers of German to the results of a reading test as well as a placement test. On the other hand, participants' indications about how confident they felt answering each comprehension question of the reading test are analysed. The study revealed that the majority of intermediate and advanced learners of German assessed their German reading competence one level below their overall German competence as measured by the placement test. Correlations between self-assessment and test results were moderate. Furthermore, our data reveal that learners' confidence when answering reading comprehension questions increases on the higher levels of German competence and is correlated with the correctness of their answers. However, with regard to some very intricate reading items participants often fail to realise that a comprehension problem exists.

1. Selbsteinschätzungen in der Fremdsprachenerwerbsforschung

1.1 Selbsteinschätzung im Fremdspracherwerb und -unterricht

Einschätzungen der eigenen Leistungen durch die Lernenden selbst werden seit jüngerer Zeit sowohl in der Fremdsprachenforschung als auch in der Fremdsprachendidaktik eingesetzt. Aus der Perspektive der Fremdsprachendidaktik gilt die Selbsteinschätzung heute als probates Mittel, die Lernenden zur Selbstreflexion anzuregen (vgl. Harris 1997) und sie somit in die Steuerung des Lernprozesses mit einzubeziehen. Auch im Rahmen von Einstufungsverfahren wird manchmal auf Selbsteinschätzungen zurückgegriffen, etwa um einen Ausgangspunkt für adaptive Computertests festzulegen (vgl. Chalhoub-Deville & Deville 1999).

1 Korrespondenzadresse: Dr. Elisabeth Peyer, Studienbereich Mehrsprachigkeitsforschung und Fremdsprachendidaktik, Universität Freiburg, Criblet 13, CH-1700 Freiburg, elisabeth.peyer@unifr.ch

2 Korrespondenzadresse: Dr. Irmtraud Kaiser, Studienbereich Mehrsprachigkeitsforschung und Fremdsprachendidaktik, Universität Freiburg, Criblet 13, CH-1700 Freiburg, irmtraud.kaiser@unifr.ch

Selbsteinschätzungen werden in vielen Bildungsbereichen eingesetzt und ihre Validität im Vergleich zu 'objektiven' Testinstrumenten wurde vielfach untersucht. In einer Meta-Analyse von 57 Studien aus verschiedenen Bildungsgängen (Medizin, Psychologie etc.) zeigte sich eine breite Übereinstimmung zwischen den subjektiven und den alternativen Bewertungsmethoden, allerdings mit hoher Varianz (vgl. Falchikov & Boud 1989). Gleichzeitig zeigte sich allerdings auch die Abhängigkeit der Qualität der Selbsteinschätzungen von einer hohen Anzahl psychologischer und situativer Faktoren (vgl. auch Abschnitt 1.2).

Auch im Bereich der Fremdsprachenforschung stellen wir eine hohe Variabilität der Ergebnisse fest. In einer Meta-Analyse von 60 Studien zur Selbsteinschätzung der allgemeinen Sprachkompetenz (über alle vier Fertigkeiten hinweg) berechnete Ross (1998) eine durchschnittliche Korrelation von 0,63 zwischen Selbst- und Fremdeinschätzung. Dieser moderaten Korrelation (mit hoher statistischer Signifikanz) steht jedoch eine sehr hohe Streuung der Resultate gegenüber. Die Gründe dafür liegen vermutlich in unterschiedlichen Studienparametern (Material, Methode, Versuchspersonen) und unterschiedlichen situativen Bedingungen, in denen die Selbsteinschätzung und die Testdurchführung stattfanden.

Unterschiedlich sind jedoch häufig nicht nur die (statistischen) Ergebnisse an sich, sondern vielmehr ihre Bewertung durch die Studienautoren. Diese hängt davon ab, wo man die Grenze für eine akzeptable Korrelation zwischen Selbst- und Fremdeinschätzung ansetzt. In der Konsequenz gelangen die verschiedenen Autoren zu unterschiedlichen Schlüssen hinsichtlich der Reliabilität und Validität von Selbsteinschätzungen sprachlicher Kompetenzen.

LeBlanc & Painchaud (1985: 673) etwa kommen nach der Durchführung einer Reihe von Experimenten mit den Fremdsprachen Englisch und Französisch zum Schluss: "self-assessment must be considered a very valuable tool as a placement instrument". Die Mehrheit der in diesem Bereich Forschenden scheint ähnlicher Meinung zu sein und kommt zu positiven Bewertungen hinsichtlich der Aussagekraft von Selbsteinschätzungen (vgl. Alderson 2005; Bachman & Palmer 1989; Blanche 1988; Brantmeier 2005; Oscarson 1997; Raasch 1979).

Andere Autoren jedoch mahnen im Zusammenhang mit der Verwendung von Selbsteinschätzungen im Sprachunterricht zur Vorsicht. Ihre Ergebnisse deuten auf nur schwache oder gar keine Zusammenhänge mit objektiveren Kompetenzmaßen hin (vgl. Blue 1994; Janssen-van Dieten 1989; Peirce, Swain & Hart 1993; Vogel & Vogel 1980; Wesche, Paribakht & Ready 1993). In Untersuchungen zum Zusammenhang zwischen Selbsteinschätzungen und internationalen standardisierten Tests für Englisch als Fremdsprache (IELTS und TOEFL) berechnet Blue (1994) keine signifikanten Korrelationen. Ebenso beobachtet er nur wenig Übereinstimmung zwischen der Eigenbewertung der Lernenden und den Einschätzungen durch die Lehrenden. Blue stimmt dies nachdenklich: "As we believe

accurate self-assessment to be an important factor in the continued language learning (including needs analysis and setting of goals) of such learners, this finding [...] gives cause for some concern" (Blue 1994: 27).

1.2 Potenzielle Einflussfaktoren auf die Selbsteinschätzung

In zahlreichen Studien konnte der Einfluss verschiedener Faktoren beobachtet werden, die mit der Selbsteinschätzung der Sprachkompetenz interagieren. Moritz (1996: 42) identifiziert sechs Kategorien von Einflussfaktoren: Frageinterpretation, sprachbiographischer Hintergrund und Sprachlernerfahrung, herangezogene Vergleichsmaßstäbe, Strategien für die Beantwortung eines Fragebogens, Antwortssicherheit und Selbstbewusstsein.

Die Frageinterpretation ist dabei sicherlich auch testseitig von der Frageformulierung abhängig. Besonders konkrete Fragen haben sich in diesem Zusammenhang bewährt. In mehreren Studien konnte gezeigt werden, dass Selbsteinschätzungen besser mit Testresultaten korrelieren, wenn die Lernenden mit konkreten Beschreibungen von möglichst lebensnahen sprachlichen Aufgabensituationen konfrontiert werden, deren subjektive Bewältigbarkeit sie einschätzen sollen (Brantmeier & Vanderplank 2008; LeBlanc & Painchaud 1985; Peirce, Swain & Hart 1993; Ross 1998). Globale Angaben (etwa zu den vier Fertigkeiten) scheinen hingegen nur wenig Aussagekraft zu besitzen.

Unter bestimmten Bedingungen können Versuchspersonen zudem auf unterschiedliche Vergleichsstandards zurückgreifen, was einen deutlichen Einfluss auf die Selbsteinschätzung hat (vgl. Moritz 1996; Peirce, Swain & Hart 1993). Werden sie etwa gebeten, ihre Sprachkompetenz auf einer Skala von 'sehr gut' über 'durchschnittlich' bis 'sehr schlecht' einzuschätzen, können sie als Vergleichsmaßstab die soziale Gruppe, zu der sie sich zugehörig fühlen, heranziehen, oder aber auf eine bestimmte, subjektiv als besonders vorbildlich oder markant erlebte Einzelperson (z.B. einen native speaker), die eigene Lernentwicklung oder den aktuellen sozialen Kontext (z.B. die Kursgruppe) zurückgreifen (Moritz 1996: 42).

In Bezug auf den sprachbiographischen Hintergrund würde man zunächst erwarten, dass sich das Ausmaß der Sprachlernerfahrung auch in der Korrektheit der Selbsteinschätzungen niederschlägt. Diese Vermutung konnte jedoch in einem Experiment von Blanche (1985, zit. in 1988: 84) und in einer Studie von Janssen-van Dielen (2000) nicht bestätigt werden. Weder das Ausmaß der gesteuerten Lernerfahrung noch des Kontakts mit Fremdsprachen hing mit der Korrektheit der Selbsteinschätzungen zusammen. Widersprüchliche Ergebnisse existieren im Hinblick auf den Einfluss der Kompetenz in der betreffenden Fremdsprache selbst auf die Korrektheit der Selbsteinschätzung (d.h. auf die Korrelation zwischen

Selbsteinschätzung und Testresultat). Janssen-van Dielen (1989: 41) stellt hier keinen Zusammenhang fest, während Poggia et al. (2009: 8) beobachten, dass sich Lernende auf höheren Niveaus korrekter einstufen können als Anfänger und Mittelstufen-Lernende. Sie führen dies auf mehr Erfahrung mit Tests auf Seiten der fortgeschrittenen Lernenden zurück. Einige Studien deuten hingegen auf eine Tendenz der Leistungsstärkeren hin, sich zu unterschätzen, während sich die Leistungsschwächeren eher überschätzen (Blanche 1988; Blue 1994; Brantmeier & Vanderplank 2008).

Eine wichtige Rolle im Zusammenhang mit Selbsteinschätzungen spielt ganz offensichtlich die Persönlichkeitsstruktur der Versuchspersonen (vgl. Blanche 1988; Moritz 1996; Vogel & Vogel 1980). Vogel & Vogel (1980: 84) haben die Ergebnisse eines Persönlichkeitstests mit den Selbsteinschätzungen ihrer Probanden verglichen und stellen fest: "[S]elbstkritische, offene Lerner haben sehr häufig die Tendenz, sich selbst zu unterschätzen, stark extrovertierte, gesellige Lerner hingegen neigen eher zur Selbstüberschätzung".

In Studien von MacIntyre, Noels & Clément (1997) und Young (2000) hat sich zudem gezeigt, dass der Faktor "Foreign Language Anxiety" einen nicht zu unterschätzenden Einfluss auf die Selbsteinschätzung von Fremdsprachen-Lernenden haben kann, bzw. dass ängstliche Versuchspersonen im Vergleich zu anderen ihre Fremdsprachen-Fertigkeiten zu schlecht einstufen.

In einigen wenigen Studien wurden weitere potenzielle Einflussvariablen wie Geschlecht, Alter oder Kulturraum betrachtet; sie lassen jedoch noch keine eindeutigen Schlüsse zu.

1.3 Selbsteinschätzungen zur Untersuchung des fremdsprachlichen Leseverstehens

Ähnlich wie für die selbst eingeschätzte allgemeine Sprachkompetenz (vgl. 1.1) lässt sich in der einschlägigen Literatur auch in Bezug auf die Validität der Selbsteinschätzung des fremdsprachlichen Leseverstehens eine hohe Variabilität feststellen. So weisen Resultate von Studien von Brantmeier (2006) oder Peirce, Swain & Hart (1993) keinen bzw. nur einen schwachen Zusammenhang zwischen Selbsteinschätzungen des fremdsprachlichen Leseverstehens und Testresultaten auf. Demgegenüber kommen z.B. Brantmeier & Vanderplank (2008) oder Le-Blanc & Painchaud (1985) in ihren Studien zum Schluss, dass Selbsteinschätzungen des fremdsprachlichen Leseverstehens verlässlich seien. Verschiedene lerner- und testseitige Faktoren dürften für diese unterschiedlichen Resultate verantwortlich sein. So wurde denn sowohl der Leseverstehenstest als auch der Selbsteinschätzungsfragebogen in den zitierten Studien jeweils unterschiedlich realisiert.

Fragwürdig ist die Realisierung des Leseverstehenstests insbesondere in Peirce, Swain & Harts (1993) Studie, in welcher die Selbsteinschätzungen mit den Resultaten eines Cloze-Tests verglichen wurden. Für das Messen des Leseverstehens wird die Validität dieses (auch) produktiven Testverfahrens seit längerem angezweifelt (vgl. z.B. Alderson 2000: 208f.; Urquhart & Weir 1998: 157). In Bezug auf die Selbsteinschätzungen scheinen kriteriums-basierte, kontextualisierte Items bessere Einschätzungen zu liefern als sehr allgemein gehaltene Selbsteinschätzungsfragen. So wurden sowohl in LeBlanc & Painchauds (1985) als auch in Brantmeier & Vanderblanks (2008) Studie jeweils detaillierte kontextualisierte Selbsteinschätzungsitems verwendet. Während LeBlanc & Painchaud (1985) konkrete Beschreibungen möglicher alltäglicher Lesesituationen zur Selbsteinschätzung benutzten, basierten Brantmeier & Vanderplank (2008) ihren Selbsteinschätzungsfragebogen auf den Kann-Beschreibungen des Gemeinsamen europäischen Referenzrahmens. Brantmeier (2006: 31) interpretiert denn auch den nicht gefundenen Zusammenhang zwischen Selbsteinschätzung und Testresultat in ihrer Studie dahingehend, dass die von ihr gestellten Selbsteinschätzungsfragen zu allgemein gewesen seien.

Nebst der Frage nach der Validität von Selbsteinschätzungen des fremdsprachlichen Leseverstehens wird von der Forschung ebenfalls die Frage diskutiert, ob Lernende ihre produktiven und rezeptiven Fertigkeiten gleich gut einschätzen können. Auch dazu liegen in der einschlägigen Literatur unterschiedliche Resultate vor. So konnten Versuchspersonen in Studien von Raasch (1979) und Alderson (2005) ihre produktiven Sprachfertigkeiten besser einschätzen als ihre rezeptiven. Allerdings verglich Alderson in seiner Studie nicht Korrelationen zwischen Testresultaten und Selbsteinschätzungen der produktiven und rezeptiven Fertigkeiten, sondern wie gut eine allgemeine Selbsteinschätzung der Sprachkompetenz mit der Selbsteinschätzung der einzelnen Teilkompetenzen übereinstimmte (Alderson 2005: 105). Ob man aufgrund dieses Vergleichs aber tatsächlich den Schluss ziehen darf, dass Lernende ihre produktiven Fertigkeiten besser einschätzen können als ihre rezeptiven, erscheint uns zweifelhaft. Im Gegensatz zu Raasch (1979) und Alderson (2005) kommt Ross (1998) in seiner Metastudie zu Korrelationen zwischen Sprachtests und Selbsteinschätzungen der vier Sprachfertigkeiten zum Schluss, dass sich Versuchspersonen in den rezeptiven Kompetenzen leicht besser einschätzen können als in den produktiven. So lag der Durchschnitt der 23 ausgewerteten Leseverstehens-Korrelationen bei $r = 0,61$ und derjenige der 18 Hörverstehens-Korrelationen bei $r = 0,65$ (Ross 1998: 5 bzw. 7). Demgegenüber fielen die Korrelationen für das Sprechen ($r = 0,55$) und Schreiben ($r = 0,52$) etwas niedriger aus (Ross 1998: 8 bzw. 9). Die Homogenität der Korrelationen war zudem für das Leseverstehen am höchsten. Die höhere Validität der Selbsteinschätzungen des Leseverstehens begründet Ross damit, dass im schulischen

Erwerbskontext das Lesen normalerweise diejenige Fertigkeit sei, die am meisten geübt werde. Die größere Erfahrung mit dem fremdsprachlichen Lesen mache sich dann auch bei der Selbsteinschätzung bemerkbar (Ross 1998: 6).

Auch für die selbst eingeschätzte Lesekompetenz soll die Frage nicht ausgeklammert werden, ob sich die Lernenden je nach erreichtem Sprachniveau eher über- oder unterschätzen. Hierzu liegen jedoch erst wenige Resultate vor, die sich zum Teil erheblich widersprechen, vermutlich aufgrund unterschiedlich realisierter Parameter wie Methodik und Versuchspersonen. So beobachten Van Onna und Jansen in ihrer Studie, dass eine deutliche Mehrheit ihrer niederländischen Versuchspersonen ihre Lesekompetenz in Französisch eine oder sogar zwei GeR³-Stufen über dem mit DIALANG⁴ getesteten Niveau einschätzen (Van Onna & Jansen 2006: 177f.). Demgegenüber stellt Alderson für seine ebenfalls mit DIALANG getesteten Versuchspersonen unterschiedlicher L1 fest, dass sich die Mehrheit der fortgeschrittenen Sprachlernenden bezüglich der Lesekompetenz um eine GeR-Stufe unterschätzt, während die Mehrheit der Versuchspersonen auf Anfänger- und Mittelstufenniveau sich richtig einstuft (Alderson 2005, 108). Auch in Brantmeier & Vanderplanks Studie haben fortgeschrittene Versuchspersonen die Tendenz, ihre mit Hilfe eines Tests gemessene Lesekompetenz zu unterschätzen (Brantmeier & Vanderplank 2008: 471), während fortgeschrittene Studierende in einer Studie von Brantmeier (2005) ihre mit einem "written recall task" überprüfte Lesekompetenz mehrheitlich richtig einschätzten.

Während es also einige Studien gibt, die Selbsteinschätzungen des fremdsprachlichen Leseverstehens mit Testresultaten vergleichen, sind uns Studien zur Einschätzung des subjektiven Verständnisses einzelner Sätze und Strukturen nicht bekannt. Der Blick auf das selbst eingeschätzte Verständnis einzelner Sätze erscheint uns jedoch ebenfalls lohnenswert, kann doch das Bewusstsein bzw. das Erkennen einer Schwierigkeit unter Umständen ein erster Schritt im Erwerbsprozess einer Struktur sein.

2. Methodik

Bei der vorliegenden Arbeit handelt es sich um eine Teilstudie aus einem Forschungsprojekt⁵ zum Lesen in Deutsch als Fremdsprache und zum Verständnis verschiedener grammatischer Strukturen. Ein eigens konstruierter Lesetest sollte

3 GeR = Gemeinsamer europäischer Referenzrahmen für Sprachen (Europarat 2001).

4 Bei DIALANG handelt es sich um ein Diagnoseinstrument zur Feststellung des GeR-Sprachniveaus für bestimmte Teilfertigkeiten. Vgl. www.dialang.org

5 Das Projekt wurde finanziert vom Schweizerischen Nationalfonds; Projekt-Nr. PP001-106634, Projektleiter: Raphael Berthele.

einerseits zu Aussagen über den relativen Schwierigkeitsgrad spezifischer grammatischer Strukturen des Deutschen führen und andererseits zu Einsichten in die Selbsteinschätzung des Leseverständnisses einzelner Sätze/Strukturen durch die Versuchspersonen. Nebst dem eigentlichen Lesetest wurden als weitere Erhebungsinstrumente ein Selbsteinschätzungsfragebogen zu rezeptiven Kompetenzen in verschiedenen Sprachen verwendet sowie ein Deutsch-Einstufungstest, der eine grobe Zuordnung der Versuchspersonen zu den Niveaus des europäischen Referenzrahmens erlaubte. Die Ergebnisse dieser verschiedenen Arten von Selbsteinschätzungen sollen – besonders hinsichtlich ihres Zusammenhangs mit den Testresultaten – im Folgenden diskutiert werden.

2.1 Fragestellung

Die zentrale Fragestellung der vorliegenden Studie bezieht sich auf das Verhältnis zwischen objektiven Testresultaten und der Einschätzung von Kompetenzen und Leistungen durch die Lernenden selbst. Im Zentrum des Interesses stand damit auch die Frage, inwieweit sich Lernende ihrer Schwierigkeiten beim Leseverstehen in Deutsch als Fremdsprache bewusst sind – und zwar im Allgemeinen und im Hinblick auf spezifische Lesetest-Items.

Die übergreifende Fragestellung wurde demnach auf einer globaleren und einer spezifischeren Ebene betrachtet. Auf globaler Ebene erfolgte die Operationalisierung über den Vergleich zweier Testergebnisse (Einstufungstest; Leseverstehenstest) mit der Selbsteinschätzung der allgemeinen Lesekompetenz der Lernenden in Deutsch. Auf der spezifischeren Ebene vergleichen wir für die einzelnen Sätze unseres Leseverstehenstests die Korrektheit des Leseverständnisses mit der Antwortsicherheit der Versuchsperson. Die zu vergleichenden Messgrößen sind demnach:

- allgemein, objektiv: Deutsch-Einstufungstest
- lesespezifisch (allgemein), objektiv: Lesetest zum Verstehen grammatischer Strukturen
- lesespezifisch (allgemein), subjektiv: Selbsteinschätzung der Lesekompetenz in Deutsch als Fremdsprache
- lesespezifisch (Einzelsatz), objektiv: Korrektheit des Leseverständnisses bei jedem Einzelsatz
- lesespezifisch (Einzelsatz), subjektiv: Antwortsicherheit bei der Verstehensfrage zu jedem Einzelsatz.

Während der Vergleich von Messgrößen auf globaler Ebene bereits eine gewisse Tradition in der Selbsteinschätzungsforschung hat (vgl. Kap. 1), ist die Betrachtung von Selbsteinschätzungen des Leseverständnisses auf Ebene des ein-

zelen Lese-Items unseres Wissens neu. Im Gegensatz zur globalen Selbsteinschätzung der Lesekompetenz in Deutsch als Fremdsprache sollten die Daten zur Antwortsicherheit auf Einzelsatzebene Hinweise auf die wahrgenommene Schwierigkeit bestimmter Sätze bzw. ihrer grammatischen Struktur geben.

Darüber hinaus sollen die Zusammenhänge zwischen Antwortsicherheit und anderen Variablen wie Gesamt-Schwierigkeitsgrad des Items oder allgemeine Deutsch-Kompetenz erkundet werden. Dadurch soll das Konstrukt 'Antwortsicherheit' konkreter fassbar gemacht werden.

Unsere spezifischen Fragen, die in den folgenden Abschnitten diskutiert werden, lauten:

- (1) Wie hängen die selbst eingeschätzte Deutsch-Lesekompetenz und die getestete allgemeine Deutsch-Kompetenz zusammen?
- (2) Wie stark korreliert die Anzahl der Fehler beim Lesetest mit der selbst eingeschätzten Lesekompetenz?
- (3) Wie hängt die subjektive Antwortsicherheit mit dem Deutsch-Kompetenzniveau der Versuchsperson zusammen?
- (4) Wie hängen Antwortsicherheit und Korrektheit der individuellen Antwort zusammen?
- (5) Wie hängt die subjektive Antwortsicherheit mit der Schwierigkeit des jeweiligen Satzes (gemessen an der Gesamtfehlerquote) zusammen?
- (6) Welche Gründe für die subjektive (Un-)Sicherheit beim Beantworten spezifischer Lesetest-Items lassen sich identifizieren?

Schließlich soll die Diskussion aber auch Schlussfolgerungen darüber erlauben, welchen Beitrag Selbsteinschätzungen zur Erforschung des fremdsprachlichen Leseverstehens leisten können.

2.2 Selbsteinschätzungsfragebogen

Den ersten Teil der Untersuchung stellte ein Fragebogen dar, mit dessen Hilfe zum einen die L1 der Testpersonen und zum anderen rezeptiv beherrschte Fremdsprachen erfasst wurden. Die Versuchspersonen wurden dabei gebeten, ihre Lesekompetenz in Deutsch, Englisch und weiteren Fremdsprachen mittels des "Rasters zur Selbstbeurteilung" des Europäischen Sprachenportfolios genauer zu bestimmen. Von diesem Raster wurden lediglich die Niveau-Beschreibungen für das Lesen verwendet. Dieses sechsstufige Raster bietet konkrete Kann-Beschreibungen, anhand deren die Lernenden sich einer der sechs Stufen A1 ("Breakthrough") bis C2 ("Mastery") zuordnen können. Der Fragebogen war – je nach Erstsprache der Testpersonen – auf Französisch bzw. Italienisch verfasst.

2.3 Leseverstehenstest

Das zentrale Testinstrument waren zwei eigens entwickelte Leseverstehens-tests, die jeweils aus einem Lesetext und einem methodisch vielfältigen Verstehenstest bestanden. Bei den Texten handelte es sich um die Beschreibungen von zwei imaginären Tieren ('Flundodil' und 'Humpfhorn') in Form von Lexikonartikeln. Diese Textsorte bot eine Reihe von Vorteilen: So konnten etwa die Faktoren Vor-/Weltwissen und Textsortenschemata auf diese Weise kontrolliert werden. Die Textsorte 'Lexikonartikel' dürfte der gewählten Probandenzielgruppe (Studierende) hinlänglich bekannt sein. Auch ein grundsätzliches Schemawissen zum Konzept 'Tier' dürfte angenommen werden. Da es sich jedoch um die Beschreibung eines imaginären Tieres handelte, konnte man gleichzeitig ausschließen, dass unterschiedliches Vor-/Weltwissen der Versuchspersonen eine relevante Einflussgröße darstellte. Durch die Arbeit mit einem zusammenhängenden Text war außerdem der Kontext und dessen Relevanz für die Verarbeitung der aktuellen sprachlichen Struktur vorgegeben. In Summe war diese Situation wohl einer authentischen Lesesituation recht nahe: Will man einen Fachtext lesen, so bringt man in der Regel ein gewisses Vorwissen mit (sonst würde man sich für den Text nicht interessieren), will aber dennoch dem Text neue Informationen entnehmen (denn auch sonst würde man sich für den Text nicht interessieren). Da die vorliegende Untersuchung ausschließlich auf Verstehensschwierigkeiten auf der grammatischen Ebene abzielte, wurden die Inhaltswörter in die L1 der jeweiligen Versuchspersonen übersetzt und somit auch gleichzeitig der Faktor Lexikon (zumindest teilweise) kontrolliert.

Die Liste der grammatischen Strukturen, deren Schwierigkeit für das Lesen die Untersuchung überprüfen sollte, ist zum einen aus Pilotstudien hervorgegangen, die in DaF-Kursen der Universitäten Fribourg und Bern durchgeführt wurden. Zum anderen basiert sie auf der Analyse von sog. 'rezeptiven Grammatiken' des Deutschen (vgl. z.B. Heringer 1987, 2001; Bernstein 1990; Stalb 1993) sowie empirischen Studien zu grammatischen Schwierigkeiten beim DaF-Leseverstehen (u.a. Lutjeharms 1998, Fabricius-Hansen 2002). Überprüft wurden das erweiterte Linksattribut, die OVS-Topologie, das Konditionalgefüge ohne Subjunkt, die Verbalklammer, die Verb-Topikalisierung, der Subjektsatz ohne Korrelat sowie das 'werden'-Passiv. Um die Schwierigkeit der Strukturen zu überprüfen, wurde mit Parallelversionen der Texte gearbeitet. Jedem Satz mit der zu überprüfenden Zielstruktur entsprach ein inhaltlich gleicher Satz, in dem aber eine alternative grammatische Struktur gebraucht wurde. Jeder Zielstruktur wurde also eine alternative Struktur zugeordnet:

Zielstruktur – Alternativstruktur:

- Linksattribut – Relativsatz
- OVS-Topologie – SVO-Topologie
- Konditionalgefüge ohne Subjunktor – Konditionalgefüge mit Subjunktor
- Verbalklammer – keine Verbalklammer
- Verb-Topikalisierung – keine Verb-Topikalisierung
- Subjektsatz ohne Korrelat – (zumeist) Subjektsatz mit Korrelat
- 'werden'-Passiv – Aktiv

Die Lexikonartikel wurden demnach in jeweils zwei verschiedenen Versionen verfasst und die Zielstrukturen bzw. die alternativen Strukturen gleichmäßig auf die beiden Versionen verteilt. Auf diese Weise sollte die direkte Vergleichbarkeit der Verstehensschwierigkeit von Ziel- und Alternativstruktur gewährleistet werden.

Der gesamte Testteil war in der jeweiligen L1 der Versuchspersonen verfasst und auch die Testpersonen selbst waren angehalten, in ihrer L1 zu antworten, um nicht das Verständnis der Fragestellung oder die Formulierung in der Fremdsprache zum Hindernis bzw. zum Verzerrungsfaktor werden zu lassen. Da die Testfragen jeweils auf das Verständnis eines Satzes/einer grammatischen Konstruktion zielten, konnten die grammatischen Schwierigkeiten relativ isoliert voneinander überprüft werden.

Jeder Testfrage war zudem eine Frage nach der Antwortsicherheit zugeordnet. Diese Frage lautete in der italienischen Testversion:

Quanto siete sicuri?

- ☐ 25 %
- ☐ 50 %
- ☐ 75 %
- ☐ 100 %.

So wird mit der Frage nach der Antwortsicherheit gleichzeitig die Selbsteinschätzung des Verstehenserfolgs erhoben, und zwar immer bezogen auf eine konkrete Verstehensfrage. Mit Hilfe dieser Zusatzfrage sollten einerseits Anhaltspunkte über das Ausmaß des Ratens gewonnen werden, aber auch über das Bewusstsein der Probandinnen und Probanden für die Schwierigkeiten im Satz. Die Methode ist unseres Wissens in dieser Form neu und somit können wir unsere Ergebnisse auch zu keinen bereits vorliegenden Untersuchungen in Beziehung setzen. Sie hat sich jedoch unseres Erachtens als durchaus empfehlenswert erwiesen und bedeutende Zusatzerkenntnisse geliefert.

2.4 Deutsch-Einstufungstest

Im Anschluss an den Leseverstehenstest wurden die Versuchspersonen gebeten, den Online-Einstufungstest des Goethe-Instituts zu lösen.⁶ Dies ermöglichte eine grobe Zuordnung der Testpersonen zu den Niveaus des europäischen Referenzrahmens. Der Einstufungstest des Goethe-Instituts besteht aus insgesamt dreißig Multiple-Choice-Fragen. Es handelt sich dabei um Sätze mit Lücken, zu denen drei oder vier Lösungsvorschläge zur Auswahl stehen. Die Hälfte dieser Sätze ist dabei kontextfrei und hat je vier Multiple-Choice-Optionen. Die anderen fünfzehn Sätze entstammen drei kurzen Texten. Für die Lücken in diesen Sätzen stehen jeweils nur drei Multiple-Choice-Optionen zur Auswahl. Die meisten Multiple-Choice-Items fragen grammatisches Wissen ab (18), einige lexikalisches Wissen (5) und circa ein Viertel (7) der Items testet Wissen über Kollokationen.

2.5 Durchführung der Studie und Versuchspersonen

Die Untersuchung wurde an verschiedenen französischen und italienischen Universitäten mit 506 Studierenden französischer oder italienischer L1 durchgeführt (312 italophon; 194 frankophon). Der Großteil der Probandinnen und Probanden waren Studierende der Germanistik oder eines Wirtschafts- bzw. Tourismusfachs mit Nebenfach in Fremdsprachen. Einige jedoch waren in anderen Fächern inskribiert. Manche hatten noch nie Deutsch gelernt. Die drei Untersuchungsteile wurden in folgender Reihenfolge durchgeführt:

1. Selbsteinschätzungsfragebogen,
2. Leseverstehenstest und
3. Deutsch-Einstufungstest.

Während von allen 506 Versuchspersonen Angaben zur Selbsteinschätzung ihrer Deutsch-Lesekompetenz vorliegen, verfügen wir nur von 323 Versuchspersonen auch über Angaben zu ihrer Antwortsicherheit bei jeder Frage.

2.6 Qualitative Studie

Eine qualitative Studie in kleinerem Rahmen sollte die Erkenntnisse aus der quantitativen Erhebung ergänzen. Während die Lesetest-Studie Aussagen über das Endprodukt des Verstehensprozesses (d.h. richtig bzw. falsch verstandene

⁶ <http://www.goethe.de/cgi-bin/einstufungstest/einstufungstest.pl>

Teile) und quantitative Aussagen über Selbsteinschätzungen und Antwortsicherheit erlaubt, sollte die qualitative Zusatzuntersuchung eine Basis für Erkenntnisse über den Prozess des Leseverstehens liefern und über die Gründe für Unsicherheiten. Sie sollte es erlauben, die Hindernisse, aber auch die Hilfestellungen auf dem Weg des Verstehens genauer zu identifizieren und so auch hilfreiche Hinweise darauf geben, wie die quantitativen Ergebnisse zu interpretieren sind.

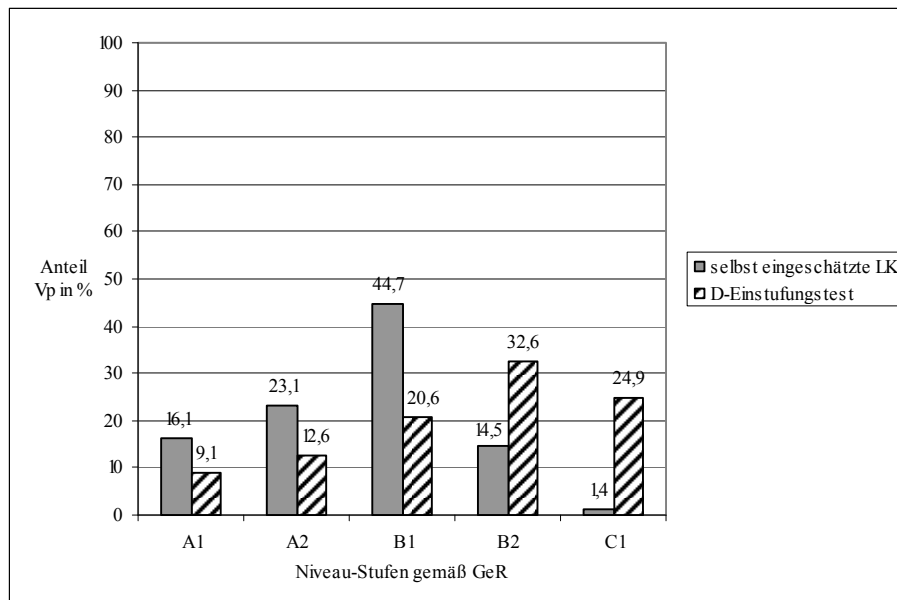
Die gewählte Vorgangsweise integriert Elemente verschiedener Methoden. Die Versuchspersonen wurden jeweils in Zweier- oder Dreiergruppen mit den Texten zu 'Humpfhorn' und 'Flundodil', wiederum mit übersetzten Inhaltswörtern, konfrontiert. Sie wurden dann gebeten, diese in Zusammenarbeit in ihre L1 zu übersetzen und möglichst viele Denkprozesse, Probleme und Gedanken dabei zu artikulieren. Dieses Gespräch in der Gruppe wurde aufgezeichnet. Durchgeführt wurden die qualitativen Nacherhebungen mit 26 Gruppen.

3. Resultate und Diskussion

3.1 Selbst eingeschätzte Lesekompetenz, allgemeines Deutsch-niveau und Verständnis des Lesetests

Eine erste zentrale Frage betrifft den Zusammenhang zwischen selbst eingeschätztem Lesekompetenz-Niveau und dem mittels Einstufungstest gemessenen allgemeinen Deutschniveau (Frage 1). Beide Werte beziehen sich auf die gleiche Messskala – den europäischen Referenzrahmen.

Eine erste Berechnung der Korrelation zwischen der Selbsteinschätzung der Lesekompetenz unserer Probandinnen und Probanden und ihrem Ergebnis im allgemeinen Einstufungstest ergibt, dass diese beiden Resultate statistisch signifikant miteinander korrelieren ($r = 0,676$; $p < 0,001$). Ein genauerer Blick auf das Verhältnis zwischen diesen beiden Messgrößen zeigt jedoch dennoch bedeutende Unterschiede auf.


 Abbildung 1: Selbst eingeschätzte Lesekompetenz und -Einstufungstest⁷

Aus Abbildung 1 wird deutlich, dass die meisten unserer Versuchspersonen ihre Lesekompetenz zwischen A1 und B1 einschätzen.⁸ Nur rund ein Sechstel schätzen ihre Lesekompetenz auf Niveau B2 oder höher ein. Gemäß den Resultaten des Deutsch-Einstufungstests befinden sich aber 57,7 % auf Niveau B2 oder höher. Unsere Versuchspersonen haben also die Tendenz, ihre Lesekompetenz unterhalb ihres allgemeinen, getesteten Deutsch-Niveaus einzustufen. Dies gilt insbesondere für Versuchspersonen auf höheren Deutsch-Niveaus.

⁷ A1: S (Selbsteinschätzung): n (VPs) = 81, T (Test): n = 46; A2: S: n = 116, T: n = 50; B1: S: n = 225, T: n = 118; B2: S: n = 73, T: n = 200; C1: S: n = 7, T: n = 91; C2: S: n = 1, T: n = 1.

⁸ Die Versuchspersonen konnten sich auf die Niveaus A1-C2 einstufen. Ein Einstufen als Nullanfänger war nicht möglich. Lediglich eine Person hat sich auf Niveau C2 eingestuft. Aus diesem Grund wird das Niveau C2 in diesem Artikel jeweils nicht dargestellt. Von drei Versuchspersonen liegen keine Selbsteinschätzungen der Deutsch-Lesekompetenz vor.

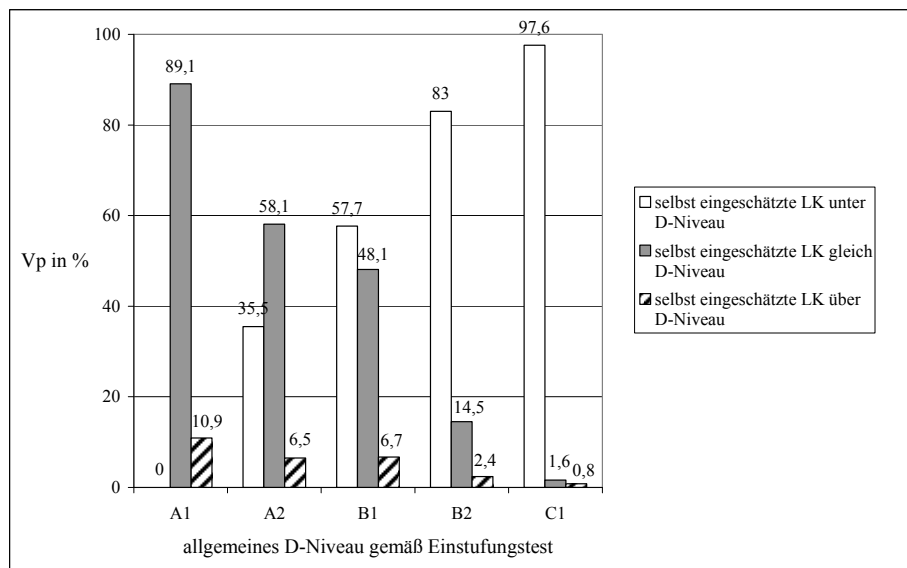


Abbildung 2: Selbst eingeschätzte Lesekompetenz auf den D-Niveaus⁹

Wie aus Abbildung 2 ersichtlich wird, schätzen Versuchspersonen auf Niveau A1 und A2 ihre Lesekompetenz noch mehrheitlich auf dem gleichen Niveau wie ihre allgemeine Deutsch-Kompetenz ein. Der Anteil der sich korrekt Einstufenden nimmt mit steigendem Deutsch-Niveau jedoch kontinuierlich ab und beträgt bereits auf Niveau B2 noch lediglich 14,5 %. Demgegenüber steigt der Prozentsatz jener, die ihre Lesekompetenz im Vergleich zu ihrem allgemeinen Deutsch-Niveau schlechter einstufen, bei steigendem Deutsch-Niveau kontinuierlich an. Bereits auf Niveau B1 schätzt eine knappe Mehrheit der Versuchspersonen ihre Lesekompetenz unterhalb ihrer getesteten Deutsch-Kompetenz ein, auf Niveau B2 sind es bereits rund vier Fünftel der Versuchspersonen. Erstaunlich ist im Weiteren, dass viele Versuchspersonen ihre Lesekompetenz um mehr als nur eine Niveau-Stufe unterhalb des getesteten allgemeinen Deutsch-Niveaus einschätzen. So stufen 52,8 % (n = 66) von den Versuchspersonen mit einem getesteten Deutsch-Niveau von C1 ihre Lesekompetenz auf B1 ein und einige (11,2 %, n = 14) sogar auf A2.

Will man dieses erstaunliche Resultat interpretieren, so gilt es natürlich zu beachten, dass der Einstufungstest nicht spezifisch das Leseverstehen, sondern ver-

⁹ A1: U (Selbsteinschätzung unter D-Niveau): n = 0, G (Selbsteinschätzung gleich D-Niveau): n = 41, Ü (Selbsteinschätzung über D-Niveau) n = 5; A2: U: n = 22, G: n = 36, Ü: n = 4; B1: U: n = 47, G: n = 50, Ü: n = 7; B2: U: n = 137, G: n = 24, Ü: n = 4; C1: U: n = 122, G: n = 2, Ü: n = 1; C2: U: n = 1, G: n = 0, Ü: n = 0.

schiedene Bereiche wie Wortschatz- und Grammatikkenntnisse misst. Dennoch ist es bemerkenswert, dass so viele Versuchspersonen ihre Lese-Kompetenz unterhalb ihrer gemessenen Deutschkompetenz einstufen.¹⁰ Ob die Versuchspersonen ihre Lesekompetenz zumeist unterschätzen, kann aufgrund unserer Daten nicht mit Sicherheit geklärt werden, da wir zusätzlich zur Selbsteinschätzung der Leseverstehenskompetenz keinen standardisierten Leseverstehenstest durchgeführt haben. Möglicherweise dürfte aber der Ruf des Deutschen als schwierig zu erlernende Sprache in die Selbsteinschätzungen eingeflossen sein. Eine andere Erklärung wäre, dass der Einstufungstest nicht das bereits erreichte Deutsch-Niveau angibt, sondern jenes des Kurses, der zum betreffenden Niveau hinführt. Dies würde bedeuten, dass der Einstufungstest eher etwas zu hohe Niveaus angeben würde.¹¹

Der Leseverstehenstest wurde generell relativ gut gelöst. Die übersetzten Inhaltswörter ermöglichten es bereits den Anfängerinnen und Anfängern, einen Großteil der Sätze zu verstehen: Bei insgesamt 21 möglichen Fehlern lag die durchschnittliche Fehlerquote auf Niveau A1 bei 7,7 Fehlern. Es wurden also rund 2/3 aller Sätze bereits auf Niveau A1 richtig verstanden (Fehlerquoten auf den übrigen Deutsch-Niveaus: A2: 5,9; B1: 5,3; B2: 4,0; C1: 2,7 Fehler). Bezüglich des Lesetest-Resultats interessierte uns insbesondere der Zusammenhang zwischen diesem und der selbst eingeschätzten Lesekompetenz (Frage 2). Dabei war von einer negativen Korrelation auszugehen, d.h. dass mit steigender selbst eingeschätzter Lesekompetenz die Anzahl der Fehler beim Lesetest abnehmen würde. Dies bestätigte sich in unseren Daten: Eine Korrelationsberechnung zwischen der selbst eingeschätzten Deutsch-Lesekompetenz und der Fehleranzahl beim Lesetest ergab einen sehr signifikanten Zusammenhang ($p < 0,001$), der bei einem Korrelationskoeffizienten von $r = -0,393$ jedoch relativ gering ist. Dies dürfte unter anderem an der hohen Streuung der Fehleranzahl innerhalb eines selbst eingeschätzten Lesekompetenz-Niveaus liegen, wie sie auch in Abbildung 3 deutlich wird.

10 Da die Selbsteinschätzung zeitlich vor dem Einstufungstest stattfand, waren die Versuchspersonen in ihrer Einschätzung jedoch unbeeinflusst vom Testresultat.

11 Der Website des Einstufungstests des Goethe-Instituts ist nicht eindeutig zu entnehmen, ob hier das bereits erreichte oder das anzuzielende Niveau bekanntgegeben wird, sondern dass ein Kursbesuch auf der jeweiligen Stufe empfohlen wird.

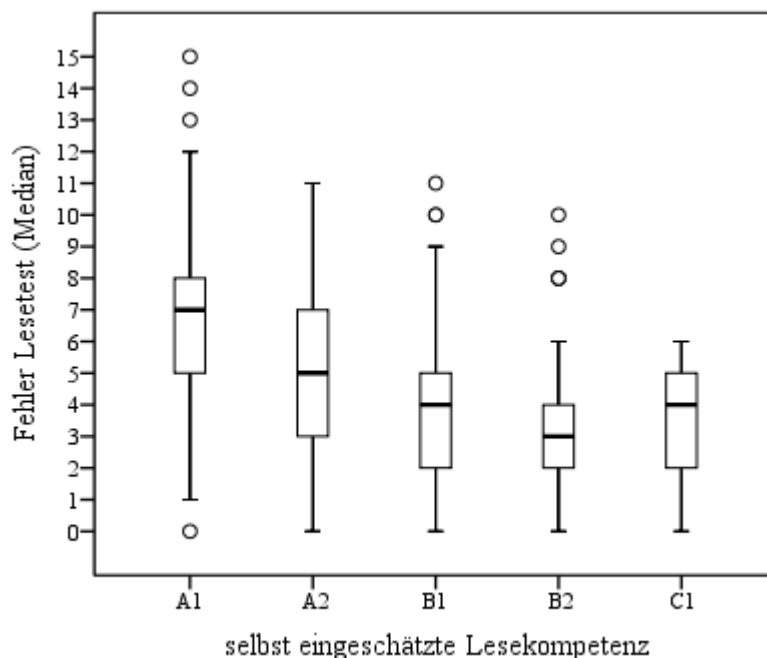


Abbildung 3: Box-Plot¹² Fehler beim Lesetest nach selbst eingeschätzter Lesekompetenz¹³

Da unser Lesetest eigens konzipiert wurde, einen spezifischen Teil des Leseverstehens, nämlich das grammatische Satzverstehen, zu überprüfen, ist diese schwache Korrelation jedoch nur auf den ersten Blick erstaunlich. Sie dürfte eine Folge davon sein, dass der Lesetest tatsächlich nicht das Leseverstehen in seinem vollen Umfang überprüft, sondern gewisse Aspekte wie die Worterkennung und das Erstellen eines mentalen Modells nur am Rande berücksichtigt. Dieser Zusammenhang zwischen selbst eingeschätzter Lesekompetenz und Lesetest-Resultat erweist sich zudem als deutlich schwächer als jener zwischen selbst eingeschätzter Lesekompetenz und allgemeinem Einstufungstest, der ja – wie vorne erwähnt – einen Korrelationskoeffizienten von $r = 0,676$ ($p < 0,001$) aufweist. Dies ist zumindest insofern erstaunlich, als sich beide Messgrößen auf dieselbe Teilkompetenz beziehen.

¹² Box-Plot: Erläuterung:

- Fett ausgezogene horizontale Linie innerhalb der Boxen: Median.
- Boxen: Interquartilbereich mit den mittleren 50 % der Fälle.
- Senkrechte Linien mit Füßchen: Streuung.
- Kreise: Ausreißer (1,5 bis 3 Boxenhöhen weg vom jeweiligen Boxenende).

¹³ A1: n (VP) = 81; A2: n = 116; B1: n = 225; B2: n = 73; C1: n = 7; (C2: n = 1).

3.2 Selbst eingeschätzte Antwortsicherheit und Verständnis des Lesetests

Wie oben bereits erwähnt, haben wir einen Teil unserer Versuchspersonen¹⁴ gebeten, für jede Antwort auf eine Verstehensfrage anzugeben, wie sicher sie sich dabei fühlen. Insgesamt überwog über alle Strukturen hinweg die Sicherheit unserer Probandinnen und Probanden: 69,7 % der Fragen wurden mit 100-prozentiger Sicherheit beantwortet, nur bei 30,3 % der Fragen waren die Leserinnen und Leser unsicher.¹⁵

Diese erstaunlich hohe Sicherheit steht teilweise nicht im Einklang mit der Fehlerhaftigkeit der Antworten (s.u.). Hinsichtlich der Verlässlichkeit unseres Testdesigns bedeutet sie jedoch, dass wir nur mit einem relativ geringen Anteil schlichten Ratens bei der Fragebeantwortung rechnen müssen.

Eine Aufschlüsselung der Sicherheitsangaben nach Deutsch-Niveau gemäß Einstufungstest demonstriert einen klaren Anstieg der subjektiven Antwortsicherheit mit höherem Deutsch-Niveau. Überwiegt auf Niveau A1 noch die Unsicherheit bei den Probandinnen und Probanden, so übernimmt die Sicherheit ab Niveau A2 die Oberhand.¹⁶ Auf Niveau C1 werden schließlich über drei Viertel der Fragen mit absoluter Sicherheit beantwortet.

14 Wir verfügen über Sicherheitsangaben von 323 Versuchspersonen.

15 Die Probandinnen und Probanden konnten bei jeder Frage ankreuzen, wie sicher sie sich bei der Beantwortung fühlen: zu 25%, zu 50%, zu 75% oder zu 100%. Für die Auswertung wurden die Angaben zweistufig umkodiert: Die Angabe 100% gilt als 'sicher', alle anderen Angaben als 'unsicher'.

16 Hier ist bereits der Unterschied zwischen den Niveaus A1 und A2 statistisch signifikant ($\chi^2(1) = 7,361$; $p < 0,01$).

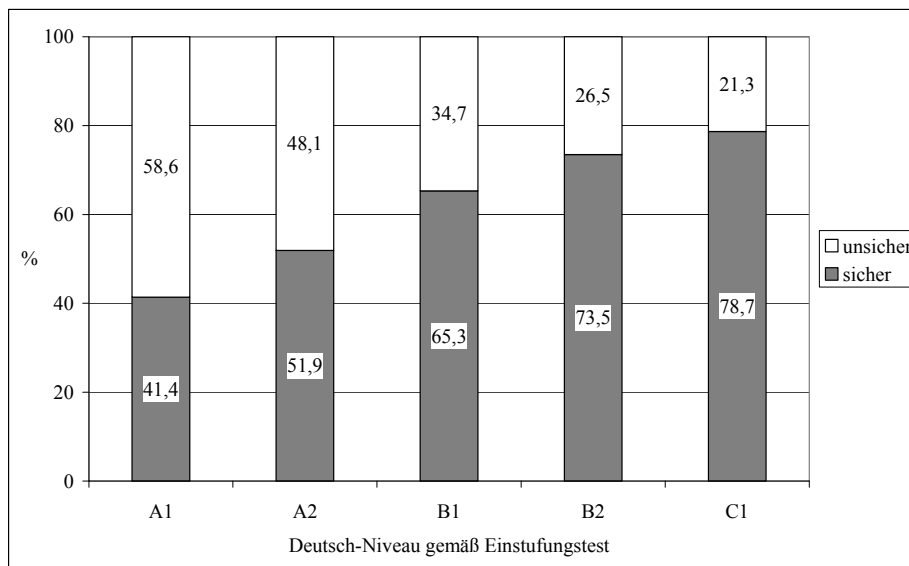


Abbildung 4: (Un-)Sicherheit nach Deutsch-Niveau (Einstufungstest)¹⁷

Ein sogar noch deutlicheres Bild entsteht, wenn man nicht die Daten des Einstufungstests, sondern jene der Selbsteinschätzung der Kandidatinnen und Kandidaten heranzieht. Wiederum fühlen sich nur jene Personen mehrheitlich bei der Beantwortung der Fragen unsicher, die angeben, Deutsch lediglich auf Niveau A1 lesen zu können. Schon auf Niveau A2 kehrt sich das Verhältnis um und die Lesenden fühlen sich zu fast zwei Dritteln sicher.¹⁸

¹⁷ A1: n (Sätze) = 292; A2: n = 389; B1: n = 809; B2: n = 1708; C1: n = 1368.

¹⁸ Der Unterschied ist bereits zwischen den selbst eingeschätzten Niveaus A1 und A2 statistisch signifikant ($\chi^2(1) = 122,837$; $p < 0,001$).

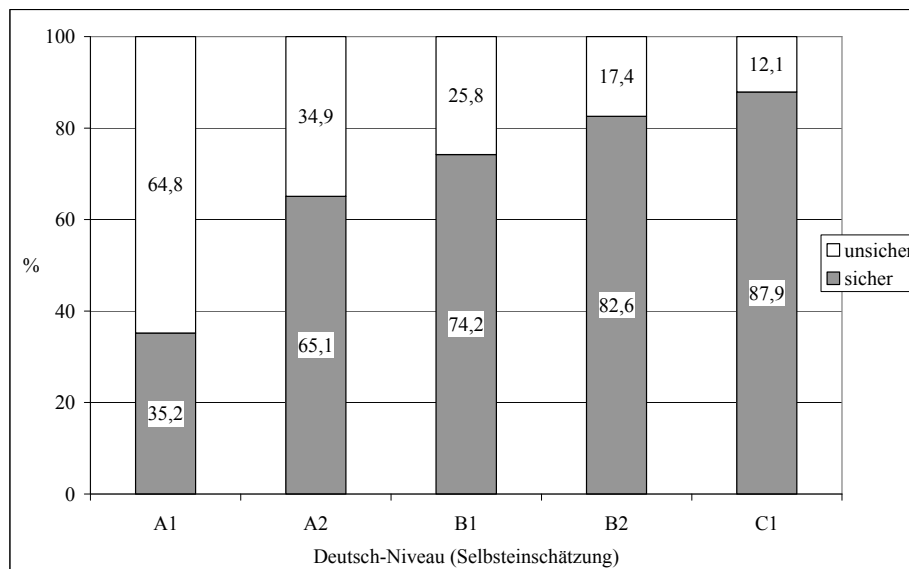


Abbildung 5: (Un-)Sicherheit nach Deutsch-Niveau (Selbsteinschätzung)¹⁹

Hinsichtlich unserer Frage (3) können wir also feststellen, dass sich Personen mit höherem Deutsch-Kompetenzniveau, und zwar sowohl hinsichtlich des Testresultats als auch hinsichtlich des selbst eingeschätzten Deutsch-Niveaus, sicherer bei der Beantwortung unserer Testfragen fühlen.

Vergleicht man die Angaben zur Sicherheit mit der Korrektheit der Antwort auf die Verstehensfrage, so stellt man fest, dass hier – relativ wenig überraschend – ein Zusammenhang existiert.

¹⁹ A1: n = 514; A2: n = 1001; B1: n = 2083; B2: n = 867; C1: n = 58.

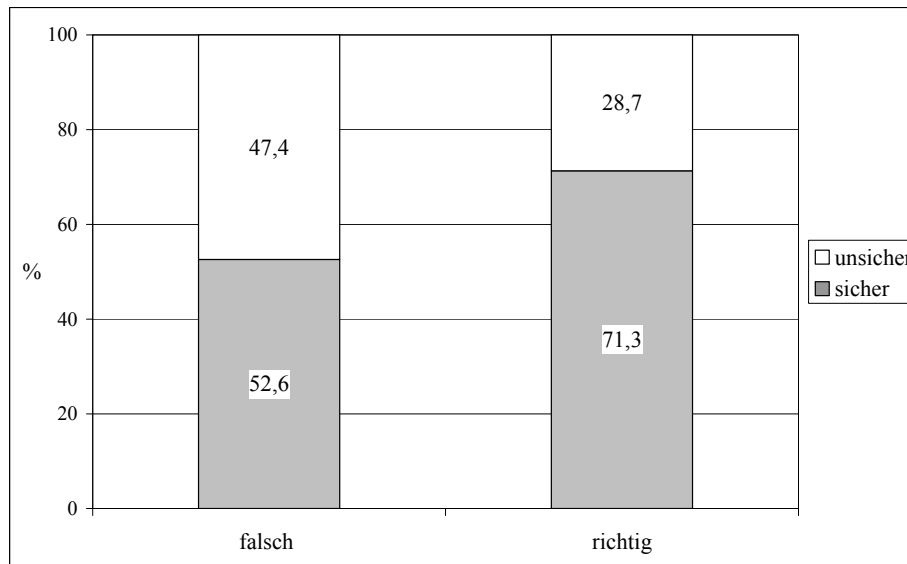


Abbildung 6: (Un-)Sicherheit nach Korrektheit der Antwort (falsch: n = 775; richtig: n = 3791)

Von jenen Personen, die richtig auf die gestellte Verstehensfrage antworteten, fühlten sich 71,3 % bei ihrer Antwort sicher. Von jenen, die letztlich falsch antworteten, waren sich weniger, nämlich 52,6 % sicher. Auch wenn dieser Unterschied deutlich signifikant ist,²⁰ scheint die subjektive Sicherheit unter jenen, deren Antwort letztlich falsch war, doch erstaunlich hoch. Um dies zu verdeutlichen: Mehr als die Hälfte der Versuchspersonen war sich zu 100 % sicher, richtig zu antworten, wenn ihre Antwort tatsächlich falsch war! Das bedeutet schließlich nichts anderes, als dass diese Versuchspersonen sich beim Lesen der gegebenen Sätze so sicher fühlten, dass sie ihr Verständnis nicht hinterfragt haben und auf keinerlei Hilfsmittel zurückgegriffen hätten, auch wenn dies möglich gewesen wäre. Offensichtlich sind sich die Lesenden einiger Verstehensschwierigkeiten des Deutschen gar nicht bewusst. Dass dies allerdings nicht im Besonderen – wie man vielleicht erwarten würde – ein Problem der Personen auf den unteren Deutsch-Niveaus ist, zeigt die Analyse der Selbsteinschätzungen bei allen falsch beantworteten Sätzen über die verschiedenen Deutsch-Niveaus hinweg. Mit steigendem Deutsch-Niveau steigt die Antwortsicherheit auch bei jenen Sätzen deutlich, die falsch beantwortet werden (vgl. Abb. 7). Nach diesen Daten zu urteilen,

²⁰ $\chi^2(1) = 103,471$; $p < 0,001$.

tendieren sogar eher die höheren Niveaus dazu, sich ihrer Sache manchmal zu sicher zu sein. Diese Tendenz gilt im Übrigen genauso für schwierige wie für einfache Sätze (gemessen an der Gesamtfehlerquote).

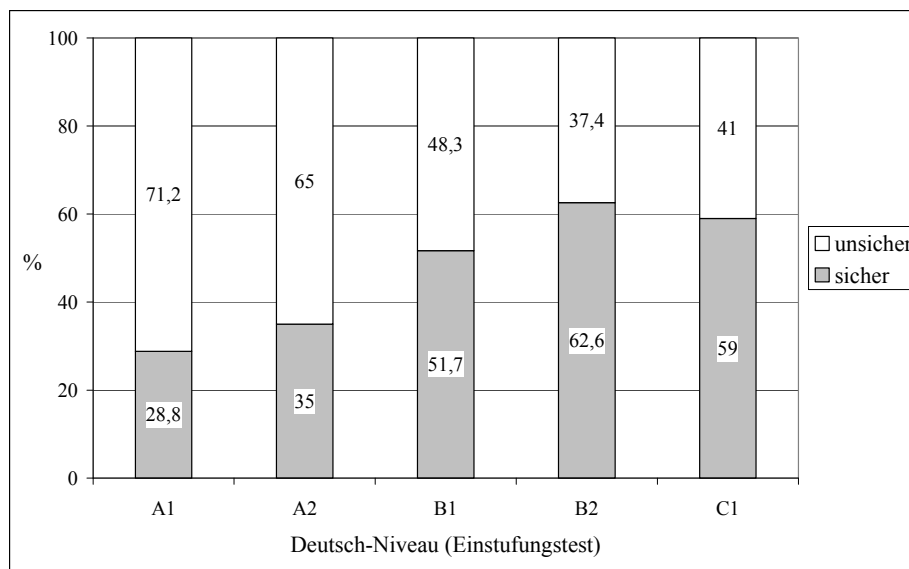


Abbildung 7: (Un-)Sicherheit bei falscher Antwort²¹

Zwischen den Sicherheitsangaben und der Antwortrichtigkeit herrscht besonders bei den allerschwierigsten Sätzen eine deutliche Diskrepanz. Bei den zehn schwierigsten Sätzen fühlen sich tatsächlich in Relation mehr Falsch-Antwortende sicher als Richtig-Antwortende (63,4 % vs. 55,1 %), während bei den anderen Sätzen das Verhältnis – wie eigentlich zu erwarten – umgekehrt ist (44,6 % vs. 74,1 %). Unter den Sätzen, bei denen sich besonders viele Versuchspersonen falsch einschätzten,²² befanden sich sechs Konditionalsätze (ohne und mit Subjunktor) und sechs Subjektsätze (mit und ohne Korrelat). Besonders auffallend ist die falsche Selbsteinschätzung bei zwei Sätzen: einem Konditionalgefüge ohne Subjunktor und einem Subjektsatz ohne Korrelat. Die beiden Sätze lauteten wie folgt:²³

21 A1: n = 73; A2: n = 100; B1: n = 180; B2: n = 278; C1: n = 144.

22 Mehr als 50 % der Versuchspersonen, die falsch antworteten, waren sich bei diesen Sätzen absolut sicher, die richtige Antwort zu geben.

23 Hier werden die Sätze so dargestellt wie in der französischsprachigen Testversion (d.h. mit französischer Interlinearübersetzung der Inhaltswörter).

Konditionalgefüge:

Nach neuesten Forschungen wären die Tiere vom Aussterben bedroht,
nouveau recherche animal disparition menacer
 hätten sie nicht ein ausgeklügeltes Abwehrsystem entwickelt.
ingénieux système de défense développer

Verstehensfrage²⁴:

Les flundodils sont menacés d'extinction.

v□ f□

Subjektsatz:

Widerlegt ist zudem, dass die Männchen und Weibchen gemeinsam die Eier
réfuter mâle femelle ensemble oeuf
 in einer Höhle im Meeresgrund vergraben .
trou fond de la mer enterrer

Verstehensfrage²⁵:

Pendant la période de reproduction, les mâles et les femelles enterrent les œufs ensemble.

v□ f□

Es handelt sich dabei um zwei Sätze mit äußerst hohen Fehlerquoten (78 % bzw. 69,1 %). Dabei gaben 58,9 % der Personen, die die Frage bezüglich des Konditionalgefüges falsch beantworteten, an, absolut sicher zu sein. Hingegen waren von jenen Personen, die letztlich richtig antworteten, nur 23,1 % sicher, die richtige Antwort zu geben. Beim Subjektsatz waren sogar 79,6 % der Personen, deren Antwort auf die Verstehensfrage falsch war, völlig sicher, richtig zu liegen. Jene, die mit ihrer Antwort tatsächlich richtig lagen, hatten dagegen zu 50 % Zweifel an der Korrektheit ihrer Antwort. Was diese beiden Sätze gemeinsam haben, sind offenbar Schwierigkeiten, die leicht übersehen werden können. Im Konditionalgefüge scheint dies in erster Linie der Irrealis zu sein. Auch andere Konditionalgefüge mit Irrealis führten zu deutlich höheren Fehlerquoten im Vergleich zu jenen ohne Konjunktiv. Im Subjektsatz lag das Problem vermutlich darin, dass das Verb im Matrixsatz den Nebensatz verneinte. Da das Verb allerdings interlinear übersetzt war, handelt es sich um kein Wortschatzproblem im engeren Sinne. Das Problem mag vielmehr in der begrenzten Verarbeitungsspanne – besonders bei DaF-Anfängern – liegen, die es verhinderte, dass ein adäquater Zusammenhang zwischen Haupt- und Nebensatz hergestellt wurde.

24 Die Flundodile sind vom Aussterben bedroht. wahr □ falsch □

25 Während der Paarungszeit vergraben die Männchen und Weibchen die Eier gemeinsam. wahr □ falsch □

Was also Frage (4) betrifft, den Zusammenhang zwischen Antwortsicherheit und Korrektheit/Inkorrektheit der Antwort, zeigt sich ein sehr differenziertes Bild. Der globale Zusammenhang zwischen sicherem Antwortverhalten und korrekter Fragebeantwortung ist zwar klar gegeben, dennoch bleibt ein bemerkenswert hoher Anteil an Sicherheit bei falscher Antwort – und zwar auf allen Deutsch-Niveaus und besonders bei den allerschwierigsten Test-Items.

Es soll jedoch an dieser Stelle nicht der Eindruck entstehen, dass sich die Versuchspersonen durchgängig bei allen schwierigen Sätzen überschätzten. Häufig wurden die Schwierigkeiten durchaus erkannt, wie wir sowohl an den Angaben zur Antwortsicherheit als auch in unseren Übersetzungsprotokollen feststellen können. Der Vergleich der Sicherheitsangaben zu den zehn schwierigsten und den zehn einfachsten Sätzen unseres Lesetests (gemessen an der Fehlerquote) zeigt einen klaren Zusammenhang zwischen dem Schwierigkeitsgrad und der Antwortsicherheit der Probandinnen und Probanden ($\chi^2(1) = 73,261$; $p < 0,001$). Das heißt, deutlich mehr Versuchspersonen sind sich bei ihrer Antwort bezüglich eines der einfachsten Sätze sicher (80,5 %) als bezüglich eines der zehn schwierigsten Sätze (59,8 %). Frage (5) lässt sich also klar beantworten: In der Regel gilt, dass bei hohem Schwierigkeitsgrad des Items auch der Grad der Antwortsicherheit hoch ist.

3.3 Gründe für subjektive (Un-)Sicherheit

Es zeigt sich beim Blick auf die Einzelsätze, bei denen sich die Leserinnen und Leser am sichersten bzw. unsichersten fühlten, Ähnliches wie hinsichtlich des Einflusses der Struktur auf die Korrektheit des Leseverstehens: Es spielte in vielen Fällen für die Antwortsicherheit keine Rolle, ob die Personen mit der Ziel- oder der Alternativstruktur konfrontiert waren. So finden sich unter den Sätzen, bei denen sich über 85 % der Probandinnen und Probanden absolut sicher fühlten, sowohl die Aktiv- als auch die Passivversion des gleichen Satzes und Versionen desselben Satzes in der Objekt-Verb-Subjekt-Anordnung und der Subjekt-Verb-Anordnung. Und unter den Einzelsätzen, bei denen sich mindestens 50 % der Lesenden unsicher fühlten, befinden sich Versionen eines Konditionalgefüges sowohl mit als auch ohne Subjunktior, genauso wie die Versionen desselben Satzes mit und ohne Satzklammer.

Weshalb haben sich nun besonders viele Personen bei gewissen Sätzen sicher bzw. unsicher gefühlt (s. Frage 6)? Betrachtet man jene Sätze etwas genauer, bei denen sich die Mehrheit der Lesenden unsicher fühlte, so stellt man fest, dass es sich um Sätze mit mittlerer bis höherer Fehlerquote handelt (nicht dabei sind allerdings die Sätze mit den allerhöchsten Fehlerquoten, siehe vorne). Bei der

Mehrheit der Sätze handelt es sich um längere Gefüge aus zwei Teilsätzen. Und in den meisten Sätzen kann man eine besondere Schwierigkeit isolieren, die – wie die Übersetzungsprotokolle zeigten – in der Regel auch bemerkt wurde. Zu diesen Schwierigkeiten gehörten etwa Funktionswörter, die nicht verstanden wurden, Unsicherheiten bei der funktionellen Zuordnung oder beim Erkennen von Flexionsformen oder der richtige Anschluss eines Relativsatzes. Eine ungewöhnliche Semantik kann diese Probleme noch verschärfen. Manchmal lag allerdings die Schwierigkeit auch in der Verstehensfrage, so erforderte eine Frage z.B. eine Inferenzleistung der Lesenden.

Auf Seiten der Sätze, bei denen sich die große Mehrheit (mindestens 85 %) der Lesenden sicher fühlten, finden sich hingegen mehrheitlich einfache Sätze (im grammatischen Sinne). Bei diesen Sätzen handelt es sich um Sätze mit äußerst geringen Fehlerquoten (von 0 % bis 9,2 %). In der Regel lässt das Kombinieren von (einigen oder allen) Inhaltswörtern bei diesen Sätzen keinen Raum für Missverständnisse und führt damit zum richtigen Beantworten der Verstehensfrage.

4. Zusammenfassung

Die italophonen und frankophonen Versuchspersonen unserer Studie scheinen in Bezug auf ihre Deutsch-Lesekompetenz sehr selbstkritisch zu sein: Eine große Mehrheit der Versuchspersonen stuft ihr Leseverständnis in Deutsch eine GeR-Stufe unterhalb ihres mit Hilfe eines Einstufungstests gemessenen allgemeinen Deutsch-Niveaus ein. Über die Gründe dafür kann hier nur spekuliert werden. Sie mögen z.B. in der Persönlichkeitsstruktur der Versuchspersonen liegen oder aber im Ruf des Deutschen als schwer zu erlernende Sprache. Besonders bescheiden in der Einschätzung ihrer Lesekompetenz sind in unserer Studie Lernende auf höheren Deutsch-Niveaus. Auf Stufe B2 schätzen gut vier Fünftel der Versuchspersonen ihre Lesekompetenz eine Stufe unter dem getesteten allgemeinen Deutsch-Niveau ein und auf Stufe C1 sind es dann so gut wie alle. Da wir keinen standardisierten Lesetest durchgeführt haben, können wir aufgrund dieser Beobachtungen jedoch nicht endgültig darauf schließen, dass unsere Versuchspersonen auf den höheren Deutsch-Niveaus ihre Lesekompetenz zumeist unterschätzen. Unsere Daten scheinen aber dennoch in diese Richtung zu deuten, zumal das Phänomen der sich unterschätzenden fortgeschrittenen Lernenden in Studien zum Leseverstehen von Alderson (2005) und Brantmeier/Vanderplank (2008) bereits beobachtet wurde.

Nebst dem Vergleich von selbst eingeschätzter Lesekompetenz und allgemeinem Deutsch-Niveau interessierte uns insbesondere der Vergleich von selbst eingeschätzter Lesekompetenz und Verständnis des eigens konzipierten Lesetests.

Dieser war so gestaltet, dass er uns Aufschluss über die rezeptive Schwierigkeit verschiedener grammatischer Strukturen sowie die Wahrnehmung dieser Schwierigkeit durch die Leserinnen und Leser geben konnte. Korrelationsanalysen zeigten, dass die selbst eingeschätzte Lesekompetenz nur schwach mit den Resultaten des Lesetests korreliert, schwächer als mit den Resultaten des Deutsch-Einstufungstests. Da dieser allgemeine Deutsch-Test zu einem großen Teil grammatisches Wissen erfasst, sehen wir dies als eine Bestätigung dafür, dass unser Lesetest tatsächlich vor allem das Verständnis der grammatischen Strukturen misst.

Neben dem Verhältnis von selbst eingeschätzter, allgemeiner DaF-Lesekompetenz und objektiven Messgrößen interessierte uns insbesondere die subjektiv empfundene Sicherheit der Versuchspersonen bei der Beantwortung der einzelnen Verstehensfragen und ihr Zusammenhang mit der Deutsch-Kompetenz sowie der Korrektheit der Antwort. Hierzu lässt sich zunächst einmal feststellen, dass sich die Mehrheit der Testpersonen zumeist sehr sicher fühlt. Die Sicherheit nimmt zudem bei steigendem Deutsch-Niveau stetig zu. Lediglich auf Niveau A1 fühlen sich die Versuchspersonen zumeist unsicher. Im Weiteren bestätigte sich die Vermutung, dass die Sicherheit mit der Korrektheit der Antwort zusammenhängt: Bei falsch verstandenen Sätzen sind sich die Testpersonen in der Regel ihrer Antwort weniger sicher als bei richtig verstandenen Sätzen. Viele grammatische Schwierigkeiten beim Lesen in Deutsch als Fremdsprache werden demnach durchaus als solche wahrgenommen. Dennoch ist die Sicherheitsquote von 52,6% der Personen, die schließlich falsch antworteten, überraschend hoch. Besonders bei den Items, die sich in unserem Test als sehr schwierig erwiesen, zeigte sich das Phänomen, dass viele der Versuchspersonen offenbar die darin liegende Schwierigkeit nicht erkannten und sich bei der Beantwortung der Verstehensfrage sicher fühlten. Einige grammatische Phänomene (wohl nicht nur des Deutschen) bergen demnach offenbar Schwierigkeiten, die den fremdsprachlichen Leserinnen und Lesern nur wenig bewusst sind. Zu denken ist dabei u.a. an den Irrealis oder an die Möglichkeit einer komplexen semantischen Verschränkung zwischen Teilsätzen eines Satzgefüges. Erstaunlicherweise wiegen sich tendenziell nicht die Anfängerinnen und Anfänger, sondern die fortgeschrittenen Lernenden besonders oft in falscher Sicherheit.

Gerade auch aufgrund dieser Feststellung ist die selbst eingeschätzte Lesekompetenz unserer Meinung nach nicht unbedingt als Ersatz für eine 'objektive' Messung zu sehen, sondern vielmehr als eine wertvolle Ergänzung dazu. So können etwa besonders gezielt gestellte Fragen zur Selbsteinschätzung und Sicherheit in Bezug auf bestimmte Phänomene weiterführende Auskunft darüber geben, welche Probleme erst ins Bewusstsein der Versuchspersonen gehoben werden müssen, um überhaupt als solche angegangen werden zu können. Besonders in Bezug auf das Leseverstehen scheinen solche Rückmeldungen an die Lernenden

wichtig, da andernfalls Falschverstehen oftmals unbemerkt und unkorrigiert bleiben kann. Mit dieser Untersuchung sollte ein Anstoß gegeben werden, besonders diese Linie weiterzuverfolgen. Doch auch weitere Untersuchungen hinsichtlich der vielfältigen Zusammenhänge von objektiven und subjektiven Kompetenzmaßen und potenzieller Einflussvariablen wären selbstverständlich wünschenswert.

Eingang des revidierten Manuskripts: 28.7.2010

Literaturverzeichnis

- Alderson, J. Charles (2000), *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J. Charles (2005), *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London/New York: Continuum.
- Bachman, Lyle F. & Palmer, Adrian S. (1989), The construct validation of self-ratings of communicative language ability. *Language Testing* 6: 1, 14-29.
- Bernstein, Wolf Z. (1990), *Leseverständnis als Unterrichtsziel. Sprachliches und methodisches Grundwissen für den Lehrer im Fach Deutsch als Fremdsprache*. Heidelberg: Julius Groos.
- Blanche, Patrick (1988), Self-assessment of foreign language skills: implications for teachers and researchers. *RELC Journal* 19, 75-93.
- Blue, George M. (1988), Self-assessment – the limits of learner independence. In: Brookes, Arthur (Hrsg.) (1988), *Individualisation and Autonomy in Language Learning*. ELT Documents 131. London: Modern English Publications/British Council, 100-118.
- Blue, George M. (1994), Self-assessment of foreign language skills: does it work? *CLE Working Papers* 3, 18-35.
- Brantmeier, Cindy (2005), Nonlinguistic variables in advanced second language reading: learners' self-assessment and enjoyment. *Foreign Language Annals* 38: 4, 494-504.
- Brantmeier, Cindy (2006), Advanced L2 learners and reading placement: self-assessment, CBT, and subsequent performance. *System* 34: 1, 15-35.
- Brantmeier, Cindy & Vanderplank, Robert (2008), Descriptive and criterion-referenced self-assessment with L2 readers. *System* 36: 3, 456-477.
- Chalhoub-Deville, Micheline & Deville, Craig (1999), Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics* 19, 273-299.
- DIALANG (*Diagnostic Language Test*), www.dialang.org.
- Europarat (Hrsg.) (2001), *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lehren, lernen, beurteilen*. Berlin et al.: Langenscheidt.
- Fabricsius-Hansen, Cathrine (2002), Texte in der Fremdsprache lesen und verstehen: Überlegungen zu einem vernachlässigten Thema. *SPRIKreports* 16, 1-17.
- Falchikov, Nancy & Boud, David J. (1989), Student self-assessment in higher education: a meta-analysis. *Review of Educational Research* 59: 4, 395-430.
- Harris, Michael (1997), Self-assessment of language learning in formal settings. *English Language Teaching Journal* 51: 1, 12-20.
- Heringer, Hans Jürgen (1987), *Wege zum verstehenden Lesen. Lesegrammatik für Deutsch als Fremdsprache*. München: Hueber.
- Heringer, Hans Jürgen (2001), *Lesen lehren lernen: eine rezeptive Grammatik des Deutschen*. 2. durchgesehene Auflage. Tübingen: Niemeyer.

- Janssen-van Dieten, Anne-Mieke (1989), The development of a test of Dutch as a foreign language: the validity of self-assessment by inexperienced subjects. *Language Testing* 6: 1, 30-46.
- Janssen-van Dieten, Anne-Mieke (2000), Alternative assessment: self-assessment beyond the mainstream. *Melbourne Papers in Language Testing* 9: 2, 18-29.
- Kruger, Justin & Dunning, David (1999), Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77, 1121-1134.
- LeBlanc, Raymond & Painchaud, Gisèle (1985), Self-assessment as a second language placement instrument. *TESOL Quarterly* 19: 4, 673-687.
- Lutjeharms, Madeline (1998), Die syntaktische Verarbeitung bei der Rezeption von Sprache. In: Klein, Eberhard; Schierholz, Stefan J. (Hrsg.) (1998), *Betrachtungen zum Wort*. Tübingen: Stauffenburg, 117-152.
- MacIntyre, Peter D.; Noels, Kimberly A. & Clément, Richard (1997), Biases in self-ratings of second language proficiency: the role of language anxiety. *Language Learning* 47, 265-287.
- Moritz, Christine E. B. (1996), Student Self-Assessment of Language Proficiency: Perceptions of Self and Others. Paper presented at the Annual Meeting of the American Association for Applied Linguistics (18th, Chicago, IL, March 23-26, 1996).
- Oscarson, Mats (1997), Self-assessment of foreign and second language proficiency. In: Clapham, Caroline; Corson, David (Hrsg.) (1997), *Language Testing and Assessment*. Dordrecht et al.: Kluwer Academic Publishers, 175-187.
- Pearce, Bonny N.; Swain, Merrill & Hart, Doug (1993), Self-assessment, French immersion and locus of control. *Applied Linguistics* 14: 1, 25-42.
- Poglia, Edo; Galeandro, Cristina; Stoks, Gé & Tamagni Kathya (2009), Compétences en langues secondes. Acquisition, utilisation et conséquences identitaires, professionnelles et sociales pour les jeunes tessinois. Rapport final. http://www.nfp56.ch/d_projekt.cfm?Projects.Command=details&get=12&kati=2 [15.08.2010].
- Raasch, Albert (1979), Selbsteinschätzungstests für den Lerner: eine Hilfe für den Französischlehrer. *Zielsprache Französisch* 1, 1-7.
- Ross, Steven (1998), Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing* 15: 1, 1-20.
- Stalb, Heinrich (1993), *Deutsch für Studenten: Lesegrammatik*. Ismaning, München: Verlag für Deutsch.
- Urquhart, Sandy & Weir, Cyril J. (1998), *Reading in a Second Language: Process, Product and Practice*. London, New York: Longman.
- Van Onna, Bert & Jansen, Carel (2006), How multilingual are the dutch really? On proficiency in Dutch, English, French, and German in Dutch organizations. *BELL – Belgian Journal of English Language and Literatures* 4, 169-180.
- Vogel, Klaus & Vogel, Sigrid (1980), Möglichkeiten und Grenzen von Selbsteinschätzungstests: Bemerkungen zu ihrer Konstruktion und Verwendung im Französischunterricht. *Zielsprache Französisch* 2, 81-89.
- Wesche, Marjorie; Paribakht, Sima. T. & Ready, Doreen (1993), A Comparative Study of Four Placement Instruments. Paper presented at the Annual Language Testing Research Colloquium (15th, Cambridge, England, and Arnhem, The Netherlands, August 2-8, 1993).
- Young, Dolly Jesusita (2000), An investigation into the relationship between L2 reading anxiety and L2 reading comprehension, and self-reported level of comprehension, topic familiarity, features of an L2 text and reading ability in the L1 and L2. In: Leow, Ronald P. & Sanz, Cristina (Hrsg.) (2000), *Spanish Applied Linguistics at the Turn of the Millennium*. Somerville, MA: Cascadilla Press, 15-23.