



Research paper

Teachers' judgment accuracy of students' subjective well-being in school: In search of explanatory factors



Carmen L.A. Zurbriggen^{a,*}, Lena Nusser^b, Mireille Krischler^c, Monja Schmitt^b

^a University of Fribourg, Switzerland

^b Leibniz Institute for Educational Trajectories, Bamberg, Germany

^c University of Luxembourg, Luxembourg

ARTICLE INFO

Keywords:

Subjective well-being
Teacher judgments
Judgment accuracy
CT-C(M-1) model

ABSTRACT

This study investigated teachers' judgment accuracy of students' subjective well-being and characteristics explaining the specificity in teacher reports (i.e., non-agreement with student self-reports). We drew from the self-reports of 2592 grade six students and the ratings of 432 teachers. Findings show that teachers' judgement accuracy of students' subjective well-being is low to moderate, indicating considerable specificity in teacher reports. Students' gender, special educational needs and achievement predicted the specificity. Teaching experience was negatively associated with the specificity for academic self-concept, while teachers' self-efficacy, attitudes towards inclusion and responsibility were positively related to the specificity for emotional well-being and social inclusion.

Students' well-being is considered both an essential prerequisite for various educational outcomes and an outcome in itself. For instance, numerous studies have reported positive associations between students' well-being with their academic achievement (Amholt et al., 2020), psychosocial adjustment (Piqueras et al., 2019) and global life satisfaction (Suldo et al., 2011). Recognized as an important educational outcome next to academic achievement, school-related well-being has gained considerable attention both in research and educational practices over the last decade (e.g., Carter & Andersen, 2019; Hascher, 2010; Orkibi et al., 2014; Suldo, 2016).

It might not be immediately obvious, but students' well-being also serves as a central indicator for teaching, inasmuch as it affects a teacher's decisions on how to design classroom interventions and instructions (Pielmeier et al., 2018). By deriving a student's well-being via observing behaviours or other cues, teachers implicitly or explicitly adapt their teaching according to the students' individual differences and social-emotional needs. Adapting teaching methods by accounting for inter-individual differences is generally considered a key component of effective teaching (Fairbanks et al., 2010; Parsons et al., 2018). Based on their judgments about students' characteristics, teachers adjust their communication, teaching material or instructional strategies to better support students in their learning processes (Helmke & Schrader, 1987; Urhahne & Wijnia, 2021). Due to the increasing diversity of the student

body – not to mention the expansion of inclusive education around the globe – adaptive teaching has become even more relevant in today's education (Hardy et al., 2019; Schipper et al., 2017).

Accurate teacher judgment is supposed to be a necessary condition for adaptive teaching, which in turn promotes each student's learning and academic development (Machts et al., 2016; Praetorius et al., 2015). When determining whether a student is feeling well at school, teachers frequently (need to) rely on their own judgment, although subjective well-being corresponds to a person's inner state and should be preferably assessed by means of self-report (Eid, 2018; Pavot, 2008; Schwarz & Strack, 1999). However, teachers' accuracy in judging different aspects of students' well-being is known to be low or at least only moderate (cf. Review by Urhahne & Wijnia, 2021).

Against this backdrop, the present study aimed first to investigate the accuracy of teacher reports of students' subjective well-being in school. Considering the broad scope of the construct of subjective well-being and its multidimensionality (Hascher, 2010), we focused on three components: emotional well-being in school, social inclusion in class and the academic self-concept. Prior research has indicated that certain student characteristics – more specifically, gender, special educational needs (SEN) and academic achievement – may induce a bias in the teachers' estimates of the three aforementioned components (Schwab et al., 2020; Urhahne & Zhu, 2015; Venetz et al., 2019). Although

* Corresponding author. University of Fribourg, 21 Rue St. Pierre-Canisius, 1700 Fribourg, Switzerland.

E-mail address: carmen.zurbriggen@unifr.ch (C.L.A. Zurbriggen).

theoretically proposed by Funder's (1995; 2012) realistic accuracy model, studies on teachers' characteristics relating to their judgment accuracy of students' subjective well-being are scarce. Thus, the second goal of this study was to examine whether specific student and teacher characteristics can explain teachers' judgment (in)accuracy.

1. The relevance of teachers' judgment accuracy

Teacher judgment plays an important role in students' academic and personal development in many respects. First and foremost, teachers' assessment of students' academic achievement has implications for grades, grade retention and placement decisions (e.g., regarding ability groups) and can thus crucially influence students' educational progress and trajectories (Begeny et al., 2008, 2011; Francis et al., 2017; Parsons & Hallam, 2014; Südkamp et al., 2012; Wiliam & Bartholomew, 2004). In the shorter term, a teacher's evaluation of students' academic and social-emotional characteristics serves as a basis for various instructional decisions and teaching practices (Alvidrez & Weinstein, 1999; Clark & Peterson, 1986; Hoge & Coladarci, 1989), such as the selection of classroom activities, the difficulty of assigned tasks or the compilation of student learning groups. Moreover, studies have revealed that teachers' perceptions of students' characteristics are associated with students' social-emotional outcomes. Students who were perceived as more competent by their teachers were reported to be more motivated in school compared to students who were underestimated by their teachers (Urhahne, 2015).

The accuracy of teacher judgment relates to the teacher's ability to adequately assess relevant student characteristics and appropriately evaluate learning requirements (Artelt & Gräsel, 2009). Given the important implications, teacher judgment competence is considered a core competence of teachers' professional skills (Artelt & Gräsel, 2009; Herppich et al., 2018). According to Helmke and Schrader (1987), high judgment or assessment accuracy – combined with a high-frequency use of appropriate instructional techniques (e.g., providing structuring cues or individual support) – is a critical aspect of teaching and is particularly relevant for a student's educational progress.

Given the significant role played by teachers' judgment on students' characteristics, numerous studies have examined its accuracy. The main focus of related research has been on academic achievement. Corresponding findings were synthesized in an early meta-analysis by Hoge and Coladarci (1989) and in a more recent one by Südkamp and colleagues (2012), showing a positive and fairly high correlation between teacher judgments of student achievement and students' actual achievement (i.e., relative accuracy). However, until now, only a few studies have specifically dealt with the accuracy of teachers' judgments of their students' subjective well-being.

2. Teachers' judgment accuracy of students' subjective well-being

In general terms, subjective well-being can be described as "people's cognitive and affective evaluations of their lives" (Diener, 2000, p. 34). Most often, however, subjective well-being is defined as a broad area of interest rather than a specific concept (Diener et al., 2005; Pavot, 2008), and its application vary considerably depending on the context or field. Hascher (2010; 2012) defined students' subjective well-being *in school* as a multi-component construct encompassing positive attitudes toward school, enjoyment in school, the absence of worries related to school, a positive academic self-concept, the absence of social problems in school as well as the absence of physical complaints in school. The appropriateness of a differentiated understanding of students' well-being with several meaningful components has been underpinned by differential empirical findings (Hascher, 2010), particularly, in relation to teachers' judgment accuracy of students' self-reports concerning inclusion (Schwab et al., 2020; Venetz et al., 2019).

To determine the degree of teacher judgment accuracy related to

subjective well-being and other internal student characteristics, teacher reports are compared with students' self-reports. Compared to student achievement, internal constructs are way more difficult to judge for external observers, as there is no objective measure available. Thus, teachers need to rely on hints related to certain aspects of subjective well-being that are not necessarily easy to detect. Based on this methodological peculiarity, *agreement* has been proposed as an alternative term for *judgment accuracy* of more internal constructs, such as subjective well-being, emotions, motivation or academic self-concept (e.g., Friedrich et al., 2013; Urhahne & Wijnia, 2021).

One of the few studies that explicitly address teachers' judgment accuracy of students' *subjective well-being* is the one by Urhahne and Zhu (2015). More specifically, they examined the teachers' skills to not only assess students' academic self-concept and attitudes towards school but also their emotional and social aspects, such as enjoyment at school, worries about school and social problems. Their findings revealed only small within-class correlations (i.e., rank components) between the teachers' judgments and the self-reports of students' worries, enjoyment, physical complaints and social problems, indicating that teachers "were not able to estimate students' well-being very well" (Urhahne & Zhu, 2015, p. 230).

Teachers' judgment of the students' *academic self-concept* as the cognitive aspect of subjective well-being has been studied more intensely compared to students' social and emotional aspects. Various studies documented a moderate accuracy for teachers' judgments about students' academic self-concept (e.g., Givvin et al., 2001; Pielmeier et al., 2018; Praetorius et al., 2013; Urhahne et al., 2011). Nevertheless, when teachers were informed about the students' achievement, they assessed the students' academic self-concept more accurately (Helm et al., 2018).

In studies on judgement accuracy related to *emotional well-being*, the term has often not been referred to or indicated as such. Most previous research on judgement accuracy have considered specific negative or positive emotions (e.g., test anxiety, worries about school, enjoyment), which were frequently related to other constructs such as *motivation*. Givvin et al. (2001) investigated teachers' judgments about students' motivation in math by assessing positive and negative emotions, learning orientation and academic self-concept. Regarding emotions, they found small correlations between teacher ratings and self-reports of students' positive and negative emotions in math generally but higher correlations, and thus higher accuracy for specific math contexts (e.g., fractions). Furthermore, their findings revealed that teachers made more global judgments of students' characteristics while students differentiated more in their self-perceptions regarding positive and negative emotions, learning orientation and academic self-concept (Givvin et al., 2001).

The findings regarding different *social aspects of subjective well-being* are inconclusive, ranging from small to relatively high accuracy or agreement levels. While the previously cited study by Urhahne and Zhu (2015) reported low correlations for social problems, the study by Gomez (2014) demonstrated moderate correlations between teacher ratings and adolescents' self-reports of prosocial behaviour and even high correlations in terms of peer problems. Considering the peer ratings of social aspects, studies showed small to moderate positive correlations between teacher- and peer-assessed social skills (Kwon et al., 2012; Südkamp et al., 2018) and moderate correlations regarding social acceptance assessed via peer nominations (Südkamp et al., 2018).

3. Factors explaining teachers' judgment accuracy

As outlined in the previous section, teachers' judgments of students' subjective well-being are not necessarily accurate or compliant with students' self-reports. This raises the question: which factors can be assumed to affect teachers' judgement accuracy?

Südkamp and colleagues (2012) proposed a heuristic model of teacher judgment accuracy that postulates that *judgment, test, student*

and *teacher* characteristics are factors with theoretical relevance for teachers' judgment accuracy, possibly influencing the correspondence between teachers' judgments of students' academic achievement and the students' actual test performance. Although the model by Südkamp and colleagues (2012) was developed in terms of academic achievement, it can also serve as a basis for explaining the judgment accuracy of subjective well-being. Nevertheless, one important difference has to be noted. While studies related to students' academic achievement typically rely on achievement tests as a reference for assessing teachers' judgement accuracy, emotional and motivational characteristics as well as certain aspects of subjective well-being are generally assessed by means of students' self-reports and then compared with the teacher ratings (Urhahne & Wijnia, 2021). This methodological peculiarity is one of the main factors responsible for the fact that teachers' judgement accuracy is usually lower for students' subjective well-being than for academic achievement (Südkamp et al., 2012; Zhu & Urhahne, 2014). As it corresponds to an internal state that is not easily observable, teachers are facing more difficulties in finding suitable information to accurately assess students' subjective well-being.

Funder's realistic accuracy model (1995; 2012) illustrates the process of accurate personality judgement and the circumstances under which accurate judgements are most likely to occur. Funder identified four key conditions of judgment accuracy – relevance, availability, detection and utilization – which evolve in a complex interplay between the “environment” and the “perceiver”. In a nutshell, “accuracy in such judgments is achieved when relevant behavioral information is available to and detected by a judge, who then utilizes that information correctly” (Funder, 2012, p. 177). That is, if behavioural cues are not available or not used properly, the teachers' reports are most likely to not result in accurate judgements. Accordingly, the visibility of traits in terms of behavioral cues proved to be a strong predictor of accuracy (Kenny & West, 2010). Thus, not all aspects of school-related well-being are equally observable or visible. For instance, the academic self-concept includes better observable cues than emotional well-being, which is reflected in a higher judgement accuracy for the academic self-concept than for emotional aspects (Urhahne & Wijnia, 2021).

The realistic accuracy model also postulates basic moderators that determine the degree of judgment accuracy (Funder, 2012). The moderating variables can be grouped into four categories: *target*, *judge*, *trait* and *information*. In the following, we focus on factors related to the “target” and the “judge”, which in our case means a focus on student and teacher characteristics.

3.1. Student characteristics explaining teachers' judgment accuracy

Research has shown that teacher judgment accuracy of students' subjective well-being at school is affected by the students' gender. Venetz and colleagues (2019) found small positive effects of gender on teacher ratings of eighth-grade students' academic self-concept, i.e., teachers overestimated girls' self-concept compared to boys. In Grade four, a similar gender bias was shown for emotional well-being and social inclusion (Schwab et al., 2020). In the study by Urhahne and Zhu (2015), teachers perceived gender differences with regard to students' attitudes towards school, worries and physical complaints.

Furthermore, several findings indicate that teachers tend to systematically underestimate students belonging to marginalized groups, for example, those from ethnic minorities or those having special needs (Meissel et al., 2017; Rubie-Davies et al., 2012). However, teachers tended to overestimate bilingual students' performance in linguistically complex tasks compared to monolingual students with a migrant background or students of the majority language (Hachfeld et al., 2010). In terms of subjective well-being, Venetz and colleagues (2019) showed a small negative effect for emotional well-being and a moderate negative effect for the academic self-concept, which could be interpreted as a negative bias towards students with SEN compared to students without SEN. Further results highlighted that teacher ratings of students'

emotional well-being and academic self-concept, but also regarding social inclusion, seem to be negatively biased for students with SEN (Schwab et al., 2020). In terms of social acceptance, a similar negative bias for students with SEN was found in the study by Südkamp et al. (2018).

The meta-analysis of Machts et al. (2016) revealed that teachers' judgment accuracy of students' cognitive abilities is strongly associated with students' academic achievement. Whether teacher reports of students' subjective well-being are also biased towards students' academic achievement is an open question. Nevertheless, some important clues are provided in the study by Zhu and Urhahne (2021). Their findings showed that teacher judgement of students' emotion and academic self-concept (as one indicator of motivation) was a good predictor of teacher judgement of students' achievement, while self-reported students' emotion and motivation could predict actual students' achievement far less accurately. Furthermore, teachers tended to judge students' achievement, motivation and emotion similarly (indicated by high correlations), while students differentiated more between the constructs. In the light of these findings, it can be assumed that students' academic achievement might explain to some extent teachers' judgement accuracy of students' subjective well-being.

3.2. Teacher characteristics explaining teachers' judgment accuracy

Teacher characteristics are thought to influence their judgment at various stages of the judgment process. In literature, a broad range of characteristics, such as job experience (Impara & Plake, 1998), beliefs (Shavelson & Stern, 1981) and professional goals, have been associated with teachers' judgment processes (Rogers & Biesanz, 2019). As *teachers' professional expertise* results from medium-to long-term systematic and reflective teaching experiences (Baumert & Kunter, 2006), teachers with more professional experience have been shown to be better able to judge their students' performance (McElvany et al., 2009). Accordingly, the more experienced the teachers are, the better they can integrate contradictory issues, which leads to more accurate judgments (Krolik-Schwerdt et al., 2009). However, more recent studies have also shown that more experienced teachers tend to judge students' work more strictly compared to novice teachers or student teachers (Jansen et al., 2021; Möller et al., 2022). In the context of inclusive education, it may be beneficial that teachers possess experience or feel well-prepared in teaching students with SEN or disability (Gilmour & Wehby, 2019). Thus, teachers with more professional experience should be able to judge their students' subjective well-being more accurately than teachers with less professional expertise. However, according to a study by Praetorius et al. (2011), teachers' professional experience was found to not affect their judgement accuracy of students' academic self-concept.

In search of relevant teacher characteristics related to the judgement accuracy of students' subjective well-being, it may be fruitful to consider variables that have been proposed to be important for successful adaptive teaching in (inclusive) classrooms with a heterogeneous student body. Presumably, the two most frequently studied constructs in this regard are teachers' self-efficacy (Klassen & Tze, 2014; Van Mieghem et al., 2020) and teachers' attitudes towards inclusion (Yada et al., 2022). *Self-efficacy* in general terms is defined as a person's subjective expectation about how they will manage to attain their goals based on their individual abilities (Bandura, 1997). In the context of teaching, self-efficacy can be described as teachers' beliefs in their capabilities to positively influence students' engagement and learning, “even among those students who may be difficult or unmotivated” (Tschannen-Moran & Woolfolk Hoy, 2001, p. 783). Teachers' self-efficacy is thought to be highly relevant for effective teaching behaviours in general (Zee & Koomen, 2016), the adoption of new teaching strategies (Ghaith & Yaghi, 1997) and the implementation of inclusive education (e.g., Bosse & Spörer, 2014; Savolainen et al., 2020; Sharma et al., 2018). According to a review by Zee and Koomen (2016), teachers' self-efficacy is

positively related to several teacher characteristics, including a higher level of commitment to teaching. Following this line of thought, it could be argued that the higher a teacher's self-efficacy in dealing with diverse educational needs, the more willing they would be to spend (more) effort in accurately assessing each student's subjective well-being.

Teachers' attitudes toward inclusion are argued to play a crucial role in the successful implementation of inclusive education (Avramidis & Norwich, 2002; de Boer et al., 2011; Van Miegheem et al., 2020). For instance, positive attitudes towards inclusion are associated with more diverse teaching practices that are aimed at adapting the learning environment to the needs of the students (Hellmich et al., 2019; Monsen et al., 2014). To implement adaptive teaching successfully, teachers need to accurately assess their students' specific needs. Thus, teachers with a positive attitude towards inclusion might be more attuned to their students' well-being on a regular basis.

Similarly, theoretical considerations suggest that *teachers' responsibility towards every student* might influence their judgement accuracy of the students' well-being. Teacher responsibility incorporates a sense of inner obligation and commitment to achieve or prevent certain outcomes (Lauermaann & Karabenick, 2011, 2013). According to Jussim et al. (1996), judgment bias is less likely when a person (the "judge") is motivated to develop an accurate impression of another person (the "target") or when the judge's outcomes depend on the target. Being responsible for a judgment with important consequences may also result in high attention to and a careful integration of all available information (Lee et al., 1999; Lerner & Tetlock, 1999). Thus, it could be deduced that a high sense of responsibility for every student in a classroom may affect the teachers' judgement accuracy of students' subjective well-being.

4. The present study

The present study aimed to investigate teachers' judgment accuracy of students' subjective well-being in school and to examine whether specific student and teacher characteristics could explain the judgment accuracy. Considering the broad scope of the construct subjective well-being in school and its multidimensionality (Hascher, 2010), we focused on three components: emotional well-being, social inclusion and academic self-concept.

First, we investigated the teachers' judgment accuracy of students' subjective well-being in school by analysing the consistency between teacher reports and students' self-reports (i.e., agreement or shared view) of the three components as well as the specificity of the teacher reports (i.e., non-agreement or the part in teacher reports that is not shared with the student reports). Based on prior results, we hypothesised that the consistency would be rather low for emotional well-being and social inclusion (Schwab et al., 2020; Venetz et al., 2019) and moderate for academic self-concept (Praetorius et al., 2015; Urhahne & Wijnia, 2021).

Second, building upon our first research question and following Funder's (1995; 2012) realistic accuracy model, we examined whether the specificity in teacher reports (i.e., non-agreement between student and teacher reports) could be explained by student characteristics. In previous studies, students' gender in particular has been empirically associated with teachers' judgment accuracy (e.g., Gentrup et al., 2018; Südkamp et al., 2012). Past findings led us to further expect that the status SEN (Schwab et al., 2020; Venetz et al., 2019), students' language backgrounds (Hachfeld et al., 2010) and academic achievement (Zhu & Urhahne, 2021) would be linked to the teachers' judgement accuracy of students' subjective well-being at school. Finally, we accounted for teacher characteristics that may explain the judgment accuracy of students' well-being. Based on theoretical considerations, we assumed that teachers' professional experience, their self-efficacy, their responsibility for every student and their attitudes towards inclusion might be associated with the specificity in teacher reports regarding students' subjective well-being. So far, there has been no empirical evidence underpinning this assumption.

5. Method

5.1. Participants and procedure

This article draws on data from the first measurement occasion of the first cohort of the longitudinal project Inclusive Education in Lower Secondary Schools in Germany (INSIDE; Schmitt et al., 2020). The Federal Ministries of Education in Germany approved the study. Furthermore, ethical standards were approved by the responsible persons within the data protection team. Participation in the study was voluntary. In accordance with the Declaration of Helsinki, the corresponding educational institution, the students and their parents gave informed consent to participate in the study. The paper-based survey was conducted in the school context and was guided and supervised by trained personnel in a standardised manner. Homeroom teachers (i.e., the main responsible teacher for the class) were also given a questionnaire to provide information about themselves (i.e., teacher characteristics) and to assess the subjective well-being of all students in their class who participated in the study so that the teachers' information could be attributed to each student.

The initial student sample consisted of 3772 grade six students (49.5% girls) from 622 classes in 231 schools. On average, six students per class participated in the study (after obtaining the parents' consent). Students on average were 12.6 years old ($SD = 0.6$). For 11.5% of the students, German (the teaching language) was not their primary language. Some 11.9% had received the diagnosis of SEN. In total, 432 homeroom teachers responded to the questionnaire, corresponding to a response rate of 69.5%. Moreover, they were asked to judge the subjective well-being of all students in their class participating in the study. Most of the teachers were female (77.2%). On average, teachers were 42 years old ($SD = 10.8$). Thus, complete data for both student and teacher reports are available for 2592 grade six students from 432 classes, with an average of six students to be rated by teachers per class. Compared with the initial sample, the analysis sample thus has a reduced size. This is mainly due to the fact that the participation rate of teachers was not as high as that of the children.

5.2. Measures

Student's subjective well-being. The German version of the Perception of Inclusion Questionnaire (PIQ; Venetz et al., 2015) was utilised to assess the students' emotional well-being, social inclusion and academic self-concept from both the student's and teacher's perspectives. The student version (PIQ-S) has been validated repeatedly (Knickenberg et al., 2022; Venetz et al., 2014). Among others, measurement invariance across the groups of several student characteristics (e.g., learning difficulties, primary language, gender) has been established (DeVries et al., 2018; Knickenberg et al., 2020; Venetz et al., 2014; Zurbruggen et al., 2019). Internal consistency of the three scales of the teacher version (PIQ-T) proved to be good (Schwab et al., 2020; Venetz et al., 2019). In the present study, each construct was assessed by three Likert-type items (0 = *not at all true*, 1 = *rather not true*, 2 = *somewhat true*, and 3 = *certainly true*) instead of four, as in the original PIQ scales. The scales were primarily shortened for reasons of research economics. Due to a possible method bias for students with learning disabilities, the three negatively-worded items (one in each scale) were selected for the reduction (Knickenberg et al., 2020).

Status SEN. To identify students with SEN, teachers were asked about the official status according to the local school authority.

Student's gender. Gender was assessed via teacher reports as a dichotomous variable, "girls" and "boys". Missing information was replaced with information taken from the student or parent questionnaire.

Student's language background. Teachers reported on the primary language that the students were speaking at home. The variable was dichotomised as "German" and "other languages" since German was the

teaching language in all classes.

Student's academic achievement. We used two indicators to account for academic achievement: reading comprehension and mathematics achievement. Both tests are based on test development for the longitudinal assessment of competences within the Nationale Educational Panel Study (NEPS, Artelt et al., 2013). Reading comprehension incorporated four to five different texts with five to eight comprehension questions (Gehrer et al., 2013). Mathematical test included 20 items from different mathematical content areas (Neumann et al., 2013). Reliability of weighted maximum likelihood estimates (WLE; Warm, 1989) based on a partial credit model (PCM; Masters, 1982), and distribution of item difficulties in relation to students' competences are very good (Stegwallner-Schütz et al., 2022).

Professional expertise. Teachers' professional expertise was operationalised by their years of teaching practice. In addition, teachers were asked whether they have worked in inclusive classes where students with and without SEN were taught together (yes/no) and whether they felt well-prepared for teaching in an inclusive class (1 = *not well-prepared*, 2 = *rather not well-prepared*, 3 = *somewhat well-prepared*, 4 = *well-prepared*).

Teachers' self-efficacy. Teachers' self-efficacy was measured by three items from a scale by Bosse and Spörer (2014) measuring self-efficacy regarding inclusive education. The three items selected focused on students with SEN (item example: "I know that I can prepare a varied lesson topic that students with SEN can also actively participate in"). The internal consistency was acceptable, $\omega = 0.789$ (95% CI = [0.756, 0.823]).

Teacher responsibility. To assess teacher responsibility for every student, we used three items of the Teacher Responsibility Scales (TRS) by Lauermann and Karabenick (2013). The items were translated into German and adapted in terms of positive valence (item example: "I feel responsible for ensuring that each of my students manages to learn the subject matter"). The internal consistency of the subscale with the three items proved to be good, $\omega = 0.828$ (95% CI [0.795, 0.861]).

Attitudes towards inclusion. The six items measuring teachers' attitudes towards inclusion were selected in consultation with the two authors of the scale "Attitudes towards inclusion in the population" (Lüke & Grosche, 2018). An item example is: "I suppose that an inclusive school system could improve education for all children". The internal consistency of the short scale was high, $\omega = 0.901$ (95% CI [0.871, 0.931]).

5.3. Statistical analyses

To assess teachers' judgement accuracy of students' subjective well-being, we applied a correlated trait-correlated method minus one (CT-C [M-1]) model (Eid, 2000; Eid et al., 2003). This is a confirmatory factor analysis multi-trait-multi-method (CFA-MTMM) approach for structurally different (i.e., not interchangeable) methods (Eid et al., 2016). Basically, MTMM approaches are used to study convergent validity of different methods or informants, and thus suitable to assess the degree of agreement between teachers' and students' perspectives on students' noncognitive characteristics (e.g., Friedrich et al., 2013; Gomez, 2014; Schwab et al., 2020). CFA-MTMM approaches including explanatory variables allow researchers to investigate whether those external variables explain (or moderate) teachers' judgement accuracy (Koch, Kelava, & Eid, 2018). More specifically, the CT-C(M-1) model allows for the examination of the consistency of two or more methods (i.e., convergent validity) as well as of the specificity (i.e., method effect) by contrasting one (or more) method(s) with a reference method. The specificity corresponds to the part of a trait measured by a specific method that cannot be explained by the reference method (Eid et al., 2003). As we focused on students' subjective well-being, we selected students' self-reports as the reference method that was contrasted with the teacher reports as the non-reference method. In our study, *consistency* thus corresponds to the part of teacher reports that are shared with students' self-reports, whereas *specificity* represents the part of the

teacher reports that is *not* shared with the student reports.

The CT-C(M-1) model has several advantages over other CFA-MTMM models. Most importantly, the CT-C(M-1) offers the possibility of calculating variance components indicating the degree of consistency versus method specificity (Koch, Eid, & Lochner, 2018). In our study, the *consistency coefficient* refers to the proportion of observed variance in teacher reports that is common with students' self-reports (i.e., agreement), for example, that of emotional well-being, whereas the *method specificity coefficient* is equivalent to the proportion of observed variance in teacher reports that is *not* shared with the self-reports (i.e., non-agreement). The variance coefficients (consistency, method specificity and reliability) for the aggregated CT-C(M-1) model were calculated using formulas provided by Eid et al. (2003; Appendix B).

To address our second research question, we added as hypothesised, the categorical covariates students' gender, SEN, language background and the two variables measuring academic achievement to the CT-C (M-1) model. More specifically, the trait factors of the reference method (student reports) and trait-specific method factors of the non-reference method (teacher reports) were regressed on gender (male vs. female), SEN (without SEN vs. with SEN), language background (German as primary language vs. other languages) as well as reading comprehension and mathematics achievement.

Third, we related the teacher characteristics to the CT-C(M-1) model as explanatory variables (each treated using a separate model). The variables concerning professional expertise were entered as manifest covariates. The three latent variables – teacher self-efficacy, teacher responsibility and attitudes towards inclusion – were transformed so that the explanatory variables were not correlated to the latent trait factors. This transformation was performed to free the latent explanatory variables from confounding latent trait or method influences, and thus to avoid a suppression structure and related methodological problems (Koch, Kelava, & Eid, 2018). We specified the restricted CT-C(M-1) model with transformed latent explanatory variables according to the syntax provided by Koch et al. (2018b; Appendix B).

The fit of the models was assessed with the χ^2 test as well as by means of the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR). For the CFI and TLI, values greater than 0.95 are generally recommended (Hu & Bentler, 1999; Marsh et al., 2004). RMSEA values less than 0.06 refer to a close fit to the data, while SRMR values below 0.08 represent a good fit.

Parameters were calculated using the weighted least square means and variances (WLSMV) estimator, except for the restricted CT-C(M-1) models with transformed latent explanatory variables for which the robust maximum likelihood (MLR) estimator was employed. Since students were clustered in classrooms and teachers rated the students from their class, we used the design-based complex sample option to account for the clustering by adjusting standard errors accordingly. The intra-class correlation coefficients for the nine PIQ-S items ranged from 0.03 to 0.12 ($M = 0.07$) and that for the nine PIQ-T items ranged from 0.08 to 0.28 ($M = 0.16$). The reliability of the scales was evaluated by McDonald's ω (McDonald, 1999).

Data was analysed using the software Mplus Version 8.6 (Muthén & Muthén, 1998–2017). Examples of our model syntaxes are presented in the online supplementary material [<https://osf.io/pytbm/>]. The data, including all survey materials and instruments of the project, will be made available to the scientific community free of charge as of 2023 via the Data Center (www.neps-data.de/Data-Center). The data will be provided after closing a Data Use Agreement with the Leibniz Institute for Educational Trajectories (LifBi).

6. Results

6.1. Factorial validity, reliability and correlations of the PIQ-S and PIQ-T

As a preliminary analysis, we tested the factorial validity of the

reduced PIQ. Overall, the fit indices for the three-factor CFA models of the PIQ-S and the PIQ-T showed a good fit to the data (Table 1, Model 1 & 2). The reliability coefficients for the PIQ-S ranged from $\omega = 0.83$ (95% CI = [0.82, 0.84]) to $\omega = 0.91$ (95% CI = [0.90, 0.92]) and for the PIQ-T from $\omega = 0.96$ (95% CI = [0.95, 0.96]) to $\omega = 0.98$ (95% CI = [0.97, 0.98]), referring to good internal consistency of the scales (Table 2).

Table 2 shows the correlations in a simple CFA-MTMM model with a separate latent factor for each trait-method unit (TMU), allowing all TMU factors to correlate (Geiser et al., 2014; Marsh & Hocevar, 1988). This baseline CFA-MTMM model fitted well with the data (Table 1, Model 3). The monotrait-heteromethod correlation was highest for academic self-concept, followed by social inclusion, and lowest for emotional well-being. The heterotrait-monomethod correlations were higher for teacher reports than for student self-reports, indicating that teacher reports were in general less sensitive to the differences in the traits (i.e., lower discriminant validity).

76.2.2. Consistency and specificity of teacher reports with student self-reports

The CT-C(M-1) model also showed a good fit to the data (Table 1, Model 4). The loadings of the nonreference method on the trait factors of the reference method are indicative of convergent validity between the teacher reports and the student self-reports at the item level. These trait factor loadings varied between 0.31 and 0.60, with lower factor loadings for emotional well-being ($0.31 \leq \lambda \leq 0.36$) and social inclusion ($0.36 \leq \lambda \leq 0.41$) than for the academic self-concept ($0.51 \leq \lambda \leq 0.60$). The method factor loadings ranged between 0.73 and 0.92, referring to a high level of specificity on item level for teacher reports.

The estimated variance components of the aggregated CT-C(M-1) model reported in Table 3 correspond to the consistencies, (method) specificities and reliabilities of the total scales. For the aggregation, the indicators belonging to the same TMU were included (Eid et al., 2003). The consistency coefficients show that only 12% in emotional well-being, 18% in social inclusion and 33% in academic self-concept of the variance in teacher reports were shared with the student self-reports. Thus, the consistencies were rather low for emotional well-being and social inclusion and moderate for academic self-concept. Since previous studies on teachers' judgment accuracy of different aspects of subjective well-being mostly reported correlations (Urhahne & Wijnia, 2021) rather than the consistency coefficients, we also calculated the latent correlations (i.e., correlations corrected for measurement errors) by extracting the square root of the consistency coefficients (Table 3, last column). The latent correlation between teacher reports and student self-reports was .35 for emotional well-being, 0.42 for social inclusion, and 0.57 for the academic self-concept. Additional multilevel analyses revealed similar results for relative judgement accuracy (i.e., rank components; results are presented in the online supplementary

Table 1 Summary of goodness-of-fit statistics of the models.

Model	χ^2_{wlsmv}	df	p	CFI	TLI	RMSEA [90% CI]	SRMR
(1) CFA PIQ-S	369.94	24	<.001	.990	.985	.064 [.058, .069]	.027
(2) CFA PIQ-T	638.25	24	<.001	.994	.991	.097 [.091, .104]	.032
(3) CFA MTMM	1083.04	120	<.001	.992	.990	.046 [.044, .049]	.031
(4) CT-C (M-1)	1058.27	114	<.001	.993	.990	.047 [.044, .049]	.030

Note. CT-C(M-1) = correlated trait-correlated method minus one; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval; SRMR = Standardized Root Mean Square Residual.

Table 2 Descriptive Statistics, Reliabilities (McDonald's ω) of the PIQ-S^a and PIQ-T^a and Correlations in the CFA MTMM Model.

	Student self-report			Teacher report		
	EMO	SOC	ASC	EMO	SOC	ASC
<i>Student</i>						
EMO	(.91)					
SOC	.47***	(.91)				
ASC	.51***	.39***	(.83)			
<i>Teacher</i>						
EMO	.35***	.25***	.37***	(.98)		
SOC	.24***	.42***	.25***	.70***	(.96)	
ASC	.20***	.12***	.58***	.57***	.48***	(.97)
M	8.56	9.88	8.62	9.20	9.04	7.84
SD	2.50	1.98	2.08	1.92	1.99	2.64

Note. EMO = emotional well-being, SOC = social inclusion, ASC = academic self-concept.

Main diagonal (values in parentheses): reliabilities (McDonald ω). Monotrait-heterotrait correlations are marked in grey. *** $p < .001$.

^a Both PIQ versions without negatively worded items (4, 8, 12).

material).

In Table 4, the correlations of the trait factors (student self-reports) and trait-specific method factors (teacher reports) in the CT-C(M-1) model are reported. The significant moderate to high correlations between the three trait-specific method factors ($0.51 \leq r \leq 0.69$) indicate that the method effects can be generalised to some extent across the traits. That is, teachers tend to over- or under-estimate, for instance, students' emotional well-being and social inclusion in a similar manner. Five out of the six correlations between the trait factors and the trait-specific method factors were significant. For instance, the significant positive correlation between the trait factor for academic self-concept and the teacher reports of students' emotional well-being ($r = 0.21$) shows that teachers overestimate the emotional well-being of students with a higher academic self-concept, and vice versa, they underestimate the emotional well-being of students with a lower academic self-concept. The small negative correlation between the method factor for academic self-concept and the trait factor for social inclusion ($r = -0.13$), in turn, indicate that teachers tended to underestimate the academic self-concept of students who reported to be more socially included, and vice versa.

6.3. Student characteristics as explanatory variables

Focusing in Table 5 on the significant effects of student characteristics on the trait factors of student self-reports, girls reported higher emotional well-being and lower academic self-concept than boys. The status SEN had a small negative effect on the self-reports of social inclusion and the academic self-concept but not on emotional well-being, whereas German as primary language had a small negative effect on the students' ratings of their emotional well-being. Reading comprehension had a small positive effect on students' ratings of both emotional well-being and academic self-concept, being larger for the latter, while mathematics achievement had a positive effect only on the self-reported academic self-concept.

The standardised regression coefficients concerning the trait-specific method factors (i.e., specificity of teacher reports) were all significant with two exceptions: German as primary language had a significant small positive effect on only the method factor for students' academic self-concept and both gender and the status SEN could predict the method effects of teacher reports. Teachers rated all three traits systematically higher for girls than for boys compared to their self-reports. In turn, the status SEN was negatively associated with method effects of teacher reports. That is, compared to students without SEN, teachers rated the subjective well-being of students with SEN systematically lower than the students themselves. The effect was most pronounced for

Table 3
Estimated variance components in the aggregated CT-C(M-1) model.

Rating	Observed variables			True-score variables		
	Reliability	Consistency	Method specificity	Consistency	Method specificity	Latent correlation ^a
<i>Emotional well-being</i>						
Self	.91	.91		1.00		
Teacher	.96	.12	.84	.12	.88	.35
<i>Social inclusion</i>						
Self	.89	.89		1.00		
Teacher	.94	.17	.77	.18	.82	.42
<i>Academic self-concept</i>						
Self	.83	.83		1.00		
Teacher	.97	.32	.65	.33	.67	.57

Note.

^a Latent correlation with the reference method ($\sqrt{\text{consistency}}$).

Table 4
Correlations of the trait and trait-specific method factors in the CT-C(M-1) model.

	Trait factors (Student self-report)			Trait-specific method factors (Teacher report)		
	EMO-S	SOC-S	ASC-S	EMO-T	SOC-T	ASC-T
<i>Trait factors</i>						
EMO-S	1.00					
SOC-S	.47***	1.00				
ASC-S	.51***	.39***	1.00			
<i>Trait-specific method factors</i>						
EMO-T		.09***	.21***	1.00		
SOC-T	.04		.10***	.69***	1.00	
ASC-T	-.11***	-.13***		.51***	.51***	1.00

Note. EMO = emotional well-being, SOC = social inclusion, ASC = academic self-concept; S = self-report, T = teacher report. Empty cells indicate non-admissible correlations that were fixed to 0. ****p* < .001.

the academic self-concept. Both measures of academic achievement could predict the method effects of teacher reports, indicating that the higher a student’s reading comprehension and mathematics achievement, the higher were the teacher’s ratings – compared to the student’s self-reports – of all three aspects of subjective well-being. Again, these method effects were most pronounced for the students’ academic self-concept.

6.4. Teacher characteristics as explanatory variables

As shown in Table 6, there was no significant effect of teacher characteristics on all trait factors of student self-reports. None of the three manifest variables representing professional expertise had an effect on trait-specific method factors (i.e., specificity of teacher reports), except for one, as the number of years of teaching practice was negatively associated with the method factors for students’ academic self-

Table 5
Standardized regression coefficients and standard errors (in parentheses) of the student characteristics as covariates on trait and method factors in the CT-C(M-1) model.

	Trait factors (Student self-report)			Trait-specific method factors (Teacher report)		
	EMO	SOC	ASC	EMO	SOC	ASC
Gender (female)	.129** (.019)	.012 (.019)	-.131*** (.023)	.121*** (.023)	.101*** (.022)	.142*** (.022)
SEN (with)	-.018 (.019)	-.090*** (.019)	-.114*** (.020)	-.079*** (.020)	-.156*** (.020)	-.289*** (.021)
Primary language (German)	-.050** (.020)	-.003 (.021)	.029 (.023)	.029 (.023)	.042 (.023)	.102*** (.023)
Reading comprehension	.086*** (.020)	.007 (.020)	.274*** (.020)	.158*** (.025)	.178*** (.024)	.410*** (.020)
Mathematics achievement	.021 (.020)	.008 (.020)	.328*** (.018)	.166*** (.025)	.142*** (.024)	.415*** (.020)

Note. EMO = emotional well-being, SOC = social inclusion, ASC = academic self-concept; SEN = special educational needs. Coding: Gender 0 = male, 1 = female; SEN 0 = without, 1 = with; Primary language 0 = other languages (than German), 1 = German (teaching language).

p* < .05. *p* < .01. ****p* < .001.

concept. That is, with increasing teaching experience, the teachers tended to underestimate the students’ academic self-concept. The three latent explanatory variables, in turn, were all positively related to the trait-specific method factors of emotional well-being and social inclusion. The small positive effect of teacher self-efficacy indicate that the more a teacher believed in themselves or their abilities to teach, in particular to the context of inclusive education, the more positively they rated a student’s emotional well-being and social inclusion compared to the self-reports. No effects were found for method effects regarding the academic self-concept. A similar pattern emerged with regard to teachers’ responsibility for every student as well as teachers’ attitudes towards inclusion. However, the effect sizes were all rather small.

7. Discussion

The present study investigated teachers’ judgment accuracy of students’ subjective well-being in school, focusing on the three components: emotional well-being in school, social inclusion in class and the academic self-concept. We drew from the self-reports of 2592 grade six students as well as the corresponding ratings of 432 homeroom teachers. To assess teachers’ judgment accuracy of students’ subjective well-being, we evaluated the consistency between self-reports and teacher reports (i.e., agreement) as well as the specificity of teacher reports (i.e., non-agreement) by applying a CT-C(M-1) model. Second, we examined whether specific student and teacher characteristics could explain the specificity in teacher reports. To this end, we included student and teacher characteristics as manifest or latent explanatory variables to the CT-C(M-1) model.

Results showed that the teachers’ judgment accuracy of the students’ subjective well-being was low to moderate. Only 12% of the variance in teacher reports of students’ emotional well-being was shared with self-reports, 18% was shared in terms of social inclusion and 33% in terms of the academic self-concept. Hence, for all three aspects of subjective well-being, the consistency between teacher reports and students’ self-

Table 6

Standardized regression coefficients and standard errors (in parentheses) of the teacher characteristics as explanatory variables on trait and method factors in the classical and restricted CT-C(M-1) model.

	Trait factors (Self-report)			Trait-specific method factors (Teacher report)		
	EMO	SOC	ASC	EMO	SOC	ASC
<i>Classical CT-C(M-1) Model with Manifest Covariates</i>						
Years of teaching practice	-.028 (.031)	-.007 (.028)	-.021 (.025)	-.035 (.041)	-.045 (.036)	-.054* (.027)
Professional experience in inclusive education (yes)	.034 (.026)	.009 (.025)	.032 (.021)	.012 (.039)	.027 (.032)	-.002 (.027)
Feeling prepared for inclusive education	.006 (.027)	-.003 (.024)	-.001 (.024)	.062 (.039)	.014 (.034)	.006 (.029)
<i>Restricted CT-C(M-1) Model with Transformed Latent Explanatory Variables</i>						
Self-efficacy	.011 (.029)	-.005 (.035)	-.041 (.029)	.139** (.041)	.069* (.031)	.026 (.030)
Responsibility	-.031 (.032)	-.046 (.033)	-.003 (.028)	.092* (.042)	.069* (.034)	.042 (.032)
Attitudes towards inclusion	.032 (.029)	.001 (.030)	.051 (.027)	.130*** (.036)	.075** (.030)	.030 (.027)

Note. EMO = emotional well-being, SOC = social inclusion, ASC = academic self-concept. Coding: Professional experience in inclusive education 0 = no experience, 1 = experience. * $p < .05$. ** $p < .01$. *** $p < .001$.

reports was lower than the specificity of teacher reports. In search of explanatory factors for the specificity in teacher reports (i.e., non-agreement with student reports), we accounted for student and teacher characteristics that are considered as relevant for adaptive teaching, especially in the context of inclusive education. Students' gender, their status SEN as well as reading comprehension and mathematics achievement were all associated with the specificity in teacher reports for all three aspects of subjective well-being. The effects were most pronounced for the academic self-concept. Similarly, German as a primary language had only a significant small positive effect on the specificity in teacher reports for students' academic self-concept. Teaching experience, as the number of years of teaching practice, was negatively associated with the specificity in teacher reports for students' academic self-concept. In turn, teachers' self-efficacy, their attitudes towards inclusion as well as their responsibility for every student were positively related to the specificity in teacher reports of students' emotional well-being and social inclusion but not to the academic self-concept. However, the effect sizes for the teacher characteristics taken into account as explanatory variables were all small.

7.1. Specificity in teacher reports of students' subjective well-being

As expected, teachers' judgement accuracy of students' subjective well-being was moderate or even low, in particular for the emotional component. Only slightly more than a tenth of the variance in teacher reports of students' emotional well-being was shared with self-reports, pointing to the fact that it is difficult for teachers to gauge students' emotional well-being. Emotions or feelings correspond to a person's internal states, to which solely they have immediate access to (Eid, 2018; Schwarz & Strack, 1999), thus leaving others to decode emotional well-being from external clues, such as behavior and verbal or facial expressions.

When it comes to social inclusion, the reference method may be less clear. As we focused on *subjective* well-being, *self-perceived* social inclusion was assessed in our study. Compared to the study by Urhahne and Zhu (2015), the correlation between students' self-reports and the

teacher reports of the social aspect of subjective well-being was higher. It should be noted that their findings were based on (transformed) within-class correlations and multilevel analyses (i.e., relative judgement accuracy) for social problems, while we calculated latent correlations of social inclusion based on the consistency components of the CT-C(M-1) model by accounting for the clustering of the data. Our additional multilevel analyses showed a similar degree for relative judgement accuracy (cf. supplementary material). Although our analytical approach was similar to that adopted by Gomez (2014), the latent correlations found in his study were considerably higher. Again, social problems (i.e., peer problems) were assessed, thus the constructs investigated in both studies were not the same.

Our results on teachers' judgement accuracy of students' academic self-concept are in line with previous findings. As hypothesised, moderate judgement accuracy was found (e.g., Givvin et al., 2001; Pielmeier et al., 2018; Praetorius et al., 2013; Urhahne et al., 2011). Indeed, the students' academic self-concept seems to be a relatively "good trait" (Funder, 1995; 2012), as it is easier for the teachers to detect than the two other components of students' subjective well-being. In contrast to emotional well-being, both judgements (i.e., self-report and teacher report) of a student's academic self-concept might be obtained on a relatively common ground, because the academic self-concept is influenced especially by evaluations, such as grading by teachers or achievement tests. This assumption is underscored by the finding that the teachers' assessment accuracy of the students' academic self-concept was higher when they were informed about the students' achievements (Helm et al., 2018).

Overall, our findings show that the specificity of teacher reports is higher than the consistency between the teacher reports and self-reports for all three aspects of students' subjective well-being, pointing to considerable method effects (i.e., non-agreement). Furthermore, the method effects can be generalised to some extent across the three traits, meaning that teachers tend to over- or under-estimate the three aspects of subjective well-being in a similar manner. The method effects can also be considered as method bias (Koch, Eid, & Lochner, 2018). It should be emphasised that we do not classify teacher reports as 'wrong' and their specificity as a judgmental error to be omitted, when applying the terms method effects or method bias. Rather, we deem method effects to be a vital source of information. If method effects are detected, it is relevant then to understand why teacher reports differ from self-reports about the students' subjective well-being and what variables can explain these differences.

7.2. Gender, SEN and achievement as moderators of teachers' judgement accuracy

As expected, the students' gender and the status SEN could predict teachers' judgement accuracy of students' subjective well-being in school. Teachers rated (in relation to self-reports) all three components of students' subjective well-being systematically higher for girls than for boys. For students with SEN, teachers estimated the subjective well-being (again in relation to self-reports) as systematically lower than those compared to students without SEN. This effect goes beyond the generally lower self-reports of school-related subjective well-being for students with SEN compared to those without SEN (cf. also Goldan et al., 2022). Thus, our results point to a positive bias in teacher reports for girls and a negative bias for the status SEN. The bias that distorts teachers' judgements in favour of girls can be seen as a leniency effect (e.g., Phelps et al., 1986), which may be related to the fact that girls are often perceived as more agreeable, conscientious and socially-oriented than boys (Fabes & Eisenberg, 1998). Another explanation could be that girls tend to express their positive feelings more openly than boys and hide their negative feelings (Chaplin & Aldao, 2013). Nevertheless, the overestimation of girls' subjective well-being in school can also have negative consequences (as the underestimation of boys' subjective well-being) in that teachers may overlook or underestimate those

moments when girls do not feel comfortable at school.

The negative bias related to the status SEN can be interpreted as a halo effect. This widely investigated phenomenon – which was first proposed and described in the pioneering work by Thorndike (1920) as a “constant error toward suffusing ratings of special features with a halo belonging to the individual as a whole” (p. 25) – has been repeatedly observed or suggested in research related to teachers’ judgment accuracy (e.g., Kaiser et al., 2013; Urhahne & Wijnia, 2021) or teacher attunement (Marucci et al., 2020), where it has mainly retained a positive connotation. Its negative counterpart, the negative or reversed halo effect – also called horn effect (e.g., Forgas & Laham, 2016) – seems to better align with the marked tendency of teachers to systematically underestimate the subjective well-being of students with the status SEN. One could argue that the term ‘stigma’ might be more appropriate in this regard (Jones, 1972; Katchergin, 2012). The negative effect of the status SEN was most pronounced for the academic self-concept. This finding is particularly noticeable since studies have repeatedly shown that the academic self-concept of students with SEN is lower in inclusive education than in special schools (e.g., Kocaj et al., 2018; Oh-Young & Filler, 2015; Tracey et al., 2003). As a large percentage of students with SEN demonstrate low achievement or learning difficulties, this negative effect could be associated with academic achievement, meaning that teachers, to some extent, relied their judgments of academic self-concept on the students’ academic achievement. Indeed, our findings indicate that students’ academic achievement seems to be an important moderator of teachers’ judgment accuracy of students’ subjective well-being. As expected, this effect was highest for the academic self-concept (Zhu & Urhahne, 2021). But also teacher ratings of emotional well-being and social inclusion were affected by both reading comprehension and mathematics achievement. It could be concluded that teachers color their judgements of students’ well-being to some extent by students’ level of academic achievement.

Having said that, with regard to the underestimation of the social inclusion of students with SEN, it may be that the teacher ratings are more attuned to the peers’ view than to the subjective view by the students themselves. Studies have repeatedly shown that students with SEN are generally less accepted than their peers, whereas the SEN students’ self-perception of their social inclusion is less negative or even positive (e.g., Avramidis et al., 2018; Garrote et al., 2017).

Furthermore, our findings indicate that the teachers judged the academic self-concept of students with German as a primary language higher than those with another primary language. In contrast to the other student characteristics, there was no method effect for emotional well-being and social inclusion. It could be concluded that students’ academic self-concept not only seems to be easier to rate for teachers but is also more affected by judgmental bias.

7.3. Teacher characteristics (not) explaining teachers’ judgement accuracy

Research on teacher characteristics that explains the judgement accuracy of students’ subjective well-being is scarce. As students’ subjective well-being is an important indication for teachers to adapt their communication, teaching material or instructional strategies according to the students’ individual differences, we focused on characteristics that are meaningful for adaptive teaching, especially in the context of inclusive education. In terms of the variables related to teachers’ professional expertise, neither professional experience in inclusive education nor feeling prepared for inclusive education could explain the method effects for teacher reports of students’ subjective well-being. Only the teachers’ years of experience showed an effect on their ratings, and only on those for students’ academic self-concept, although this was contrary to what was expected. However, the small effect size was in line with previous results (Urhahne & Wijnia, 2021), indicating that more years of teaching experience does not increase by itself the teachers’ judgment accuracy or diagnostic competences (Praetorius &

Südkamp, 2019).

In contrast to professional experience, the three other teacher characteristics did not affect the teachers’ ratings of students’ academic self-concept but those of emotional well-being and social inclusion positively. Given that the effect sizes were rather small, the findings should be interpreted cautiously. Having said that, a noticeable finding is that the effects of teachers’ self-efficacy, responsibility for every student and attitudes towards inclusion were slightly more pronounced for emotional well-being than for social inclusion. This could be a hint for a leniency tendency in teachers’ ratings, which could be explained by the emotional connotation of feeling self-efficacious and having positive attitudes towards inclusion. The same might apply to teachers’ responsibility towards every student, although even weaker effects were noted. In this regard, the effects might have been higher if the items considered to assess the teachers’ responsibility had focused more on responsibility in terms of peer relations in class (Harks & Hannover, 2017).

Overall, teacher characteristics could only explain to a small extent the specificity in teacher reports of students’ subjective well-being. This calls for other explanatory variables and further investigations. Since the range of teachers’ characteristics possibly relevant to explain the judgement accuracy is supposed to be rather broad, it might be worthwhile to examine whether a teacher’s judgment is based on “good information” – to borrow Funder’s terminology once more. According to Funder (2012), information can be “good” in terms of quantity or quality. Knowing more about the kind and the extent of information a teacher needs to accurately assess a student’s subjective well-being could likely to have an educational impact.

7.4. Improving teachers’ judgement accuracy for subjective well-being

As already noted by Shavelson and Stern (1981), one overarching goal of research on teacher judgements is to improve the teaching practice by enhancing the judgement accuracy. Our study aimed to contribute to this common long-standing goal by investigating factors that could explain teachers’ judgement accuracy of students’ subjective well-being in school.

In line with previous studies, our results indicate that teachers seem to be susceptible to a negative bias related to SEN and to a (small) positive bias in favour of girls compared to boys. A slight tendency for leniency was also shown in relation to the teachers’ self-efficacy and attitudes towards inclusion. While a slight overestimation might be to the benefit of the students, and even teachers themselves (Wang & Williamson, 2020), a strong overestimation and a (strong) underestimation are both possibly problematic. Therefore, interventions should focus on expanding teachers’ knowledge of judgmental processes, including information about possible bias when rating students’ well-being. A crucial first step in the promotion of teacher judgement accuracy is to raise the teachers’ awareness of the significance of students’ subjective well-being for their individual development and adaptive teaching. If a teacher does not acknowledge a student’s subjective well-being as relevant, trait-specific behaviours or other cues might not be detected by the teacher, even if that information is available.

Another issue in the promotion of teacher judgement accuracy of students’ subjective well-being is obtaining valid information about the actual subjective well-being of a student. This is way more difficult than in terms of judgement accuracy of observable constructs, such as academic achievement, for which achievement tests can be used regularly by teachers to acquire valid information about the students’ academic achievement (Südkamp et al., 2012). Classroom-implemented data collection procedures, such as observation or behaviour rating, fail to gain insights into the internal states of a student. This particularly applies to their emotional well-being. Since the ‘referee’ for a person’s actual subjective well-being is the person themselves, it is crucial to get corresponding feedback from students on a regular basis. In addition to

individual interviews with students, appropriate screening instruments could be used by teachers to detect low levels of school-related well-being or early signs of emotional or social problems (Splett et al., 2020). More importantly, the data obtained by self-reported screening instruments allows teachers to gain information about a student's perception of their subjective well-being. Through a comparison with their own view, teachers may be able to enhance their judgement accuracy.

Thus, multi-informant assessment plays a crucial role in addressing the judgement accuracy of another person's subjective well-being. Assessing the student's own view decreases the risk of making (erroneous) assumptions about the student's subjective well-being. As already mentioned, we do not claim that teacher reports are 'wrong'. However, teachers should be aware that discrepancies between their reports and the students' self-reports are very likely while judging the students' subjective well-being. Given the complexity of a judgement process, and the various stages that must be traversed successfully before an accurate judgement can be reached, we argue – in line with Funder's appeal – that instead of being astonished that the judgment accuracy of teacher reports of students' subjective well-being is relatively low, "it might be wiser to be amazed that human judgement of [subjective well-being] is ever correct" (Funder, 2012, p. 179).

7.5. Limitations and directions for future research

First, it is worth noting that we concentrated on three specific components of students' subjective well-being in school. Further studies could include all six components proposed in the definition by Hascher (2010; 2012). Furthermore, the three components of subjective well-being were assessed only with three items each, covering only certain facets of each construct (Schwab et al., 2020). Emotional well-being, for instance, primarily relates to a positive feeling towards school, corresponding to a rather narrow understanding of the construct.

Second, our findings are based on a CT-C(M-1) model (Eid, 2000; Eid et al., 2003). While this analytical approach can be considered as a strength of the study, our findings might only be somewhat comparable with those from other studies on teachers' judgement accuracy of the different aspects of subjective well-being. To address this issue, we also reported the latent correlations (based on the consistency) to provide some comparison with relative judgement accuracy, which operationalised as rank component is the most often used indicator of teacher judgement accuracy for students' emotions and academic self-concept (Urhahne & Wijnia, 2021) as well as for social aspects (e.g., Südkamp et al., 2018). Furthermore, we enclosed results of multilevel analyses with the rank component as index of relative judgement accuracy (cf. supplementary material). In this context, it should be noted that we intentionally did not use multilevel modelling so as to not further increase the complexity of the CT-C(M-1) models with transformed latent explanatory variables. Nevertheless, we applied the complex sample option as a design-based approach, which accounts for the clustering of the data by adjusting standard errors accordingly.

Third, approximately 30% of the homeroom teachers invited to participate in the study did not respond. Although this might seem a high percentage of missing values, this can be classified as an average response rate in teacher surveys. Furthermore, the teacher sample was not balanced in terms of gender. However, the relatively large percentage of female teachers in the study is reflective of their overrepresentation in secondary education in Germany.

Fourth, it needs to be stressed that we conducted secondary data analyses. The study, of which we draw the data, was not explicitly designed to investigate teachers' judgement accuracy. Although the study offers rich data, including several relevant variables, there are further explanatory variables worth investigating related to teachers' judgment accuracy of students' subjective well-being in school. Future studies on teachers' judgement accuracy might benefit by, for instance, accounting for judgment confidence (Praetorius et al., 2013), the quality

of the teacher-student relationship (Verschuere & Koomen, 2012) or the teachers' personality (for an example related to job performance: Cheng et al., 2017).

Finally, our study was cross-sectional. To investigate the predictive value of teachers' judgement accuracy of students' subjective well-being, a longitudinal design would be required. According to Funder (2012), behavioural prediction presumably corresponds to the "gold standard" among the criteria capturing the accuracy of judgements. Based on our theoretical considerations, a behavioral prediction of the teachers' judgement accuracy of the students' subjective well-being would be appropriate adaptive teaching, which in turn promotes students' individual learning processes and their social-emotional development. However, studies on such behavioural predictions of judgment accuracy are difficult to conduct and therefore need thorough planning.

Having said that, our findings offer some new insights in terms of teachers' judgement accuracy of students' subjective well-being in school. In this vein, we hope that our study might inspire future investigations in search of explanatory factors of judgment accuracy of subjective well-being.

Funding

The project was supported by the German Federal Ministry of Education and Research (BMBF) [grant numbers IN1503A, IN1503B, IN1503C, IN1503D].

Contributors

CZ, LN and MS designed and agreed on the aim of this study. CZ conducted the data analyses and wrote the manuscript. LN contributed to parts of the Introduction and the Methods section. MK contributed to parts of the Introduction. MS was responsible for the data collection and data preparation. All authors read and agreed to the final version of the submission.

Declaration of competing interest

We have no competing interest to disclose.

Data availability

Data will be made available on request.

Acknowledgements

This article is based on data from the project Inclusive Education in Lower Secondary Schools in Germany (INSIDE). The project was initiated and coordinated as cooperative project of the Leibniz Institute for Educational Trajectories (LifBi), the Institute for Educational Quality Improvement (IQB), the University of Potsdam, and the University of Wuppertal (Germany). We thank the LifBi for providing us data as well as additional data following a reviewer's request and the reviewers for their valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.tate.2023.104304>.

References

- Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731–746. <https://doi.org/10.1037/0022-0663.91.4.731>
- Amholt, T. T., Dammeyer, J., Carter, R., & Niclasen, J. (2020). Psychological well-being and academic achievement among school-aged children: A systematic review. *Child Indicators Research, 13*, 1523–1548. <https://doi.org/10.1007/s12187-020-09725-9>

- Artelt, C., & Gräsel, C. (2009). Diagnostische kompetenz von Lehrkräften [diagnostic competence of teachers]. *Zeitschrift für Pädagogische Psychologie*, 23(3), 157–160. <https://doi.org/10.1024/1010-0652.23.34.157>
- Artelt, C., Weinert, S., & Carstensen, C. H. (2013). Assessing competencies across the lifespan within the German National Educational Panel study (NEPS). Editorial. *Journal for Educational Research Online*, 5(2), 5–14. <https://doi.org/10.25656/01:8422>
- Avramidis, E., Avgeri, G., & Strogilos, V. (2018). Social participation and friendship quality of students with special educational needs in regular Greek primary schools. *European Journal of Special Needs Education*, 33(2), 221–234. <https://doi.org/10.1080/08856257.2018.1424779>
- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education*, 17(2), 129–147. <https://doi.org/10.1080/08856250210129056>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle kompetenz von Lehrkräften [keyword: Professional competence of teachers]. *Zeitschrift für Erziehungswissenschaft*, 9, 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43. <https://doi.org/10.1037/1045-3830.23.1.43>
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, 40(1), 23–38. <https://doi.org/10.1080/02796015.2011.12087726>
- de Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education*, 15(3), 331–353. <https://doi.org/10.1080/13603110903030089>
- Bosse, S., & Spörer, N. (2014). Erfassung der Einstellung und der Selbstwirksamkeit zum inklusiven Unterricht [Assessment of attitudes and self-efficacy of pre-service teachers towards inclusive education]. *Empirische Sonderpädagogik*, 4, 279–299.
- Carter, S., & Andersen, C. (2019). *Wellbeing in educational contexts*. University of Southern Queensland. <https://usq.pressbooks.pub/wellbeingineducationalcontexts/>
- Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, 139(4), 735–765. <https://doi.org/10.1037/a0030737>
- Cheng, K. H. C., Hui, C. H., & Cascio, W. F. (2017). Leniency bias in performance ratings: The Big-Five correlates. *Frontiers in Psychology*, 8, 521. <https://doi.org/10.3389/fpsyg.2017.00521>
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 255–296). Macmillan.
- DeVries, J. M., Voß, S., & Gebhardt, M. (2018). Do learners with special education needs really feel included? Evidence from the perception inclusion questionnaire and strengths and difficulties questionnaire. *Research in Developmental Disabilities*, 83, 28–36. <https://doi.org/10.1016/j.ridd.2018.07.007>
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1), 34–43.
- Diener, E., Lucas, R. E., & Oishi, S. (2005). Subjective well-being: The science of happiness and life satisfaction. In C. R. Snyder, & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 63–73). Oxford University Press.
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65(2), 241–261. <https://doi.org/10.1007/BF02294377>
- Eid, M. (2018). Statistical approaches to analyzing well-being data. In E. Diener, S. Oishi, & L. Tay (Eds.), *Handbook of well-being*. DEF Publishers. <https://www.nobascholar.com/chapters/16>
- Eid, M., Geiser, C., & Koch, T. (2016). Measuring method effects: From traditional to design-oriented approaches. *Current Directions in Psychological Science*, 25(4), 275–280. <https://doi.org/10.1177/0963721416649624>
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, 8(1), 38–60. <https://doi.org/10.1037/1082-989X.8.1.38>
- Fabes, R. A., & Eisenberg, N. (1998). Meta-analyses of age and sex differences in children's and adolescents' prosocial behavior. *Handbook of Child Psychology*, 3, 1–29.
- Fairbanks, C. M., Duffy, G. G., Faircloth, B. S., He, Y., Levin, B. B., Rohr, J., & Stein, C. (2010). Beyond knowledge: Exploring why some teachers are more thoughtfully adaptive than others. *Journal of Teacher Education*, 61, 161–171. <https://doi.org/10.1177/0022487109347874>
- Forgas, J. P., & Laham, S. M. (2016). Halo effects. In R. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in judgement, thinking and memory* (pp. 276–290). Routledge.
- Francis, B., Archer, L., Hodgen, J., Pepper, D., Taylor, B., & Travers, M.-C. (2017). Exploring the relative lack of impact of research on 'ability grouping' in England: A discourse analytic account. *Cambridge Journal of Education*, 47(1), 1–17. <https://doi.org/10.1080/0305764X.2015.1093095>
- Friedrich, A., Jonkmann, K., Nagengast, B., Schmitz, B., & Trautwein, U. (2013). Teachers' and students' perceptions of self-regulated learning and math competence: Differentiation and agreement. *Learning and Individual Differences*, 27, 26–34. <https://doi.org/10.1016/j.lindif.2013.06.005>
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102(4), 652–670. <https://doi.org/10.1037/0033-295X.102.4.652>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Garrote, A., Sermier Dessementet, R., & Moser Opitz, E. (2017). Facilitating the social participation of pupils with special educational needs in mainstream schools: A review of school-based interventions. *Educational Research Review*, 20, 12–23. <https://doi.org/10.1016/j.edurev.2016.11.001>
- Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2), 50–79. <https://doi.org/10.25656/01:8424>
- Geiser, C., Bums, G. L., & Servera, M. (2014). Testing for measurement invariance and latent mean differences across methods: Interesting incremental information from multitrait-multimethod studies. *Frontiers in Psychology*, 5, 1216. <https://doi.org/10.3389/fpsyg.2014.01216>
- Gentrup, S., Rjosk, C., Stanat, P., & Lorenz, G. (2018). Einschätzungen der schulischen Motivation und des Arbeitsverhaltens durch Grundschullehrkräfte und deren Bedeutung für Verzerrungen in Leistungserwartungen [Teachers' perceptions of students' motivation and learning behaviour and their role in biased teacher achievement expectations]. *Zeitschrift für Erziehungswissenschaft*, 21(4), 867–891. <https://doi.org/10.1007/s11618-018-0806-2>
- Ghaith, G., & Yaghi, H. (1997). Relationships among experience, teacher efficacy, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, 13(4), 451–458. [https://doi.org/10.1016/S0742-051X\(96\)00045-5](https://doi.org/10.1016/S0742-051X(96)00045-5)
- Gilmour, A. F., & Wehby, J. H. (2019). The association between teaching students with disabilities and teacher turnover. *Journal of Educational Psychology*, 112(5), 1042–1060. <https://doi.org/10.1037/edu0000394>
- Givvin, K. B., Stipek, D. J., Salmon, J. M., & MacGyvers, V. L. (2001). In the eyes of the beholder: Students' and teachers' judgments of students' motivation. *Teaching and Teacher Education*, 17(3), 321–331. [https://doi.org/10.1016/S0742-051X\(00\)00060-3](https://doi.org/10.1016/S0742-051X(00)00060-3)
- Goldan, J., Nusser, L., & Gebel, M. (2022). School-related subjective well-being of children with and without special educational needs in inclusive classrooms. *Child Indicators Research*. <https://doi.org/10.1007/s12187-022-09914-8>
- Gomez, R. (2014). Correlated trait-correlated method minus one analysis of the convergent and discriminant validities of the Strengths and Difficulties Questionnaire. *Assessment*, 21(3), 372–382. <https://doi.org/10.1177/107319112457588>
- Hachfeld, A., Anders, Y., Schroeder, S., Stanat, P., & Kunter, M. (2010). Does immigration background matter? How teachers' predictions of students' performance relate to student background. *International Journal of Educational Research*, 49(2–3), 78–91. <https://doi.org/10.1016/j.ijer.2010.09.002>
- Hardy, I., Decristan, J., & Klieme, E. (2019). Adaptive teaching in research on learning and instruction. *Journal for Educational Research Online*, 11(2), 169–191. <https://doi.org/10.25656/01:18004>
- Harks, M., & Hannover, B. (2017). Sympathiebeziehungen unter peers im klassenzimmer: Wie gut wissen lehrkräfte bescheid? [Sympathy-based peer interactions in the classroom: How well do teachers know them?]. *Zeitschrift für Erziehungswissenschaft*, 20(3), 425–448. <https://doi.org/10.1007/s11618-017-0769-8>
- Hascher, T. (2010). Wellbeing. In S. Jarvela (Ed.), *Social and emotional aspects of learning* (pp. 99–105). Elsevier.
- Hascher, T. (2012). Well-being and learning in school. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_1832
- Hellmich, F., Loeper, M. F., & Görel, G. (2019). The role of primary school teachers' attitudes and self-efficacy beliefs for everyday practices in inclusive classrooms – a study on the verification of the 'Theory of Planned Behaviour'. *Journal of Research in Special Educational Needs*, 19(1), 36–48. <https://doi.org/10.1111/1471-3802.12476>
- Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education*, 3(2), 91–98. [https://doi.org/10.1016/0742-051X\(87\)90010-2](https://doi.org/10.1016/0742-051X(87)90010-2)
- Helm, F., Müller-Kalthoff, H., Mukowski, R., & Möller, J. (2018). Teacher judgment accuracy regarding students' self-concepts: Affected by social and dimensional comparisons? *Learning and Instruction*, 55, 1–12. <https://doi.org/10.1016/j.learninstruc.2018.02.002>
- Herppich, S., Praetorius, A.-K., Foerster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Boehmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Boehmer, I., Karing, C., Kaiser, J., & Sudkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313. <https://doi.org/10.3102/00346543059003297>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Impara, J. C., & Plake, B. S. (1998). Teachers' ability to estimate item difficulty: A test of the assumptions in the angoff standard setting method. *Journal of Educational Measurement*, 35(1), 69–81. <https://doi.org/10.1111/j.1745-3984.1998.tb00528.x>
- Jansen, T., Vögelin, C., Machts, N., Keller, S., Köller, O., & Möller, J. (2021). Judgment accuracy in experienced versus student teachers: Assessing essays in English as a foreign language. *Teaching and Teacher Education*, 97, Article 103216. <https://doi.org/10.1016/j.tate.2020.103216>
- Jones, R. L. (1972). Labels and stigma in special education. *Exceptional Children*, 28(7), 553–564. <https://doi.org/10.1177/001440297203800705>
- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling

- prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 281–388). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60240-3](https://doi.org/10.1016/S0065-2601(08)60240-3).
- Kaiser, J., Retelsdorf, J., Südkamp, A., & Möller, J. (2013). Achievement and engagement: How student characteristics influence teacher judgments. *Learning and Instruction*, 28, 73–84. <https://doi.org/10.1016/j.learninstruc.2013.06.001>
- Katchergin, O. (2012). Between negative stigma (cultural deprivation) and positive stigma (learning disability): The historical development of two special education tracks. *Culture, Medicine, and Psychiatry*, 36. <https://doi.org/10.1007/s11013-012-9281-z>, 679–671.
- Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self and other perception: A meta-analysis. *Personality and Social Psychology Review*, 14(2), 196–213. <https://doi.org/10.1177/1088868309353414>
- Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001>
- Knickenberg, M., Zurbruggen, C. L. A., Gebhardt, M., & Schwab, S. (2020). Assessing dimensions of inclusion from students' perspective: Measurement invariance across students with learning disabilities in different educational settings. *European Journal of Special Needs Education*, 35(3), 287–302. <https://doi.org/10.1080/08856257.2019.1646958>
- Knickenberg, M., Zurbruggen, C. L. A., & Schwab, S. (2022). Validation of the student version of the perceptions of inclusion questionnaire in primary and secondary education settings. *Sage Open*, 12(1). <https://doi.org/10.1177/21582440221079896>
- Kocaj, A., Kuhl, P., Jansen, M., Pant, H. A., & Stanat, P. (2018). Educational placement and achievement motivation of students with special educational needs. *Contemporary Educational Psychology*, 55, 63–83. <https://doi.org/10.1016/j.cedpsych.2018.09.004>
- Koch, T., Eid, M., & Lochner, K. (2018). Multitrait-Multimethod-Analysis: The psychometric foundation of CFA-MTMM Models. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 781–846). Wiley.
- Koch, T., Kelava, A., & Eid, M. (2018). Analyzing different types of moderated method effects in confirmatory factor models for structurally different methods. *Structural Equation Modeling*, 25(2), 179–200. <https://doi.org/10.1080/10705511.2017.1373595>
- Krolak-Schwerdt, S., Böhmer, M., & Gräsel, C. (2009). Verarbeitung von schülerbezogener Information als zielgeleiteter Prozess: Der Lehrer als «flexibler Denker» [Processing student-related information as a goal-directed process: The teacher as a «flexible thinker»]. *Zeitschrift für Pädagogische Psychologie*, 23(3/4), 175–186. <https://doi.org/10.1024/1010-0652.23.34.175>
- Kwon, K., Kim, E. M., & Sheridan, S. M. (2012). A contextual approach to social skills assessment in the peer group: Who is the best judge? *School Psychology Quarterly*, 27(3), 121–133. <https://doi.org/10.1037/a0028696>
- Lauermann, F., & Karabenick, S. A. (2011). Taking teacher responsibility into account (ability): Explicating its multiple components and theoretical status. *Educational Psychologist*, 46(2), 122–140. <https://doi.org/10.1080/00461520.2011.558818>
- Lauermann, F., & Karabenick, S. A. (2013). The meaning and measure of teachers' sense of responsibility for educational outcomes. *Teaching and Teacher Education*, 30, 13–26. <https://doi.org/10.1016/j.tate.2012.10.001>
- Lee, H., Herr, P. M., Kardes, F. R., & Kim, C. (1999). Motivated search: Effects of choice accountability, issue involvement, and prior knowledge on information acquisition and use. *Journal of Business Research*, 45(1), 75–88. [https://doi.org/10.1016/S0148-2963\(98\)00067-8](https://doi.org/10.1016/S0148-2963(98)00067-8)
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255. <https://doi.org/10.1037/0033-2909.125.2.255>
- Lüke, T., & Grosche, M. (2018). Konstruktion und Validierung der Professionsunabhängigen Einstellungsskala zum inklusiven Schulsystem (PREIS) [Construction and validation of an attitude scale measuring attitudes towards inclusive education in the population]. *Empirische Sonderpädagogik*, 1, 3–20.
- Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review*, 19, 85–103. <https://doi.org/10.1016/j.edurev.2016.06.003>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*, 73(1), 107–117. <https://doi.org/10.1037/0021-9010.73.1.107>
- Marucci, E., Oldenburg, B., Barrera, D., Cillessen, A. H. N., Hendrickx, M., & Veenstra, R. (2020). Halo and association effects: Cognitive biases in teacher attunement to peer-nominated bullies, victims, and prosocial students. *Social Development*, 30, 187–204. <https://doi.org/10.1111/sode.12455>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174. <https://doi.org/10.1007/BF02296272>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.
- McElvany, N., Schroeder, S., Hachfeld, A., Baumert, J., Richter, T., Schnotz, W., Horz, H., & Ullrich, M. (2009). Diagnostische Fähigkeiten von Lehrkräften: Bei der Einschätzung von Schülerleistungen und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern [Teachers' diagnostic skills to judge student performance and task difficulty when learning materials include instructional pictures]. *Zeitschrift für Pädagogische Psychologie*, 23(3/4), 223–235. <https://doi.org/10.1024/1010-0652.23.34.223>
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Möller, J., Jansen, T., Fleckenstein, J., Machts, N., Meyer, J., & Reble, R. (2022). Judgment accuracy of German student texts: Do teacher experience and content knowledge matter? *Teaching and Teacher Education*, 119, Article 103879. <https://doi.org/10.1016/j.tate.2022.103879>
- Monsen, J. J., Ewing, D. L., & Kwoka, M. (2014). Teachers' attitudes towards inclusion, perceived adequacy of support and classroom learning environment. *Learning Environments Research*, 17(1), 113–126. <https://doi.org/10.1007/s10984-013-9144-8>
- Muthén, L. K., & Muthén, B. (1998–2017). *Mplus user's guide* (Eighth edition). Muthén & Muthén.
- Neumann, L., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., & Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online*, 5(2), 80–109. <https://doi.org/10.25656/01:8426>
- Oh-Young, C., & Filler, J. (2015). A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities. *Research in Developmental Disabilities*, 47, 80–92. <https://doi.org/10.1016/j.ridd.2015.08.014>
- Orkibi, H., Ronen, T., & Assoulin, N. (2014). The subjective well-being of Israeli adolescents attending specialized school classes. *Journal of Educational Psychology*, 106(2), 515–526. <https://doi.org/10.1037/a0035428>
- Parsons, S., & Hallam, S. (2014). The impact of streaming on attainment at age seven: Evidence from the millennium cohort study. *Oxford Review of Education*, 40(5), 567–589. <https://doi.org/10.1080/03054985.2014.959911>
- Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis, S. G., Pierczynski, M., & Allen, M. (2018). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research*, 88(2), 205–242. <https://doi.org/10.3102/0034654317743198>
- Pavot, W. (2008). The assessment of subjective well-being: Success and shortfalls. In M. Eid, & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 124–140). Guilford.
- Phelps, L., Schmidt, C. D., & Boatright, B. (1986). The effects of halo and leniency on cooperating teacher reports using Likert-type rating scales. *The Journal of Educational Research*, 79(3), 151–154. <https://doi.org/10.1080/00220671.1986.10885668>
- Pielmeier, M., Huber, S., & Seidel, T. (2018). Is teacher judgment accuracy of students' characteristics beneficial for verbal teacher-student interactions in classroom? *Teaching and Teacher Education*, 76, 255–266. <https://doi.org/10.1016/j.tate.2018.01.002>
- Piqueras, J. A., Mateu-Martinez, O., Cejudo, J., & Pérez-Gonzalez, J.-C. (2019). Pathways into psychosocial adjustment in children: Modeling the effects of trait emotional intelligence, social-emotional problems, and gender. *Frontiers in Psychology*, 10, 507. <https://doi.org/10.3389/fpsyg.2019.00507>
- Praetorius, A.-K., Berner, V.-D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *Journal of Educational Research*, 106(1), 64–76. <https://doi.org/10.1080/00220671.2012.667010>
- Praetorius, A.-K., Drexler, K., Rösch, L., Christophel, E., Heyne, N., Scheunpflug, A., Zeinz, H., & Dresel, M. (2015). Judging students' self-concepts within 30s? Investigating judgement accuracy in a zero-acquaintance situation. *Learning and Individual Differences*, 37, 231–236. <https://doi.org/10.1016/j.lindif.2014.11.015>
- Praetorius, A.-K., Karst, K., Dickhauser, O., & Lipowsky, F. (2011). Wie gut schätzen lehrer die fähigkeitsselbstkonzepte ihrer schüler ein? Zur diagnostischen kompetenz von Lehrkräften [How teachers rate their students: On teachers' diagnostic competence regarding academic self-concept]. *Psychologie in Erziehung und Unterricht*, 58(2), 81–91. <https://doi.org/10.2378/peu2010.art30d>
- Praetorius, A.-K., & Südkamp, A. (2019). Diagnostische Kompetenz von Lehrpersonen hinsichtlich motivationaler Merkmale von Schülerinnen und Schülern [Diagnostic competence of teachers regarding motivational characteristics of students]. In H. Gaspard, U. Trautwein, & M. Hasselhorn (Eds.), *Diagnostik und Förderung von Motivation und Volition* (pp. 35–50). Hogrefe.
- Rogers, K. H., & Biesanz, J. C. (2019). Reassessing the good judge of personality. *Journal of Personality and Social Psychology*, 117(1), 186–200. <https://doi.org/10.1037/pspp0000197>
- Rubie-Davies, C. M., Peterson, E. R., Flint, A., Garrett, L., McDonald, L. M., Watson, P., & O'Neill, H. (2012). Ethnicity and teacher expectations in New Zealand. *Procedia – Social and Behavioral Sciences*, 69, 256–261. <https://doi.org/10.1016/j.sbspro.2012.11.407>
- Savolainen, H., Malinen, O.-P., & Schwab, S. (2020). Teacher efficacy predicts teachers' attitudes towards inclusion: A longitudinal cross-lagged analysis. *International Journal of Inclusive Education*, 1–15. <https://doi.org/10.1080/13603116.2020.1752826>
- Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2017). Professional growth in adaptive teaching competence as a result of Lesson Study. *Teaching and Teacher Education*, 68, 289–303. <https://doi.org/10.1016/j.tate.2017.09.015>
- Schmitt, M., Robbath, H.-G., Gresch, C., Stanat, P., Böhme, K., & Grosche, M. (2020). Projekt: Inklusion in der Sekundarstufe I in Deutschland (INSIDE) [Project: Inclusion in lower secondary education in Germany (INSIDE)]. *Erziehungswissenschaft*, 31(60), 199–202. <https://doi.org/10.3224/ezw.v31i1.30>
- Schwab, S., Zurbruggen, C. L. A., & Venetz, M. (2020). Agreement among student, parent and teacher ratings of school inclusion: A multitrait-multimethod analysis. *Journal of School Psychology*, 82, 1–16. <https://doi.org/10.1016/j.jsp.2020.07.003>
- Schwarz, N., & Strack, F. (1999). Reports of subjective well-being: Judgmental processes and their methodological implications. In D. Kahneman, E. Diener, & N. Schwarz

- (Eds.), *Well-being. The foundations of hedonic psychology* (pp. 61–84). Russell Sage Foundation.
- Sharma, U., Aiello, P., Pace, E. M., Round, P., & Subban, P. (2018). In-service teachers? Attitudes, concerns, efficacy and intentions to teach in inclusive classrooms: An international comparison of Australian and Italian teachers. *European Journal of Special Needs Education*, 33(3), 437–446. <https://doi.org/10.1080/08856257.2017.1361139>
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51(4), 455–498. <https://doi.org/10.3102/00346543051004455>
- Splett, J. W., Raborn, A., Brann, K., Smith-Millman, M. K., Halliday, C., & Weist, M. D. (2020). Between-teacher variance of students' teacher-rated risk for emotional, behavioral, and adaptive functioning. *Journal of School Psychology*, 80, 37–53. <https://doi.org/10.1016/j.jsp.2020.04.001>
- Stegenwallner-Schütz, M., Obry, M., Wittmann, E., Gehrer, K., Nusser, L., & Böhme, K. (2022). *INSIDE-Studie: Dokumentation der Skalierung der Kompetenzmessungen in den Bereichen Lesen und Mathematik des ersten Messzeitpunkts in der Jahrgangsstufe 6 (Kohorte 1)*. Leibniz-Institut für Bildungsverläufe. <https://doi.org/10.5157/LfBi:WP108:1.0> (LfBi Working Paper 108).
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743–762. <https://doi.org/10.1037/a0027627>
- Südkamp, A., Krawinkel, S., Lange, S., Wolf, S. M., & Tröster, H. (2018). Lehrkräfteinschätzungen sozialer Akzeptanz und sozialer Kompetenz: Akkuratheit und systematische Verzerrung in inklusiv geführten Schulklassen [Teachers' judgments of students' social acceptance and social skills: Perceptual accuracy and bias in inclusive classrooms]. *Zeitschrift für Pädagogische Psychologie*, 32(1–2), 39–51. <https://doi.org/10.1024/1010-0652/a000212>
- Suldo, S. (2016). Promoting student happiness. In *Positive psychology interventions in schools*. Guilford.
- Suldo, S., Thalji, A., & Ferron, J. (2011). Longitudinal academic outcomes predicted by early adolescents' subjective well-being, psychopathology, and mental health status yielded from a dual factor model. *The Journal of Positive Psychology*, 6(1), 17–30. <https://doi.org/10.1080/17439760.2010.536774>
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29. <https://doi.org/10.1037/h0071663>
- Tracey, D. K., Marsh, H. W., & Craven, R. G. (2003). Self-concepts of preadolescents with mild intellectual disabilities: Issues of measurement and educational placement. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self research* (pp. 203–229). Information Age Publishing.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- Urhahne, D. (2015). Teacher behavior as a mediator of the relationship between teacher judgment and students' motivation and emotion. *Teaching and Teacher Education*, 45, 73–82. <https://doi.org/10.1016/j.tate.2014.09.006>
- Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student: The underestimated student. *British Journal of Educational Psychology*, 81(1), 161–177. <https://doi.org/10.1348/000709910X504500>
- Urhahne, D., & Wijnia, L. (2021). A review on accuracy of teacher judgments. *Educational Research Review*, 32, Article 100374. <https://doi.org/10.1016/j.edurev.2020.100374>
- Urhahne, D., & Zhu, M. (2015). Accuracy of teachers' judgments of students' subjective well-being. *Learning and Individual Differences*, 43, 226–232. <https://doi.org/10.1016/j.lindif.2015.08.007>
- Van Miegheem, A., Verschueren, K., Petry, K., & Struyf, E. (2020). An analysis of research on inclusive education: A systematic search and meta review. *International Journal of Inclusive Education*, 24(6), 675–689. <https://doi.org/10.1080/13603116.2018.1482012>
- Venetz, M., Zurbriggen, C., & Eckhart, M. (2014). Entwicklung und erste Validierung einer Kurzversion des „Fragebogens zur Erfassung von Dimensionen der Integration von Schülern (FDI 4-6)“ von Haeblerlin, Moser, Bless und Klaghofer [Development and preliminary validation of a short questionnaire based on the FDI 4-6 by Haeblerlin, Moser, Bless, and Klaghofer measuring the educational integration of pupils]. *Empirische Sonderpädagogik*, 6(2), 99–113.
- Venetz, M., Zurbriggen, C. L. A., Eckhart, M., Schwab, S., & Hessels, M. G. P. (2015). *The Perceptions of inclusion questionnaire (PIQ) (German version)*. <https://piqinfo.ch>.
- Venetz, M., Zurbriggen, C. L. A., & Schwab, S. (2019). What do teachers think about their students' inclusion? Consistency of students' self-reports and teacher ratings. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.01637>
- Verschueren, K., & Koomen, H. M. Y. (2012). Teacher-child relationships from an attachment perspective. *Attachment & Human Development*, 14(3), 205–211. <https://doi.org/10.1080/14616734.2012.672260>
- Wang, G., & Williamson, A. (2020). Course evaluation scores: Valid measures for teaching effectiveness or rewards for lenient grading? *Teaching in Higher Education*, 27(3), 297–318. <https://doi.org/10.1080/13562517.2020.1722992>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427–450. <https://doi.org/10.1007/BF02294627>
- William, D., & Bartholomew, H. (2004). It's not which school but which set you're in that matters: The influence of ability grouping practices on student progress in mathematics. *British Educational Research Journal*, 30(2), 279–293. <https://doi.org/10.1080/0141192042000195245>
- Yada, A., Leskinen, M., Savolainen, H., & Schwab, S. (2022). Meta-analysis of the relationship between teachers' self-efficacy and attitudes toward inclusive education. *Teaching and Teacher Education*, 109, Article 103521. <https://doi.org/10.1016/j.tate.2021.103521>
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801>
- Zhu, M., & Urhahne, D. (2014). Assessing teachers' judgements of students' academic motivation and emotions across two rating methods. *Educational Research and Evaluation*, 20(5), 411–427. <https://doi.org/10.1080/13803611.2014.964261>
- Zhu, C., & Urhahne, D. (2021). Temporal stability of teachers' judgment accuracy of students' motivation, emotion, and achievement. *European Journal of Psychology of Education*, 36(2), 319–337. <https://doi.org/10.1007/s10212-020-00480-7>
- Zurbriggen, C., Venetz, M., Schwab, S., & Hessels, M. G. P. (2019). A psychometric analysis of the student version of the perceptions of inclusion questionnaire (PIQ). *European Journal of Psychological Assessment*, 35(5), 641–649. <https://doi.org/10.1027/1015-5759/a00044>