FAKULTÄT FÜR EMPIRISCHE HUMANWISSENSCHAFTEN UND
WIRTSCHAFTSWISSENSCHAFT
FACHRICHTUNG BILDUNGSWISSENSCHAFTEN

SCHOLASTIC ACHIEVEMENTS AND ACADEMIC SELF-CONCEPTS IN ELEMENTARY SCHOOL STUDENTS

DISSERTATION
ZUR ERLANGUNG DES ADADEMISCHEN GRADES EINES DOKTORS DER
NATURWISSENSCHAFTEN DER FAKULTÄT HW
BEREICH EMPIRISCHE HUMANWISSENSCHAFTEN
DER UNIVERSITÄT DES SAARLANDES

VORGELEGT VON:
REBECCA SCHNEIDER
AUS WOLMIRSTEDT
SAARBRÜCKEN, 2018

Dekan:
Prof. Dr. Stefan Strohmeier, Universität des Saarlandes

Berichterstatter/innen:
Prof. Dr. Jörn R. Sparfeldt, Universität des Saarlandes
Prof. Dr. Susanne R. Buch, Bergische Universität Wuppertal

Tag der Disputation:
11.12.2018

“In many ways, conducting doctoral research and writing the dissertation is like running a marathon, enduring hills and valleys, mountains, rivers and frantic road blocks along the way.”

– (Black, 2012, p. 97) –

Acknowledgements

I am grateful to all people who supported me to master all ups and downs during the last years and who motivated me for keeping up to finally cross the finishing line:

A very special gratitude goes to Prof. Dr. Jörn R. Sparfeldt for his guidance along the way and the protecting hand whenever a storm – or more realistic: a minor headwind – came up. Thanks for all the content-related and methodological support, for testing my limits (and showing me that I can do way more than I thought), and for giving me opportunities for further educational as well as personal development.

A huge thanks goes to my untiring cheerleaders and (former) colleagues Christin Lotz and Johannes Schult for their human warmth right from the day I entered my new office, for their strong company and friendly words when I felt like crying as well as for their content-related support. Christin, I will never forget all our shared laughter and tears during work time or at night, at the university, or somewhere abroad. Thanks for taking me along with you in good times and picking me up in bad ones. It is you who made work feel a bit like home. Johannes, you made statistical miracles happen. Thank you so much for your tenacious help with obstinate statistical models that drove me crazy. It was your merit that I don't want to hide under my bed anymore if phrases like 'multi-group comparisons' or 'measurement invariance testing' are used somewhere.

Furthermore, thanks to all other colleagues and scientific assistants of my working unit as well as to all down at Educational Sciences at Saarland University for a very pleasant working atmosphere, helpful discussions, canteen visits, and celebrations. Moreover, special thanks go to my favorite companions from other universities, Sabrina Navratil, Johannes Hellenbrand, and Nils Machts. Even strenuous days were full of enjoyable and unforgettable moments due to innumerable work/dissertation-related and private talks with you, skipping sessions at conferences together, or dancing at gala dinners all night long.

I am also grateful to my family, friends, flatmates, and scout friends who have provided me through moral and emotional support, who offered well-needed relaxation, distraction as well as sweets and alcoholic drinks. You were always keen to know what I was doing and how I was proceeding, although it is likely that most of you never really grasped what 'strange things' I was doing in my dissertation project. Nevertheless, I will miss all your screams of joy, your hugs, and celebrations whenever a significant milestone was reached.

Thanks to all of you for your encouragement! Without all of you this personal marathon would have been much more difficult and boring!

Content

List of tables

List of figures

List of abbreviations

ANOVA	analysis of variance
CFI	comparative fit index
CFT 1–R	Culture Fair Intelligence Test 1–Revision
df	degrees of freedom
FIML	full information maximum likelihood
I/E model	internal/external-frame-of-reference model
M	mean
MLR	maximum likelihood robust
RMSEA	root mean square error of approximation
SD	standard deviation
SDQ I	Self-Description Questionnaire I
SPPC	Self-Perception Profile for Children
SPSS	Statistical Package of the Social Sciences
SRMR	standardized root mean square residual
TLI	Tucker-Lewis index
WRMR	weighted root mean square residual

List of publications

The present thesis is based on two articles that are published in peer-reviewed journals. These articles are available online through the respective publishing company. Furthermore, this thesis comprises one manuscript that is currently under revision.

Study 1

Schneider, R., & Sparfeldt, J. R. (2016). Zur (Un-)Genauigkeit selbstberichteter Zensuren bei Grundschulkindern [The accurarcy of self-reported grades in elementary school]. *Psychologie in Erziehung und Unterricht*, *63*, 48–59.

Study 2

Schneider. R., & Sparfeldt, J. R. (under revision). Twofold multidimensional academic self-concepts in elementary school students: Social and dimensional comparisons. *Contemporary Educational Psychology*.

Study 3

Schneider, R., Lotz, C., & Sparfeldt, J. R. (2018). Smart, confident, interested: Contributions of intelligence, self-concept, and interest to elementary school achievement. *Learning and Individual Differences*, *62*, 23–35.

In the following, I will use the term *we* when referring to these studies that my co-author(s) and I have conducted together.

Summary

Positive academic self-concepts are valued as highly desirable outcomes in educational contexts and are closely associated with scholastic achievements. With increasing scholastic experiences, academic self-concepts become more realistic. Therefore, the relation between academic self-concepts and scholastic achievements seems to be more pronounced in higher elementary school grades than in lower grades. In this thesis, it was aimed to examine the relations between scholastic achievements and academic self-concepts separately for each elementary school grade and to compare these relations across grades to provide initial evidence for developmental processes in elementary school. Furthermore, special attention is paid to the distinction of school subject-specific academic self-concepts into a competence component (self-perceived ability in a school subject) and an affect component (intrinsic motivational-affective self-perceptions).

Constituting the basis for further analyses on the relations between scholastic achievements and academic self-concepts in elementary school students, analyses regarding the assessment of reported grades as indicators of scholastic achievements (study 1; Schneider & Sparfeldt, 2016) and concerning the assessment of academic self-concepts with a differing number of response categories (excursus) were conducted. In study 1, students' self-reported grades were less accurate in grade 2 and 3 compared with fourth graders. For fourth graders, student reported grades seemed to be appropriate indicators of their actual reported grades. The results of the excursus indicated that data obtained from elementary school students did not differ fundamentally across response formats for competence and affect self-concepts (rating scales with either three, four, or five response categories).

In study 2 (Schneider & Sparfeldt, under revision), we examined the effects of scholastic achievements in mathematics and German on (non-)corresponding competence and affect self-concepts in elementary school students. Within the internal/external-frame-of-reference model (I/E model, Marsh, 1986), these effects are associated with social and dimensional comparison processes. Social comparison processes (external comparison of one's own achievement in one subject with the achievements of peers in the same subject) seem to increase during elementary school, cognitively more demanding dimensional comparison processes (internal comparison of one's academic achievement in one subject with one's own achievement in another subject) seem to occur earliest in third grade. In study 2, we extended prior research by assigning the competence-affect-distinction to the I/E model framework to simultaneously examine relations between scholastic achievements and both academic self-concept components in elementary school students. Path coefficients of reported grades on corresponding academic self-concepts indicating social comparison processes were positive and of

moderate to high magnitude in all four elementary school grades. Path coefficients of reported grades and non-corresponding competence and affect self-concepts indicating dimensional comparisons were mostly negative; some path coefficients were substantial. The relations between reported grades and corresponding competence self-concept measures were more pronounced compared to those between reported grades and affect self-concepts. Regarding grade-related differences, the relation between reported grades in mathematics and the mathematics competence self-concept was stronger in grade 4 compared to grade 1. The findings of study 2 emphasized the importance of reported grades for students' corresponding competence and affect self-concepts via social comparison processes. In contrast, dimensional comparison processes seemed to play a minor role.

In study 3 (Schneider, Lotz, & Sparfeldt, 2018), the focus was shifted towards subject-specific academic self-concepts as statistical predictors of scholastic achievements. The few prior studies with fourth graders indicated that cognitive variables, especially intelligence, are more important predictors compared to subject-specific academic self-concepts for corresponding scholastic achievements. Regarding academic self-concepts, competence self-concepts revealed higher path coefficients on corresponding scholastic achievements than affect self-concepts. Study 3 was the first to simultaneously inspect the relevance of intelligence, competence self-concept, and affect self-concept for reported grades in elementary school grades 2, 3, and 4. When all three predictors were jointly considered, competence and affect self-concepts substantially contributed to the prediction of the reported grade in mathematics and German beyond intelligence in the examined elementary school grades (with the exception of affect self-concept of fourth graders in German). As expected, competence self-concept was a stronger predictor than affect self-concept. Furthermore, we found no substantial differences in the prediction of reported grades across grades. The findings of study 3 underlined the significance of competence self-concept above and beyond intelligence, whereas affect self-concept seemed to play a minor role for the statistical prediction of reported grades.

To conclude, the results of this thesis confirmed the close relation between scholastic achievements (operationalized by reported grades) and corresponding academic self-concepts for each elementary school grade. Although competence and affect self-concepts showed very large construct overlaps within subjects in all elementary school grades, relations between reported grades and corresponding competence self-concept measures were more pronounced compared to those between reported grades and affect self-concept measures in both studies. Differences in the relations between reported grades and both self-concept components across elementary school grades were mostly non-substantial.

Zusammenfassung

Schulische Selbstkonzepte weisen eine hohe Bedeutung in schulischen Lehr- und Lernkontexten auf, u.a. sind sie eng mit schulischen Leistungen verbunden. Mit zunehmender Schulerfahrung entwickeln Schülerinnen und Schüler zunehmend realistische schulfachspezifische Selbstkonzepte. Als Folge stehen schulfachspezifische Leistungen und Selbstkonzepte in höheren Klassenstufen in zunehmend engerer Beziehung miteinander. Ziel dieser Dissertation war die Untersuchung des Beziehungsgeflechts schulischer Leistungen und schulischer Selbstkonzepte separat für die Grundschulklassenstufen 1 bis 4. Ein Vergleich dieser Beziehungen zwischen den Klassenstufen sollte Hinweise auf mögliche Entwicklungsprozesse über die Grundschuljahre liefern. Ein besonderes Augenmerk wurde in der Dissertation auf die etablierte Trennung schulfachspezifischer Selbstkonzepte in eine Kompetenzkomponente (Wahrnehmung der eigenen Fähigkeit) und eine Affektkomponente (motivational-affektive Selbstwahrnehmung) gelegt.

Grundlage für die Analysen zum Beziehungsgeflecht schulischer Leistungen und Selbstkonzepte bei Grundschulkindern bildeten Untersuchungen zur Akkuratesse schülerberichteter Zeugnisnoten (Studie 1, Schneider & Sparfeldt, 2016) und zur Messung schulischer Selbstkonzepte mit unterschiedlich gestuften Ratingskalen (Exkurs). Die Ergebnisse von Studie 1 deuten auf eine geringere Akkuratesse schülerberichteter Zeugnisnoten in den Klassenstufen 2 und 3 im Vergleich zu Viertklässlern hin. Für Viertklässler scheint eine Verwendung schülerberichteter Zeugnisnoten als Indikatoren ihrer tatsächlichen Leistungen angemessen. Ergebnisse des Exkurses legen eine adäquate sowie vergleichbare Bearbeitung von Ratingskalen mit drei-, vier- oder fünfstufigem Antwortformat bei Grundschulkindern nahe.

In Studie 2 (Schneider & Sparfeldt, in Revision) wurden Effekte schulischer Leistungen in Mathematik und Deutsch auf ihre (nicht-)korrespondierenden Kompetenz- und Affektselbstkonzepte bei Grundschülerinnen und -schülern untersucht. Im Rahmen des Internal/external-frame-of-reference-Modells (I/E Modell; Marsh, 1986) sind diese Effekte mit sozialen und dimensionalen Vergleichsprozessen assoziiert. Während soziale Vergleichsprozesse (Vergleich der eigenen schulfachspezifischen Leistung mit den Leistungen anderer im gleichen Fach) über die Grundschuljahre hinweg zunehmen, scheinen die kognitiv anspruchsvolleren dimensionalen Vergleichsprozesse (Vergleich der eigenen Leistung in verschiedenen Schulfächern) erst in der dritten Klassenstufe einzusetzen. Studie 2 verband die etablierte Kompetenz-Affekt-Differenzierung schulfachspezifischer Selbstkonzepte und das I/E Modell. Es zeigten sich substantielle, moderate bis hoch ausgeprägte Pfadkoeffizienten der Mathematiknote bzw. Deutschnote auf die korrespondierenden kompetenzbezogenen Selbstkonzepte (soziale Vergleichs-

prozesse) sowie fast durchgängig keine bedeutsamen Effekte der Zeugnisnoten auf die nicht-korrespondierenden kompetenz- und affektbezogenen Selbstkonzepte (dimensionale Vergleiche). Die Beziehungen zwischen Noten und Kompetenzselbstkonzepten im gleichen Schulfach waren stärker ausgeprägt als die korrespondierenden Noten–Affekt-Selbstkonzeptbeziehungen. Bezüglich möglicher Unterschiede zwischen Grundschulklassenstufen zeigte sich ein höherer Pfadkoeffizient von der Mathematiknote auf das Kompetenzselbstkonzept in Mathematik in Klassenstufe 4 verglichen mit Klassenstufe 1. Die Ergebnisse dieser Studie verdeutlichen die Bedeutung von Schulnoten auf die korrespondierenden Kompetenz- und Affektselbstkonzepte via sozialer Vergleichsprozesse. Dimensionale Vergleichsprozesse scheinen in der Grundschule eine eher untergeordnete Rolle zu spielen.

In Studie 3 (Schneider, Lotz, & Sparfeldt, 2018) lag der Fokus auf der statistischen Prädiktion von Schulleistung in den Fächern Mathematik und Deutsch durch korrespondierende Kompetenz- und Affektselbstkonzepte. Zumindest für Viertklässler gelten kognitive Variablen – insbesondere die Intelligenz – als bedeutendere Prädiktoren für Schulerfolg im Vergleich zu schulischen Selbstkonzepten. Studie 3 erweiterte bisherige Befunde und untersuchte die (relative) Bedeutsamkeit von Intelligenz sowie schulfachspezifischer Kompetenz- und Affektselbstkonzepte bei der gemeinsamen statistischen Vorhersage von Zeugnisnoten separat für die Fächer Mathematik und Deutsch. Bei simultaner Berücksichtigung der drei Prädiktoren waren Kompetenz- sowie Affektselbstkonzepte in allen Klassenstufen substantielle Prädiktoren der korrespondierenden Zeugnisnoten über Intelligenz hinaus (mit Ausnahme des affektiven Selbstkonzepts in Klassenstufe 4 in Deutsch). Das Kompetenzselbstkonzept war der bedeutendere Prädiktor im Vergleich zum affektbezogenen Selbstkonzept. Ein Vergleich der Pfadkoeffizienten zwischen den Klassenstufen ergab für beide Fächer keine bedeutsamen Unterschiede in der Höhe der Koeffizienten. Die Ergebnisse von Studie 3 verweisen auf die Bedeutung von Intelligenz und Motivation für Schulerfolg, auch in der Grundschule. Sie unterstreichen insbesondere die Bedeutung des Kompetenzselbstkonzepts für die Schulleistung. Affektselbstkonzepte spielen hingegen nur eine kleine Rolle bei der statistischen Prädiktion von Zeugnisnoten.

Die Ergebnisse dieser Dissertation unterstreichen die enge Beziehung zwischen schulischen Leistungen (operationalisiert durch Noten) und korrespondierenden schulischen Selbstkonzepten in den untersuchten Grundschulklassenstufen. Obwohl Kompetenz- und Affektselbstkonzepte hohe Anteile geteilter Varianz innerhalb eines Schulfachs aufwiesen, zeigten sich konsistent engere Beziehungen zwischen Noten und korrespondierenden Kompetenzselbstkonzepten als zwischen Noten und Affektselbstkonzepten im gleichen Schulfach. Alterskorrelierte Veränderungen in den Beziehungen schulischer Leistungen und Selbstkonzepte waren zumeist nicht substantiell.

1 Introduction

Elementary school as the first mandatory educational institution in Western countries plays a particularly important role in the context of lifelong learning: it is here that fundamental competencies like basic arithmetic operations, reading, and writing are taught. These competencies are essential cornerstones for secondary school and post school as well as lifelong learning and education (Kultusministerkonferenz, 2015). In the school year 2017/2018, around 2.770.000 children were enrolled in elementary schools in Germany (Statistisches Bundesamt, 2017). During their years in elementary school, these students gain manifold scholastic experiences with, for example, different contents, didactic concepts, teachers, classmates, and other students. They have to meet performance requirements, deal with achievement-related feedback (e.g., in the form of reported grades) as well as to evaluate their motivation towards different school subjects.

There certainly is a wide range of scholastic experiences in school. The present thesis will focus on two important aspects in elementary school students: scholastic achievements and academic self-concepts. Both continue to be major topics in educational and psychological research. More precisely, it was aimed to examine the relations between scholastic achievements and academic self-concepts separately for each elementary school grade and to compare these relations among grades to provide initial evidence for developmental processes.

Prior to investigating these relations, we examined the accuracy of elementary school students' self-reported grades in comparison to their corresponding teacher-reported grades (study 1). Additionally, attention was paid to the assessment of academic self-concepts with different response formats (with either three, four, or five response categories; excursus). Analyses regarding the assessment of student-reported grades as indicators of scholastic achievements and the assessment of academic self-concepts constituted the basis for further analyses on the relations between scholastic achievements and academic self-concepts in elementary school students.

Despite an increasing body of research on the relations between scholastic achievements and academic self-concepts (e.g., for different age spans, subjects, and cultures; Chen, Yeh, Hwang, & Lin, 2013; Möller, Pohlmann, Köller, & Marsh, 2009; Niepel, Brunner & Preckel, 2014; Schilling, Sparfeldt, & Rost, 2004), so far only a small number of studies dealt with elementary school students. This age span (elementary

school students are typically six to ten years old) is especially relevant due to far-reaching developmental processes like increasing cognitive abilities, presumably higher amounts of comparison processes with significant peers, and an increasing internalization of evaluative standards (e.g., Guay, Marsh, & Boivin, 2003). Therefore, grade-related differences in the relations between scholastic achievements and academic self-concepts seem plausible. For example, Wigfield & Karpathian (1991) argued that students "with high perceptions of ability would approach new tasks with confidence, and success on those tasks is likely to bolster their confidence in their ability" (p. 255). These authors assumed that this reciprocal relation is based on more firmly established, realistic competence self-concepts (self-perceived competencies in one subject) in older elementary school students. Accordingly, the relations between scholastic achievement and competence self-concepts (as one component of academic self-concepts) within one school subject should be more pronounced in higher elementary school grades compared to younger students. For affect self-concepts (subjective motivation-affective self-perceptions in one subject) as the other component of academic self-concepts, comparable grade-related differences could be presumed. However, a fundamental prerequisite to ensure meaningful comparisons of relations between scholastic achievements and academic self-concepts across elementary school grades is an analogue assessment of variables in all elementary school grades (construct and measurement equivalence).

When examining relations between scholastic achievements and academic self-concepts for each grade as well as across grades, two separate approaches can be considered: (1) the internal/external-frame-of-reference model (I/E model; Marsh, 1986) which explains effects of scholastic achievements on academic self-concepts and (2) prediction models of scholastic achievement with academic self-concepts as important motivational predictors beyond other variables. Both approaches differ regarding their theoretical foundations and their corresponding methodological approaches.

(1) Within the I/E model framework, effects of scholastic achievements on academic self-concepts are associated with social and dimensional comparison processes. Students compare their scholastic achievements in one subject to the achievements of their peers in the same subject (social comparison), thereby forming a corresponding self-concept. Dimensional comparisons refer to the contrast of a students' academic achievement in one school subject with his or her own achievement in another school subject. The joint operation and especially the relative weight of social and dimensional comparison processes play an important role for the formation of subject-specific academic self-concepts. Across elementary school grades, social comparison processes seem to increase, whereas cognitively more demanding dimensional comparison processes seem to occur earliest in third grade (e.g., Ehm, Lindberg, & Hasselhorn, 2014; Harter, 2006; Skaalvik & Skaalvik, 2002). Unfortunately, the interpretation of prior

findings regarding grade-related differences is limited by, for example, the use of different indicators of scholastic achievements and/or academic self-concepts across grades. Furthermore, as studies with elementary students have been limited to competence self-concepts, no I/E model studies with affect self-concepts were conducted so far. Thus, it seemed to be fruitful to (a) extend the traditional I/E model by the competence–affect distinction of academic self-concepts to examine effects of scholastic achievements on competence and affect self-concepts and (b) to subsequently compare these relations across grades.

(2) Furthermore, academic self-concepts are considered to be important determinants of scholastic learning behavior and performance (e.g., Craven & Marsh, 2008; Hattie, 2009; Marsh & O'Mara, 2008). Due to their motivational characteristics, academic self-concepts are supposed to explain corresponding achievement variance above and beyond other (especially cognitive) variables when predicting corresponding scholastic achievements (e.g., Kriegbaum, Jansen, & Spinath, 2015; Lotz, Schneider, & Sparfeldt, 2018; Spinath, Spinath, Harlaar, & Plomin, 2006; Steinmayr & Meißner, 2013; Steinmayr & Spinath, 2009). In fourth graders, competence self-concepts or affect self-concepts revealed to be substantial predictors of scholastic achievement beyond intelligence in prediction models with solely intelligence and one self-concept component (Spinath et al., 2006). However, the predictive power of affect self-concept vanished when the three predictors were considered simultaneously. Probably, this effect occurred due to the large amount of shared variance between competence and affect self-concepts. For elementary school students in lower elementary school grades, no such analyses were conducted so far. Hence, it seemed to be worthwhile to (a) extend prior research to elementary school grades below grade 4, and (b) to compare the prediction pattern across elementary school grades. Furthermore, unique effects of each predictor in terms of incremental predictive validity as well as the proportion of common variance explained by the predictors were inspected.

To conclude, the aim of this thesis was to examine the relations between scholastic achievements and academic self-concepts in elementary school students – separately for each elementary school grade and across grades. Thereby, particular emphasis was laid on the competence–affect distinction of subject-specific academic self-concepts. This distinction allowed the examination of possibly differential relations between scholastic achievement and the two correlated, but still distinct self-concept components within the I/E model framework and within prediction models of scholastic achievements.

2 Theoretical framework

This chapter provides a brief summary about the theoretical framework this thesis is based on. First, the theoretical background concerning scholastic achievement and academic self-concepts is provided. Second, relations between scholastic achievements and academic self-concepts in elementary school grades 1 to 4 and across these grades are portrayed. Thereby, attention is paid to the formation of academic self-concepts by comparison processes of scholastic achievements and to the incremental validity of academic self-concepts above and beyond intelligence when predicting scholastic achievements.

2.1 Doing well at school: Scholastic achievements

Good scholastic achievements are valued as desirable outcomes in education. Reported grades and standardized competence tests are the two types of evidence that are most commonly used for decisions in educational contexts (Willingham, Pollack, & Lewis, 2002). The following chapter first addresses important characteristics of reported grades as the most relevant indicator of scholastic achievement in Germany. Second, the implementation of reported grades and the grading procedure in German elementary schools are briefly described. Third, attention is dedicated to the assessment of scholastic achievement by students' self-reported grades as indicators of their actual (teacher-reported) grades.

2.1.1 Scholastic achievements

Scholastic achievement can be defined as learning processes and learning outcomes of students that are initiated by school (Ingenkamp & Lissmann, 2008, p. 131). Reported grades and standardized competence tests as typically used curricular-based indicators of scholastic achievements are supposed to signify "the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university" (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014, p. 1; also see Steinmayr, Sauer, & Gamsjäger, 2018). Reported grades and standardized competence tests are both assumed to tap aspects of skill and knowledge, but the achievement-relevant aspects of both measures seem to overlap only partly. For example, even if a scholastic competence test is curricular valid,

it cannot cover exactly the same material in detail that the averaged reported grade had assessed (see below). Therefore, moderate to high correlations between both measures of $.40 \leq r \leq .72$ are typically found (e.g., Helmke & van Aken, 1995; Krüsken, 2007; Lorenz & Artelt, 2009; Marsh, 2007; Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005).

Reported grades on (midterm) report cards are usually a conglomerate of several single reported grades in a specific subject over (half) a school year. Reported grades, therefore, reflect a broad range of scholastic achievements including written and oral tests or the oral participation in class. Hence, they are based on more comprehensive and representative performance information than standardized competence tests (Schrader & Helmke, 2001). Although reported grades have been and still are criticized frequently (cf. Birkel & Tarnai, 2018; Ingenkamp, 1971, 1995; Jäger & Lissmann, 2004), they fulfill manifold important functions in the German school system: For students, they are an immediate and salient form of feedback by teachers, they can motivate or have a disciplinary effect, they point out performance differences between students for the same subject as well as possible intraindividual performance differences between subjects and/or points of time, and make students familiar with educational standards. For parents, teachers, and the educational system in general, reported grades provide information on eventually required educational support and are of high importance to students and their parents for the promotion to the next academic year (Beutel, Lütgert, Tillmann, & Vollstädt, 1999; Birkel & Tarnai, 2018; Heine, Briedis, Didi, Haase, & Trost, 2006; Willingham et al., 2002). Moreover, they play an important role in the decision of being able to enter educational programs or institutions. For example, the important decision for a secondary school track has to be made after grade 4 in most German federal states which is typically based on students' reported grades (Baumert, Trautwein, & Artelt, 2003; Bos et al., 2004).

Single reported grades as well as reported grades on report cards in particular are heterogeneous, multi-determined measures that rely on the student's achievement in a specific subject and are additionally influenced by non-ability aspects such as motivation, volition, or effort (Ditton & Krüsken, 2009a; Harlen, 2005; Willingham et al., 2002; for prediction models of scholastic achievements see, e.g., Helmke & Schrader, 2018; Helmke & Weinert, 1997; Sauer & Gamsjäger, 1996; an overview is provided by Hasselhorn & Gold, 2009). For reported grades on report cards, motivation should be especially relevant. Students need to work constantly throughout the school year to receive good single reported grades and, in turn, good reported grades on their report cards. This consistency in learning needs higher levels of motivation. Therefore, motivation might be already indirectly assessed by reported grades resulting in substantial relations of small to medium magnitude between reported grades on report cards and motivational aspects (for elementary school, e.g., competence self-concept – reported

grade correlation in mathematics for second/third/fourth graders r = .35/.40/.52, eagerness to learn – reported grade correlation in mathematics r = .26/.32/.35; Helmke, 1997a).

In contrast, standardized competence tests are not generated by the teachers themselves, but are made available by specific publishing companies or authors. For example, the performance of elementary school students in mathematics can be assessed by the series of "Deutsche Mathematiktest für erste/zweite/dritte/vierte Klassen" (Gölitz, Roick, & Hasselhorn, 2006; Krajewski, Küspert, & Schneider, 2002; Krajewski, Liehm, & Schneider, 2004; Roick, Gölitz, & Hasselhorn, 2004) and in reading by "Ein Leseverständnistest für Erst- bis Sechstklässler" (ELFE 1–6, Lenhard & Schneider, 2006). Writing performance can be administered by "Deutscher Rechtschreibtest für das erste und zweite Schuljahr" (DERET 1–2+; Stock &Schneider, 2008a) or "Deutscher Rechtschreibtest für das dritte und vierte Schuljahr" (DERET 3–4+; Stock & Schneider 2008b). These standardized competence tests are based on respective curricular contents as well as educational standards and meet psychometric criteria in the defined scope (Heller & Hany, 2001). Furthermore and in contrast to reported grades, they are more robust against expectation effects and should not be influenced by frames of reference (see big-fish-little-pond effect; e.g., Marsh, 1987; Marsh et al., 2008; Trautwein & Lüdtke, 2005). However, standardized competence tests are one-point measures that only reflect the performance of a student in a particular test and, therefore, only cover parts of students' actual scholastic performance. Moreover, standardized competence tests play a minor role in the German school system because they are often not graded. Therefore, German students do not study extensively for such assessments with standardized competence tests.

2.1.2 Reported grades in German elementary school students

Because of the outstanding significance of reported grades in the German educational system, the grading system in German elementary schools is briefly described in this subsection. Students mostly gain their first experiences with reported grades in elementary school. In Germany, elementary school spans grades 1 to 4 (except in the federal states Berlin and Brandenburg, where elementary school includes grades 1 to 6). The specific grade in which numerical reported grades are implemented varies between federal states due to federal state-specific school and education policy regulations (see Füssel & Leschinsky, 2008). In several federal states like the State of Saxony or Bavaria students receive reported grades from second grade onwards (Bayrisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst, 2017; Sächsisches Staatsministerium für Kultus, 2017), in other states students receive reported grades no earlier than in third grade. In elementary schools in Schleswig-Holstein, teachers do not grade their students at all; however, reported grades can be used to supplement verbal evaluations in grade 3 and 4 (school-specific decision; Ministerium für Wissenschaft,

Bildung und Kultur Schleswig-Holstein, 2015). Elementary school teachers are faced with the task of familiarizing their students with reported grades and of promoting age/development-appropriate reflections about their students' learning behavior and scholastic achievements. Thus, teachers are supposed to strengthen their students' self-evaluation competencies, enable them to set and pursue their own goals as well as to interpret external performance assessments (e.g., reported grades) as learning opportunities (Kultusministerkonferenz, 2015).

The German grading system consists of reported grades from 1 (very good) to 6 (insufficient), with numerically lower reported grades indicating better scholastic achievements. In general, the majority of elementary school students receives at least satisfying reported grades (reported grades = 1, 2, or 3). Poor reported grades (reported grades = 4, 5, or 6) are relatively sparse in elementary school. For example, 80%/74% of the examined third graders (N = 585) got satisfying or better reported grades in mathematics/German on their midterm report cards (teacher-reported grades; Krüsken, 2007). Poor reported grades were rare (20%/26% in mathematics/German). For these students, the proportion of satisfying or better reported grades on their midterm report cards in grade 4 was slightly lower in comparison with grade 3 in both subjects (at least satisfying reported grades in grade 4: 73%/70% in mathematics/German). More specifically, the relative frequencies of students who received the reported grade 1 (very good) or 2 (good) on their midterm report card decreased (6% or 7%). Correspondingly, the proportion of reported grades 3 to 6 slightly increased (for the reported grade 3/4/5/6: Δ = 6%/2%/5%/1%). Valtin, Wagner, and Schwippert (2005) reported comparable results: 80% of the examined fourth graders (N = 7.633, student-reported grades from midterm report cards) got satisfying or better reported grades in mathematics or German. Again, poor reported grades were relatively sparse. Nonetheless, the results by Krüsken (2007) indicated a high temporal stability of reported grades in elementary school. Further research revealed comparable results ($.67 \leq r \leq .76$; Sauer & Gamsjäger, 1996; von Maurice, Dörfler, & Artelt, 2014) indicating that interindividual differences are quite stable across elementary school grades.

2.1.3 Assessment of reported grades

In educational and psychological research, reported grades are often used as easily accessible indicators of students' scholastic achievements; extensive use is made of the reported grades on midterm or end-of-year report cards to depict the average performance capability of a student. Furthermore, students' self-reported grades are often used as economic and efficient indicators of teacher-reported (actual) grades as presented on report cards. When examining the accuracy of students' self-reported grades in comparison to their corresponding perfectly accurate teacher-reported grades, a correlational perspective and discrepancy aspects between both reports (grades reported

by students or teachers) should be taken into account. With regard to correlations between reports, possible systematic biases of students' self-reported grades might still result in high correlations between both reports. Therefore, discrepancy aspects like average differences between student- and teacher-reported grades as well as the percentages of correctly reported grades, under-, and over-reports need to be additionally considered (Kuncel, Credè, & Thomas, 2005; Sparfeldt, Buch, Rost, & Lehmann, 2008; Sticca, Goetz, Bieg, Hall, Eberle, & Haag, 2017). Furthermore, the examination of correlations of both reports with additional criteria might emerge in an even more comprehensive picture regarding the accuracy of students' self-reported grades. Differential relations of both reports with additional criteria might provide evidence for systematic biases and, thus, limitations regarding the validity of student-reported grades (Sparfeldt et al., 2008).

Concerning prior studies, a meta-analysis by Kuncel et al. (2005) suggested some biases in self-reported GPAs (reported grades from all current classes were averaged for the grading period). In contrast, German secondary school students seem to report their grades accurately: (very) high correlations and low differences (in terms of an over-report) between student- and teacher-reported grades, (very) high amounts/percentages of correctly reported grades, and no/hardly any differential correlations with additional criteria were reported (Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008)[1]. For example, Sparfeldt et al. (2008) found a correlation between student- and teacher-reported grade in mathematics of $r = .94$; 91% of the tenth grade students indicated their reported grade correctly. Differences between student- and teacher-reported grades indicated a small overestimation of self-reported grades ($d \leq 0.18$; 1% underestimators, 8% overestimators). Furthermore, this overestimation did not correlate substantially and consistently with additional criteria.

Regarding elementary school students, one might expect self-reported grades to be less accurate in comparison with secondary school students due to lower cognitive abilities and less experience with reported grades as well as the propensity of elementary school students to overestimate their competencies (Helmke, 1999; Wild, 1991). So far, only one study reported some aspects concerning the accuracy of student-reported grades with elementary school students (Ostrop, Schmude, & Valtin, 2002). At the beginning of the school year, third ($N = 98$) and fourth graders ($N = 172$) were asked to report their grades from their last report cards (reported grades at the end of grade 2 or 3). 92% of the fourth graders reported their grade in mathematics correctly (85% for reading, sport and arts). Across all examined subjects and subject domains (e.g., reading or writing), 80% of the grades were reported correctly; over- and under-reports occurred equally with 10% each. For third graders, Ostrop et al. (2002) only stated that self-reported grades were less

[1]After study 1 was accepted by the journal, further studies regarding the accuracy of subject-specific reported grades were published. Comparable results were reported for Chinese (Feng & Rost, 2015) and Swiss secondary school students (Sticca, Goetz, Bieg et al., 2017).

accurate in comparison with fourth graders; unfortunately, no specific values were reported for third graders. Furthermore, better reported grades were reported more accurate compared to poorer grades, regardless of the subject. A replication of these findings for fourth graders as well as an extension to further accuracy aspects (e.g., average differences between student- and teacher-reported grades or correlations with additional criteria) and further elementary school grades remained research desiderata. Thus, one research aim of this thesis was to examine the accuracy of elementary school students' self-reported grades in comparison to their corresponding teacher-reported (actual) grades.

2.2 Being confident and interested: Academic self-concepts

Academic self-concepts are extensively investigated in the school context due to their close association with behavioral, emotional, and cognitive variables (e.g., Chen et al., 2013; Guay, Larose, & Boivin, 2004; Marsh, 2007; Marsh & Craven, 2006; Marsh, Hau, Artelt, Baumert, & Peschar, 2006; Pinxten, De Fraine, Van Damme, & D'Haenens, 2010). Self-concepts can be defined as cognitive-descriptive concepts of a person about him- or herself (Moschner & Dickhäuser, 2018). Concerning school, academic self-concepts refer to the self-perceptions of students that are formed by experiences with the school environment and their corresponding interpretations (Shavelson, Hubner, & Stanton, 1976). The following chapter provides a brief summary about the twofold multidimensional structure of academic self-concepts – first for students in general and second with a particular focus on elementary school students. Third, attention is paid to the assessment of academic self-concepts with rating scales that either have a three-, four-, or five-point Likert-type scale response format.

2.2.1 Academic self-concepts

Positive academic self-concepts are a highly desired goal in education and are associated with learning behavior (like planning or persistent engagement in related tasks), attribution of success and failure, scholastic achievements as well as course selections in secondary school (e.g., Bong & Skaalvik, 2003; Hattie, 2009; Marsh & O'Mara, 2008). For example, students with positive self-concepts seem to attribute success to their own ability, but failures to external factors like bad luck. By contrast, students with negative self-concepts tend to attribute success to unstable causes like effort or luck, but failures to internal causes like low abilities (e.g., Faber, 2007; Möller & Köller, 2000; Pekrun, 2000; Schwarzer & Jerusalem, 1982).

Concerning the self-concept structure, Shavelson et al. (1976) introduced a multidimensional hierarchical self-concept model with one global and very general self-concept factor at the apex, the distinction of academic and non-academic self-concepts on

the second level, and even more specific self-concept components like self-concept factors related to specific curriculum domains or school subjects on the third level (see Fig. 1). The authors postulated that subject-specific academic self-concepts should correlate substantially among each other. Yet, researchers repeatedly failed to find the assumed single second-level academic self-concept factor and the postulated substantial positive correlation between mathematics and verbal self-concepts (e.g., native language self-concept). Instead, fit indices consistently supported a model with two separate self-concept factors, academic mathematics and academic verbal (e.g., Marsh, 1990, 1992; Marsh, Byrne, & Shavelson, 1988). The reported small to negligible self-concept correlations between mathematics self-concept factors and verbal self-concept factors ($-.10 \leq r \leq .12$; Marsh, 1986) led to the Marsh/Shavelson model (Marsh, 1990; Marsh et al., 1988) that posited distinct academic mathematics and academic verbal self-concept factors instead of a single superordinate second-order academic self-concept factor. Domain-specific self-concepts of mathematics and native language loaded on "their" respective academic mathematics or academic verbal self-concept factor, whereas self-concepts in, for example, biology simultaneously loaded on the higher-order mathematics and verbal self-concept factors.

In the follwoing, the term "domain-specific" refers to the second-order academic mathematics and academic verbal self-concept factors of the Marsh/Shavelson model (Marsh, 1990; Marsh et al., 1988). Across domains, academic self-concept correlations are typically non-substantial or of small magnitude (e.g., Marsh, 1986; Möller et al., 2009). The term "subject-specific" refers to the third order self-concept factors in different school subjects like mathematics, native language, or biology, or specific aspects of these subjects like reading for the native language. Such subject-specific self-concepts are more highly correlated within domains compared to correlations between subject-specific self-concepts across domains (e.g., Möller et al., 2009; Rost, Sparfeldt, & Schilling, 2007; Schilling et al., 2004).

Apart from the subject-specificity of academic self-concepts (e.g., Arens, Yeung, Craven, & Hasselhorn, 2011; Ehm, Lindberg et al., 2014; Möller at al., 2009; Rost & Sparfeldt, 2002; Schilling et al., 2004), each subject-specific academic self-concept seems to be further subdivided into two separate but intercorrelated components: a competence component and an affect component. The competence self-concept (synonyms: ability self-concept, self-concept) refers to students' self-perceived competencies in different school subjects which are cognitive in nature. Within the framework of the well-elaborated expectancy-value theory (e.g., Eccles et al., 1983; Wigfield & Eccles, 2000), competence self-concepts reflect the expectancy component. Correlations with scholastic achievement of moderate to high magnitude were repeatedly reported ($.30 \leq r \leq .60$; Guay et al., 2003; Hattie, 2009; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Valentine,

DuBois, & Cooper, 2004). More subjective and intrinsic motivational-affective reactions are considered to be part of the affect component (Marsh, Craven, & Debus, 1999; Möller et al., 2009). The term "affect self-concept" is based on the tradition of self-concept literature dealing with its formation, structure, and assessment (e.g., Arens, Yeung et al., 2011; Marsh et al., 1999). However, affect self-concept items such as "I am interested in mathematics" from the well-established Self Description Questionnaire I (SDQ I; Marsh, 1992) can be strongly associated with interest. Some researchers even used these affect self-concept items of the SDQ I to measure interest (e.g., Schroeders, Schipolowski, Zettler, Golle, & Wilhelm, 2016). Affect self-concepts can be considered to be part of the value component within the expectancy-value model. Relations with scholastic achievement typically are of weak to medium magnitude ($.20 \leq r \leq .30$; Jansen, Lüdtke, & Schroeders, 2016; Nagy et al., 2006; Schiefele, Krapp, & Winteler, 1992) and of high magnitude with achievement-related choices or efforts (e.g., for course choice in upper secondary school: $.55 \leq r \leq .61$, Nagy et al., 2006; also see Eccles & Wigfield, 2002; Krapp, 2000; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014). Both self-concept components strongly correlate within one school subject (Eccles & Wigfield, 1995; Kadir, Yeung, & Diallo, 2017; Rost et al., 2007; Skaalvik & Valås, 1999; Trautwein et al., 2012).

Older studies examining academic self-concepts assessed by the SDQ did not distinguish between both self-concept components due to building parcels with one competence self-concept and one affect self-concept item each to (among others) reduce the effect of item wording on factor loadings (e.g. Marsh, 1986; Marsh, Craven, & Debus, 1991; Marsh, Smith, & Barnes, 1983). In more recent studies with secondary and high school students, the distinction between subject-specific competence and corresponding affect self-concepts was empirically supported (e.g., Arens & Hasselhorn, 2015; Arens, Trautwein, & Hasselhorn, 2011; Arens, Yeung et al., 2011; Arens, Yeung, & Hasselhorn, 2014; Marsh & Ayotte, 2003; Marsh et al., 1999; Möller et al., 2009). Competence and affect self-concepts emerged as separate factors in exploratory (EFA) and confirmatory (CFA) factor analyses. Both factors were substantially correlated, but formed distinct components. Reasons for the very close relation between competence and affect self-concepts within subjects might be that (a) the competence and affect component are both facets of the underlying construct of motivation, (b) both components are reciprocally related, and/or (c) both components are based on self-reported data and, therefore, share method variance. The results of a cross-lagged panel study with seventh graders by Marsh et al. (2005) gave strong hints for a reciprocal connectedness of competence and affect self-concept. Accordingly, students come to value the things they think they are good at across the years and students think they are good at in the things they value.

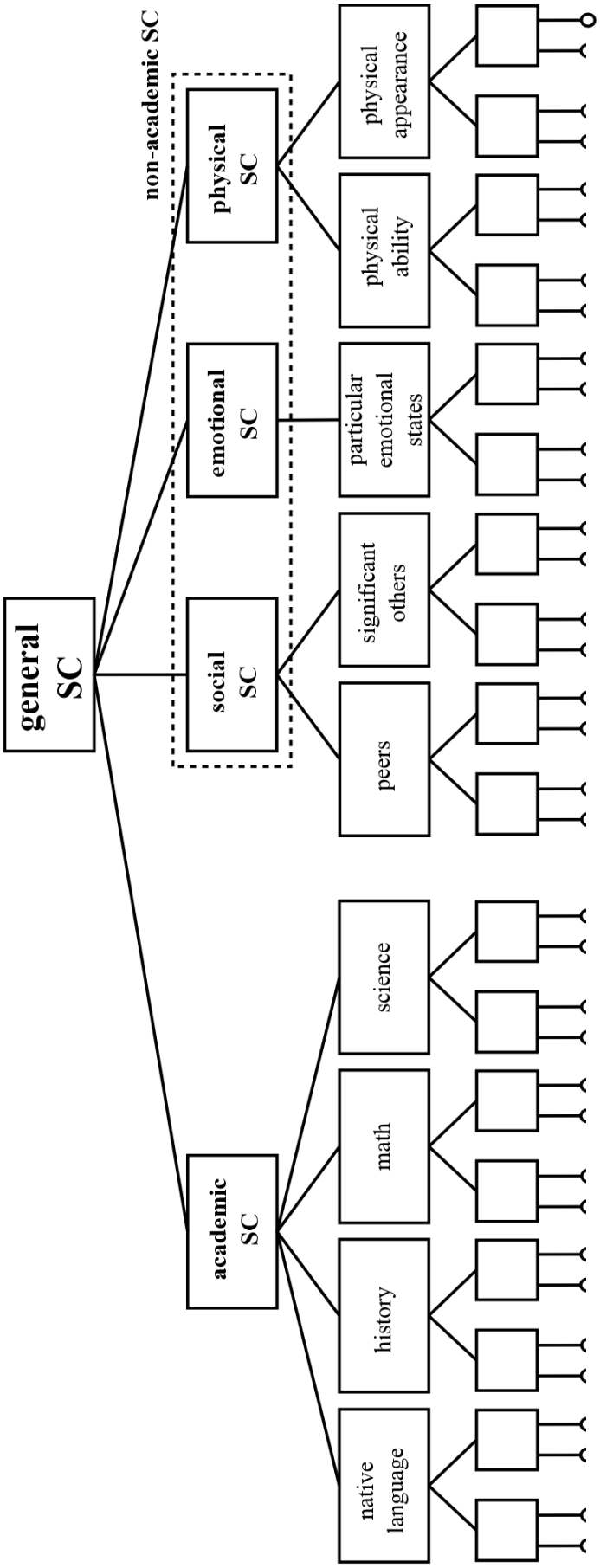


Figure 1. The multidimensional and hierarchical self-concept model of Shavelson et al. (1976; p. 413). SC = self-concept.

2.2.2 Academic self-concepts in elementary school

Elementary school is a crucial period for the development of positive academic self-concepts (e.g., Chapman & Tunmer, 1997; Marsh & Craven, 1997). On average, academic self-concepts in younger elementary school students are more positive and less realistic compared to self-concepts in older elementary school students (Helmke, 1999; Weidinger, Spinath, & Steinmayr, 2015). It is presumed that especially younger elementary school students have difficulties in distinguishing between actual and desired school-related attributes (e.g., high reading competence) and in integrating information from comparison processes with classmates into their subject-specific self-concepts, thus leading to very positive, over-optimistic academic self-concepts (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Harter, 1999; Poloczek, Karst, Praetorius, & Lipowsky, 2011). The more realistic self-concepts in older (elementary school) students are considered to be the result of higher cognitive abilities, a presumably higher amount of comparison processes of scholastic achievements with other students, and internalized evaluative standards of others (Guay et al., 2003).

Concerning self-concept correlations across domains, the correlations between mathematics and verbal self-concepts were often numerically higher for younger elementary school students at the beginning of school compared with those reported for older students, although still of rather small magnitude. For example, Ehm, Lindberg et al. (2014) reported latent correlations of mathematics and reading self-concepts of r = .20/.13/.03 for grades 1/2/3 (see also Arens, Trautwein et al., 2011; Arens, Yeung et al., 2011; Möller et al., 2009). Domain-specific self-concepts seem to be more highly correlated in early elementary school students because students of that age might be convinced that a person cannot simultaneously be good in one scholastic domain and bad in another coming along with a rather undifferentiated self-concept structure in younger students. Again, the progression in cognitive development, increasing scholastic experiences, and the onset/increase of comparison processes with increasing age should promote the differentiation of domain-specific self-concepts (Harter, 2006). This age-correlated developmental process is characterized by decreasing correlations between domain-specific self-concepts, increasing reliabilities of domain-specific (and also subject-specific) self-concept inventories, increasing model fit statistics when assuming distinct self-concept components instead of a global academic self-concept factor, and increasing correlations of domain-specific (as well as subject-specific) self-concepts with corresponding achievement measures (Guay et al., 2003; Marsh, 1986; Marsh et al., 1999; Möller et al., 2009).

As mentioned, each subject-specific self-concept can be further subdivided into a competence component and an affect component. In more recent studies, the distinction between subject-specific competence and corresponding affect self-concepts was

empirically supported for elementary school students (e.g., Arens & Hasselhorn, 2015; Arens, Trautwein et al., 2011; Arens, Yeung et al., 2011; Marsh & Ayotte, 2003; Möller et al., 2009). Within each school subject, these two self-concept components were highly correlated in all examined elementary school grades ($.71 \leq r \leq .83$; Arens & Hasselhorn, 2015; Arens, Trautwein et al., 2011; Marsh et al., 1999). Correlations of competence or affect self-concepts across domains varied substantially between studies. For example, Arens and Hasselhorn (2015) reported non-substantial correlations between competence self-concepts in mathematics and German ($r = .02$) as well as between affect self-concepts in both subjects ($r = .15$) for third graders. In the study by Marsh et al. (1999), correlations between competence or affect components in mathematics and reading were of medium to high magnitude in second to fourth graders (competence self-concept: $.38 \leq r \leq .66$; affect self-concept: $.32 \leq r \leq .55$). Nevertheless, these results further emphasized the domain specificity of both self-concept components.

Concerning grade-related differences, the competence self-concept–affect self-concept correlation within subjects does not seem to vary substantially or systematically across elementary school grades. Correlations of competence/affect self-concepts across domains seem to be of smaller magnitude in higher grades (for example, grades 2/3/4 competence self-concept in mathematics and reading: .66/.54/.38; affect self-concept: .55/.49/.32; Marsh et al., 1999). However, further research regarding the competence–affect self-concept distinction in elementary school students is needed, especially with younger elementary school students.

2.2.3 Assessment of academic self-concepts

The rating scale format is considered to be the most common response format in self-report questionnaires (Wetzel & Greiff, 2018). One issue that needs to be considered when diciding which type of rating scale researchers want to use in their questionnaires is the number of response categories. Independent of the examined construct, there is no definitive agreement among researchers about the number of response categories that optimizes the psychometric properties of a Likert-type rating scale (e.g., Lozano, García-Cueto, & Muñiz, 2008; Preston & Colman, 2000; Revilla, Saris, & Krosnick, 2014; Weng, 2004). Unfortunately, comprehensive studies relating the number of response categories with their psychometric properties are very rare. The need for such studies becomes particularly obvious if one looks at the issue of increased measurement errors when applying rating scales to students that require a finer discrimination than these students typically can easily accomplish (due to more response categories, students might not be able to distinguish reliably between adjacent categories; Adelson & McCoach, 2010; Lozano et al., 2008). Too few response categories, on the other hand, might mask individual differences (more information can be obtained by permitting greater differentiation between individual responses).

In research, questionnaires with three- to five-point Likert-type scales are typically used when applying rating scales to children (Mellor & Moore, 2013). However, such non-dichotomous scales with more than two response categories seem to be prone to response biases (see, e.g., Wetzel & Greiff, 2018): In prior studies, first and second graders mostly ticked off extreme response categories; middle categories were only seldom used (Baldering, 1993, Chambers & Johnston, 2002). In contrast, third and fourth graders seem to use extreme response categories less often and thus, tick off middle categories like *a little* or *somewhat* more often (Baldering, 1993, Chambers & Johnston, 2002). Baldering (1993) assumed that children in elementary school mainly remembered the last situation that is relevant for answering/responding to questions/items and, therefore, are prone to use unambiguous, less differentiated categories. Furthermore, these results seem to support the claim for a dichotomous response format in students at the beginning of elementary school. Reasons for response biases include less pronounced cognitive abilities and memory spans in young elementary school students compared to those in older students (Schneider & Pressley, 1989; Seitz & Rausche, 2004).

However, Baldering (1993) considered children at elementary school age not to be overtaxed by rating scales with more than two response categories in principle. In accordance with this assumption, satisfying psychometric properties were reported repeatedly with instruments assessing self-concepts. For example, Marsh et al. (1991) used a five-point Likert-type rating scale to assess self-concepts in first and second graders with the SDQ I (self-concepts in mathematics and reading, a general school self-concept, physical ability, physical appearance, peer relationships, parent relationship, general self-concept). Whereas reliabilities for all self-concept scales were at least acceptable in both grades ($\alpha \leq .73$ for group administration), fit indices for an eight-factor model in accordance with the design of the SDQ I were unacceptably low (e.g., grade 1/2: TLI = .825/.838; see Little, 2013, for guidelines regarding cut-off values of fit indices).

Focusing more specifically on academic self-concepts, Poloczek et al. (2011) used a three-point Likert-type rating scale with first graders just a few weeks after school start (self-constructed instrument; figures representing the different response categories). Reliabilities for the self-concept scales in arithmetics, reading, and writing were at least acceptable (six item measures: $.74 \leq \alpha \leq .82$; for reliabilities of self-concept scales in first graders also see Marsh et al., 1991). Furthermore, fit indices were acceptable for a self-concept structure model with three intercorrelated self-concept factors. Further studies with second (Lohbeck & Möller, 2017; using the SDQ I-GS by Arens, Yeung, Craven, & Hasselhorn, 2013; four-point Likert-type scale), third (Arens & Hasselhorn, 2015; SDQ I-GS; five-point Likert-type scale), and fourth graders (Pinxten et al., 2015; three competence self-concept items of the SDQ I; five-point Likert-type scale) typically showed good reliabilities and self-concept model fit statistics indicating an adequate

processing of items with these response formats. Ehm, Lindberg et al. (2014) even successfully used a seven-point Likert-type scale (response categories were illustrated via small stick figures representing classmates) to assess competence self-concepts in mathematics, reading, and writing in first to third graders. Moreover, criterion validities (e.g., relations with additional criteria) were as expected from prior research (Arens & Hasselhorn, 2015; Arens, Yeung et al., 2011; Ehm, Lindberg et al., 2014; Lohbeck & Möller, 2017; Pinxten et al., 2015; Poloczek et al., 2011).

Furthermore, measurement invariance across response formats can be taken into account to examine whether a measurement instrument is operating in a comparable manner for each response format under investigation (e.g., equivalence with respect to the factor pattern, factor loadings, correlations among manifest variables and/or factors, or errors of measurement; see Byrne, 1996; Christ & Schlüter, 2012; Geiser, 2013). Configural measurement invariance as the weakest form of measurement invariance only requires the factor structure (number of factors and number of factor loadings) to be constant across response formats; no specific parameters have to be set equal across response formats. For metric factorial invariance, the factor structure and factor loadings need to be the same across response formats for all indicators. This would allow a comparable interpretation of factor contents across response formats. The next level, scalar factorial invariance, requires the factor structure, factor loadings, and intercepts of the indicators to be response format-invariant. Scalar measurement invariance would indicate comparable relations between items and their corresponding characteristic to be measured. Thus, factor means could be compared meaningfully. Strict factorial invariance is satisfied when, in addition to factor loadings and intercepts, indicator residual variances are constant across response formats. Thus, indicators would have comparable reliabilities across response formats. Additionally, stricter forms of invariance can be tested (e.g., Marsh et al., 2009; Widaman & Reise, 1997), but are of rather little importance in applied research. Concerning invariance for structure models across response formats with different numbers of response categories, metric factorial invariance should be accomplished (response format-equivalent interpretation of factor content). Testing scalar factorial invariance or even stricter forms of measurement invariance and, therefore, setting (at least) item intercepts equal across response formats would not be appropriate with non-adjusted data due to different response category ranges (e.g., 1 to 3 vs. 1 to 4 vs. 1 to 5 as response categories). For example, a factor mean of $M = 2.25$ on a three-point Likert-type scale would mean something different than a factor mean of $M = 2.25$ on a five-point Likert-type rating scale.

To the best of my knowledge, there only seems to be one study that considered the issue of measurement invariance across response formats with regard to the number of response categories with elementary school students, yet. Adelson and McCoach (2010)

compared how students in grades 3 to 6 responded to a mathematics attitudes instrument with a four-point or five-point Likert-type scale showing that both response formats yielded an equal factorial structure (e.g., comparable factor loadings across response formats). Unfortunately, as the analyses were only run for the total sample, the study does not allow drawing conclusions about measurement invariance across response formats separately for each grade.

Summing up prior research, studies regarding academic self-concepts indicated an adequate use of rating scales with at least three response categories. However, there still is an important, unresolved issue surrounding the appropriate number of response categories when assessing academic self-concept in elementary school students. Thus, it appeared to be an interesting research aim to systematically compare psychometric properties (e.g., reliabilities) and to inspect measurement invariance across response formats – separately for each elementary school grade.

2.3 Interplay of scholastic achievements and academic self-concepts

The main aim of this thesis was to examine the relations between scholastic achievements and academic self-concepts – separately for elementary school grades and across grades. The following subsection first provides insights about bi- and unidirectional relations between subject-specific scholastic achievements, competence self-concepts, and affect self-concepts. Second, the effects of scholastic achievements on competence and affect self-concepts within the I/E model framework are addressed. Third, the focus is shifted towards the statistical prediction of reported grades by their corresponding competence and affect self-concepts as important motivational variables as well as intelligence as one of the most important cognitive variables.

2.3.1 Bi- and unidirectional relations between scholastic achievements and academic self-concepts

As indicated by prior study results, scholastic achievements and the two self-concept components (competence–affect distinction; see subsection 2.2.1) are substantially correlated within one school subject: Correlations between scholastic achievements and competence self-concepts within subjects are usually of moderate to high magnitude ($.30 \leq r \leq .77$; Arens, Trautwein et al., 2011; Chen et al., 2013; Guay et al., 2003; Helmke, 1997a; Kadir et al., 2017; Möller et al., 2009; Schilling, Sparfeldt, & Rost, 2006; Skaalvik & Valås, 2001; Spinath et al., 2006; Valentine et al., 2004). For affect self-concepts, typically weak to moderate relations with scholastic achievement occur ($.20 \leq r \leq .57$; Jansen et al., 2016; Kadir et al., 2017; Nagy et al., 2006; Schiefele et al., 1992). This correlational pattern is in line with expectancy-value theory: whereas subject-specific competence self-concepts seem to be more closely related to corresponding scholastic

achievements, affect self-concepts are more closely related to corresponding achievement-related choices or efforts (Arens & Hasselhorn, 2015; Eccles et al., 1983; Wigfield & Eccles, 2000). For both self-concept components, correlations between subject-specific self-concepts and corresponding indicators of subject-specific scholastic learning as well as achievement-related behavior were of higher magnitude compared to the correlations between global measures. Additionally, convergent and divergent relation patterns with subject-specific scholastic achievement measures were reported repeatedly underlining the subject-specificity of the self-concept components (e.g., Arens, Trautwein et al., 2011; Marsh & Craven, 2006; Marsh et al., 2005, 2014; Marsh & O'Mara, 2008; Rost & Sparfeldt, 2002; Skaalvik & Skaalvik, 2002).

Specifically focusing on elementary school students, correlations between scholastic achievements and corresponding competence self-concepts were of medium magnitude ($.25 \leq r \leq .50$; Helmke, 1997a; Köller, Zeinz, & Trautwein, 2008; Möller et al., 2009; Pinxten et al., 2015; Skaalvik & Valås, 2001; Spinath et al., 2006; von Maurice et al., 2014). Concerning relations between scholastic achievements and affect self-concepts (or closely associated constructs like interests or intrinsic motivation), correlation coefficients were non-substantial or of (mostly) small magnitude ($-.06 \leq r \leq .35$; Helmke, 1997a; Weidinger et al., 2015; Weidinger, Steinmayr & Spinath, 2017; von Maurice et al., 2014).

Consistent with the proposal that students' academic self-concepts become more realistic across elementary school grades (see subection 2.2.2), correlations between reported grades and both self-concept components typically are of higher magnitude in higher elementary school grades (Helmke, 1997a; Weidinger et al., 2015). For example, numerically increasing manifest correlations between reported grades and competence self-concepts were reported for second/third/fourth graders for mathematics (r = .35/.40/.52) and for the native language (r = .37/.41/.50; Helmke, 1997a). Considering eagerness to learn as an indicator of affect self-concept, manifest correlations with reported grades were numerically smaller compared to reported grade–competence self-concept correlations, but also numerically increasing for mathematics (r = .26/.32/.35) and the native language (r = .19/.31/.33).

Regarding unidirectional relations between scholastic achievements and corresponding academic self-concepts, there is an extensive literature investigating their causal ordering (e.g., Arens et al., 2017; Baumert, Schnabel, & Lehrke, 1998; Guay et al., 2003; Helmke & van Aken, 1995; Marsh & Craven, 2006; Marsh et al., 2005; Marsh & Yeung, 1997; Möller, Retelsdorf, Köller, & Marsh, 2011; Niepel et al., 2014; Valentine et al., 2004). These studies have almost consistently shown that academic self-concepts are influenced by prior scholastic achievements and that prior academic self-concepts have independent influences on later scholastic achievements within the same school subject.

Better scholastic achievements in a subject might bolster students' confidence in their scholastic competencies and their intrinsic affect while working in that subject. For example, better results can foster perceptions of self-efficacy and promote the development of corresponding interests (Köller, Baumert, & Schnabel, 2001; Lent, Brown, & Hackett, 1994). Consecutively, students with high competence self-concepts might be more likely to approach new tasks with confidence. Higher motivational-affective self-perceptions might lead to scholastic engagement (e.g., persistent engagement in tasks, practice) which again might lead to learning gains and, therefore, enhanced scholastic achievements (Spinath et al., 2006; Wigfield & Karpathian, 1991).

For elementary school students, the few attempts to specify causal models for scholastic achievements and corresponding competence self-concepts have yielded mixed results. Some studies only found significant path coefficients from prior scholastic achievements on later competence self-concepts but mostly no significant path coefficients from prior competence self-concepts on later scholastic achievements (Chapman & Tunmer, 1997; Helmke & van Aken, 1995; Skaalvik & Havget, 1990; van Aken, Helmke, & Schneider, 1997; Weidinger et al., 2015). However, three studies pointed towards a reciprocal relationship: Substantial reciprocal effects were reported in second to fourth graders using more global scholastic achievement measures (teacher ratings for students' scholastic achievements in reading, writing, and mathematics) and global academic self-concept measures (Guay et al., 2003). A comparable result pattern was found in fourth/fifth graders (scholastic achievement indicated by reported grades and a global academic self-concept; Muijs, 1997) as well as in second to fourth graders in a cross-leged model for mathematics (scholastic achievement was indicated by reported grades or test scores; Helmke & van Aken, 1995).

For scholastic achievement and corresponding affect self-concepts (or closely related constructs), there is a very limited number of studies investigating their causal ordering in elementary school students. Von Maurice et al. (2014) found that scholastic achievement in mathematics and German determined the level of subsequent affect self-concepts but not vice versa in third and fourth graders. In contrast, results by Weidinger et al. (2015) found no effects from prior reported grades/intrinsic motivation in German on later corresponding intrinsic motivation/reported grades in third and fourth graders. Skaalvik and Valås (1999) examined the causal ordering of scholastic achievement (tests, teacher ratings) in mathematics and Norwegian and corresponding "motivation" (combination of interest and investment items; e.g., "I like mathematics" or "I give up quickly if I get a difficult math task"). No separate analyses for interest as an indicator of affect self-concept were run. However, no reciprocal effects were reported in this study with third/fourth graders.

2.3.2 Internal/external-frame-of-reference model

Within the multidimensional hierarchical self-concept model by Shavelson et al. (1976; see subsection 2.2.1), subject-specific self-concepts are highly correlated and form a higher-order academic self-concept. As mentioned, factor analyses repeatedly failed to find the assumed single second-level academic self-concept factor and the postulated substantial positive correlation between mathematics and verbal self-concepts. Regarding the latter, small to negligible correlations between mathematics and verbal self-concept factors ($-.10 \leq r \leq .12$; Marsh, 1986) were reported instead. This was particularly surprising in the light of high correlations between respective achievement measures in mathematics and verbal domains (approximately .50 to .80). Theoretically, one might have expected self-concept correlations that are of comparable magnitude as the correlations between scholastic achievements because of substantial correlations between domain-specific achievement measures and corresponding self-concepts. Pursuant to these findings, Marsh (1986b) developed the I/E model. Within the I/E model framework, effects of scholastic achievements on academic self-concepts are associated with two psychological processes: social and dimensional comparisons. The formation of domain-specific academic self-concepts is explained by the joint operation and especially the relative weight of social and dimensional comparison processes.

According to the classical social comparison theory (Festinger, 1954), individuals form self-appraisals by comparing their abilities or opinions with some standard or frame of reference. Furthermore, Festinger (1954) hypothesized that people evaluate their abilities or opinions by comparison processes with other people if objective, nonsocial means are not available. Within the I/E model framework, it is assumed that students compare their scholastic achievements in one subject, for example, mathematics, with the achievements of their peers in the same subject (external comparison) to construct evaluations of their own abilities (e.g., "I am good at mathematics"). For example, a student with high scholastic achievements in mathematics should have a higher self-concept in mathematics compared to students with worse achievements in mathematics. Within the I/E model, social comparison processes should correspond with positive path coefficients of scholastic achievements on corresponding academic self-concepts (see Fig. 2; solid lines).

Dimensional comparisons refer to the contrast of one's academic achievement in one domain with one's own achievement in another domain (internal comparison; e.g., Marsh et al., 2014; Marsh et al., 2015; Möller & Marsh, 2013; Pohlmann, 2005; Pohlmann & Möller, 2009; Rost, Sparfeldt, Dickhäuser, & Schilling, 2005). To evaluate his/her own scholastic achievements in, for instance, mathematics, the student is likely to compare his/her performances in mathematics to own performances in another subject (e.g., the native language). If students achieve differently in different subjects (e.g.,

mathematics vs. native language), they tend to lower their academic self-concept in the worse-off subject while simultaneously raising their academic self-concept in the better-off subject. Within the I/E model, dimensional comparison processes should correspond with negative path coefficients of scholastic achievements on non-corresponding academic self-concepts (see Fig. 2; dotted lines).

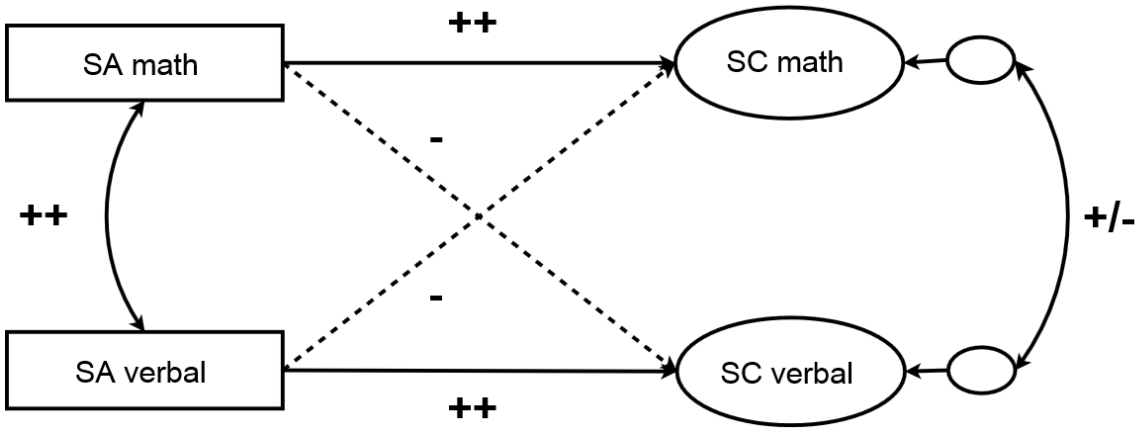


Figure 2. Internal/external-frame-of-reference model (I/E model; also see Marsh, 1986, p.134). SA = scholastic achievement, SC = self-concept, unidirectional solid line = social comparison processes, unidirectional dotted line = dimensional comparison processes.

The joint operation and especially the relative weight of both comparison processes should influence the magnitude of the resulting correlation between academic self-concepts and, therefore, should play important roles in the formation of domain-specific academic self-concepts (Marsh, 1986). The I/E model does not require the correlation between domain-specific self-concepts to be zero, but the correlation should be of substantially smaller magnitude compared with the correlation of respective scholastic achievements.

Focusing on self-concepts in mathematics and the native language within the I/E model framework, the relation pattern outlined above has been proven to be robust for different gender groups, countries, and methodologies (e.g., Möller et al., 2009; Marsh & Hau, 2004; Möller & Marsh, 2013). For example, a high academic self-concept in mathematics is expected when a student's scholastic achievements in mathematics are good compared to those of his/her classmates (social comparison processes) and high compared to the own achievements in the native language (dimensional comparison processes). Holding the scholastic achievement in mathematics constant, then, it is the

difference between both achievements that is predictive for the self-concept in mathematics. High scholastic achievements in the native language serve only to weaken a high self-concept in mathematics.

For elementary school students of all grades, substantial positive path coefficients from scholastic achievements on corresponding academic self-concepts indicating social comparison processes were repeatedly found (Ehm, Lindberg et al., 2014; Ehm, Nagler, Lindberg, & Hasselhorn, 2014; Faber, 1992; Lohbeck & Möller, 2017; Marsh, Smith, Barnes, & Butler, 1983; Möller, Kuska, & Zaunbauer, 2011; Pinxten et al., 2015; Poloczek et al., 2011; Skaalvik & Valås, 2001). In contrast, substantially negative path coefficients from scholastic achievements on non-corresponding self-concepts indicating dimensional comparison processes were found only in third and fourth grade. These substantial path coefficients were of small or medium magnitude ($-.50 \leq \beta \leq -.15$; Ehm, Lindberg et al., 2014; Ehm, Nagler et al., 2014; Faber, 1992; Marsh, Smith, Barnes, & Butler, 1983; Möller, Kuska et al., 2011; Pinxten et al., 2015; Schmidt et al., 2017; Skaalvik & Valås, 2001).

Regarding grade-related differences, social comparison processes seem to increase during elementary school years; dimensional comparison processes seem to occur earliest in third grade. Theoretically, the increasing impact of social and dimensional comparison processes in elementary school is explained by increasing experiences with scholastic achievements as well as by increasing cognitive abilities (see Guay et al., 2003; Harter, 2006; Skaalvik & Skaalvik, 2002). Increasing scholastic experiences and, therefore, increasing experiences with achievement feedback (e.g., reported grades) might go in hand with a larger number of possible social comparison processes with classmates; stable class rank orders could be established in a reference group. Furthermore, increasing experiences with achievement feedback might also lead to a larger number of cognitively more demanding dimensional comparison processes. According to Skaalvik and Skaalvik (2002), students might compare their relative rank position in two school subjects independent of each other more easily (social comparison processes; e.g., "I am one of the worst students in class in mathematics" and "I am the fourth best student in class in German"). Subsequently, the comparison of relative rank positions across subjects seems to be more difficult for elementary school students resulting in the non-substantial path coefficients that indicate dimensional comparison processes in younger elementary school students.

To the best of our knowledge, there is only one study focusing on elementary school students from different grades. Ehm, Lindberg et al. (2014) analyzed the I/E model separately for first, second, and third graders for mathematics, reading, and writing. Concerning comparison processes for scholastic achievements in mathematics and reading as a prominent learning activity of native language education in elementary

school (e.g., Arens et al., 2014; Byrne, 1996; also see subsection 4.3.2 for the discussion of instruments), all path coefficients indicating social comparison processes were substantial in all grades. Furthermore, these path coefficients were of substantially higher magnitude in higher grades (grade 1/2/3 mathematics: β = .30/.50/.58; reading: β = .27/.41/.49). Path coefficients indicating dimensional comparison processes were non-substantial in first and second grade ($\beta \leq .09$), but substantial in third grade (scholastic achievement in mathematics/reading on self-concepts in reading/mathematics: β = -.18/-.17). No grade-related differences occurred for dimensional comparison processes. The correlation between corresponding self-concept residuals was of smaller magnitude in higher grades (r = .30/.18/.15). However, the interpretation of results is limited by not taking school class affiliation and the corresponding clustering of the data (hierarchical data) into account. Moreover, the authors used different indicators of scholastic achievements (different achievement tests) across grades, thereby limiting the interpretation of comparisons of the coefficients across grades. Furthermore, comparisons of path coefficients across studies are problematic because grade-related differences in social and dimensional comparison processes on the one hand and differences of the operationalizations and constructs on the other hand could be confounded. Therefore, it was one aim of this dissertation project to assess scholastic achievements and academic self-concepts with the same instruments in all elementary grades to allow more straightforward interpretations of possible grade-related differences.

Moreover, prior I/E model-related studies with elementary school students did not take the competence–affect self-concept distinction (see subsections 2.2.1 and 2.2.2) into account. Typically, reported results refer to relations between scholastic achievements and (non-)corresponding competence self-concepts. However, the I/E model pattern with scholastic achievements and competence self-concepts should also be applicable to scholastic achievements and affect self-concepts due to the high correlation between both self-concept components within each subject (Arens & Hasselhorn, 2015; Arens, Trautwein et al., 2011; Marsh et al., 1999). Furthermore, recent research regarding a generalized internal/external-frame-of-reference model (GI/E model; Möller, Müller-Kalthoff, Helm, Nagy, & Marsh, 2016) extended the predictions of the traditional I/E model to encompass more domains on the levels of the predictors and the criteria (e.g., interest or self-regulation). So far, there is no study examining relations between scholastic achievements and affect self-concepts in the I/E model framework with elementary school students. With secondary school students, substantial path coefficients indicating social comparison processes were reported (interest or academic enjoyment as equivalent or closely related constructs to affect self-concepts; $.15 \leq \beta \leq .50$), but only some path coefficients indicating dimensional comparison processes were significantly negative ($-.25 \leq \beta \leq -.07$; Goetz, Frenzel, Hall, & Pekrun, 2008; Kadir et al., 2017;

Pohlmann, 2005; Schurtz, Pfost, Nagengast, & Artelt, 2014). These numerically smaller relations between scholastic achievements and affect self-concepts compared to the scholastic achievement–competence self-concept relations in other studies are in line with assumptions regarding expectancy-value theory (Eccles & Wigfield, 2002). An assignment of the competence–affect distinction to the I/E model framework would allow the examination of possibly differential relations between scholastic achievement and the two correlated, but still distinct self-concept components and therefore, deepen the understanding of social and dimensional comparisons on both self-concept components.

2.3.3 Statistical prediction of scholastic achievements

The question regarding determinants of scholastic achievement is still one of the key questions in the field of teaching and scholastic learning. Hence, various models for the prediction of scholastic achievement were postulated (for an overview see Hasselhorn & Gold, 2009; Helmke & Weinert, 1997). Consistently, it is assumed in all models that scholastic achievement is a heterogeneous, multi-determined measure. In general, determinants of scholastic achievement can be classified into two groups: determinants that are inherent to the student him-/herself (individual features of the student) and context factors (e.g., domestic or learning environments). Of the two, individual features of the students have frequently been found to be among the most relevant predictors of scholastic achievements (Steinmayr et al., 2014). Among these, student's scholastic achievements in a specific subject rely on cognitive factors like intelligence or prior knowledge and are additionally influenced by non-ability aspects such as motivation, volition, or effort (e.g., Ditton & Krüsken, 2009a; Harlen, 2005; Steinmayr et al., 2014; Willingham et al., 2002).

There is ample evidence that intelligence is one of the best psychological predictors of scholastic achievement (e.g., Jensen, 1998; Helmke & Weinert, 1997; Kuncel, Hezlett, & Ones, 2004; Mackintosh, 2011; Rost, 2013). Referring to reported grades as indicators of scholastic achievements, a meta-analysis revealed an average validity coefficient of $\rho = .54$ between intelligence and reported grades (Roth et al., 2015). For elementary school students, the relation was of slightly lower magnitude, $\rho = .45$. The close relation between intelligence and reported grades might be based on the high intelligence demands of scholastic learning and scholastic achievement itself. For example, conceptual overlaps between intelligence and reported grades become evident in the need for grasping concepts and meanings, learning to deal with novel material, making distinctions, recognizing patterns, or transferring knowledge and skills to new situations/problems (Jensen, 1998). However, a substantial amount of reported grade variance remains unexplained by intelligence. Academic self-concept as an important motivational variable appears to be an additional viable predictor. According to Craven and Marsh (2008) „a positive self-belief is valued as a 'hot' variable that makes good

things happen, and is fundamental to the realization of full human potential in a range of settings“ (p. 107). In educational and psychological research, competence and affect self-concepts are extensively investigated due to their importance for scholastic learning and achievement-related behavior (e.g., Guay, Ratelle, Roy, & Litalien, 2010; Helmke, 1999; Marsh, 2007; Möller et al., 2009) and, therefore, for scholastic achievements. Within prediction models of scholastic achievements, they are basic components that substantially predict scholastic achievements in the same subject beyond other variables like intelligence (e.g., Helmke, 1997b; Jansen et al., 2016; Lotz et al., 2018; Schicke & Fagan, 1994; Spinath et al., 2006; Steinmayr & Meißner, 2013; Weber, Lu, Shi, & Spinath, 2013).

Most prediction studies with elementary school students did not focus on the prediction of scholastic achievement by separate predictors, but rather examined the link between conglomerates of various cognitive or non-cognitive variables and scholastic achievement (Helmke, 1997b; Schicke & Fagan, 1994; Weber et al., 2013). Only the study by Spinath et al. (2006) examined the prediction of scholastic achievement (composite achievement scores) by intelligence, competence self-concept, and affect self-concept (referred to as interest) as separate predictors in fourth grade students. In prediction models with solely intelligence and competence self-concept or models with solely intelligence and affect self-concept (interest), both self-concept components were substantial predictors of scholastic achievement in mathematics and English as the native language beyond intelligence. Subsequent commonality analyses for a model with intelligence and competence self-concept as predictors of scholastic achievement revealed that the uniquely explained variance was numerically higher for intelligence (mathematics/English: R^2 = .17/.13) in comparison to competence self-concept (mathematics/English: R^2 = .08/.09). A total amount of R^2 = .32/.29 reported grade variance in mathematics/English was explained by both predictors, common variances accounted for R^2 = .07/.07. Commonality analyses for a model with intelligence and affect self-concept (interest) as predictors of scholastic achievement showed a comparable result pattern: the uniquely explained variance was numerically higher for intelligence (mathematics/English: R^2 = .23/.18) in comparison to affect self-concept (interest; mathematics/English: R^2 = .04/.03). A total amount of R^2 = .29/.23 reported grade variance in mathematics/English was explained by both predictors, common variances accounted for R^2 = .02/.02. Again, the numerically smaller relations between scholastic achievements and affect self-concepts compared to the scholastic achievement–competence self-concept relations (above and beyond intelligence) are in line with the assumptions regarding expectancy-value theory (Eccles & Wigfield, 2002).

Considering all three predictors – intelligence, competence self-concept, and affect self-concept – simultaneously, the predictive power of affect self-concept (interest)

vanished in mathematics and English (Spinath et al., 2006). This drop of the formerly substantial affect self-concept (interest)–scholastic achievement coefficients was probably caused by controlling for the substantial overlap between competence and affect self-concepts in the same subject (mathematics/English: r = .74/.56). Because commonality analyses by Spinath et al. (2006) were only based on two predictors (intelligence and one academic self-concept component), the analysis of specific and common variances within a simultaneous prediction of scholastic achievement by intelligence, competence self-concept, and affect self-concept still calls for further research.

Furthermore, all studies examining the prediction of scholastic achievement were conducted with fourth graders. Thus, it was one research aim to extend prior research to lower elementary school grades and to examine grade-related differences in the prediction of scholastic achievement. So far, assumptions about grade-related differences can only be derived on the basis of correlation studies. Whereas the correlation between intelligence and reported grades as the main indicator of scholastic achievement did not seem to differ between grades (e.g., Laidra, Pullmann, & Allik, 2007), subject-specific academic self-concepts showed an increasing relation to corresponding reported grades (Helmke, 1997a; Skaalvik & Valås, 1999; Weidinger et al., 2015; also see subsection 2.3.1). Subsequently, the predictive power of intelligence might remain stable across elementary school grades; path coefficients from both self-concept components on reported grades might be of higher magnitude in higher grades.

2.4 Research aims of this thesis

Based on the theoretical framework and empirical findings summarized in the previous sections, three articles and one excursus constitute the core of this thesis. The overarching research aims were as follows:

First, we aimed to examine the accuracy of elementary school students' self-reported grades in comparison to their corresponding teacher-reported (actual) grades. Second, it was one research aim to systematically compare psychometric properties (e.g., reliabilities) across competence and affect self-concept scales assessed with three-, four-, or five-point Likert-type scales and to inspect measurement invariance across these response formats separately for each elementary school grade. Analyses regarding the assessment of reported grades as indicators of scholastic achievements and the assessment of academic self-concepts constituted the basis for further analyses on the relations of scholastic achievements and academic self-concepts in elementary school students. For these analyses, the important competence-affect distinction of academic self-concepts was taken into account. As a third research aim, we explored the effects of social and dimensional comparison processes on competence and affect self-concepts and, therefore,

on the formation of domain-specific competence and affect self-concepts within the I/E model framework. The fourth aim was to examine the differential relevance of intelligence, competence self-concepts, and affect self-concepts when predicting scholastic achievements and to inspect whether the predictors uniquely explained achievement variance beyond each other. Relations between scholastic achievements and academic self-concepts were examined separately for each elementary school grade and subsequently compared across grades to provide initial evidence for developmental processes across elementary school grades.

3 Empirical Studies

In the following chapter, the three empirical studies and the excursus this thesis is based on are outlined by shortly summing up their theoretical background and research aims, methods, results, and discussions. Important findings that were not incorporated in the corresponding articles are provided in supplementary analyses sections.

The three studies and the excursus presented in this thesis are based on a large sample of first to fourth graders from German elementary schools. Data collection took place at the end of the school year in June and July 2014 (federal states: Lower Saxony and Saxony-Anhalt) and 2015 (federal state: Saarland) in cities as well as rural areas forming a quite representative sample of German elementary school students. The data collection with elementary school students, especially first and second graders, was particularly challenging because it was time-consuming: In most classes, the class teacher claimed five to ten minutes to review organizational topics, afterwards students were randomly assigned to small groups (one group per response format for the academic self-concept items) that were tested in separate rooms. The personal introduction of the experimenter and the assessment of non-verbal intelligence (Culture Fair Intelligence Test 1–R; CFT 1–R; Weiß & Osterland, 2013) took around 30 minutes and typically lasted the rest of the first school period. To motivate students, the second period started with a short movement game. Following the assessment of all SDQ I self-concept items (Marsh, 1992; 64 items) that were read aloud twice, further self-concept items were assessed with the Self-Perception Profile for Children (SPPC; Harter, 1985; 24 items). Scholastic achievements (reported grades and reported grade equivalents) and further demographical data were assessed at the end of the second period. Additionally, class teachers were asked to report their students' actual reported grades (or reported grade equivalents) as well as to evaluate their students' academic self-concepts.

For the studies and the excursus of this thesis, only data with regard to reported grades, academic self-concepts (SDQ I), and non-verbal intelligence have been used. Intercorrelations of reported grades in mathematics and German, competence and affect self-concepts in mathematics and reading, and intelligence are depicted in Appendix A.

3.1 Study 1: Assessment of reported grades

Schneider, R., & Sparfeldt, J. R. (2016). Zur (Un-)Genauigkeit selbstberichteter Zensuren bei Grundschulkindern [The accurarcy of self-reported grades in elementary school]. *Psychologie in Erziehung und Unterricht, 63*, 48–59.

In this study, we aimed to examine the accuracy of elementary school students' self-reported grades in comparison to their corresponding teacher-reported (actual) grades in mathematics, German, and sports. The sample of this study stemmed from German elementary schools in Lower Saxony and Saxony-Anhalt. In these federal states, students receive reported grades from second grade onwards. Therefore, the sample for the following analyses only comprised elementary school students from grades 2 to 4.

3.1.1 Theoretical framework and research aims

In educational and psychological research, students' self-reported grades are often used as indicators of their actual scholastic achievements. Prior studies with German secondary school students indicated an accurate report of grades in various subjects (Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008).

So far, only one study reported percentages of correctly reported grades in fourth graders (92% for the reported grade in mathematics, 85% for reading, sport and arts; Ostrop et al., 2002). Unfortunately, no further aspects regarding the accuracy of self-reported grades were examined. For third graders, Ostrop et al. (2002) only stated that self-reported grades were less accurate in comparison to those reported by fourth graders, but did not specify any values. Therefore, a replication of these findings for fourth graders as well as an extension to further accuracy aspects (e.g., average differences between student- and teacher-reported grades or correlations with additional criteria) and further elementary school grades were research desiderata.

Because it was unclear whether the economic practice of using self-reported instead of teacher-reported (actual) grades is also appropriate for elementary school students, we exploratorily compared students self-reported grades with corresponding teacher-reported grades with regard to the accuracy aspects used in studies with older students (Dickhäuser & Plenter, 2005; Kuncel et al., 2005; Sparfeldt et al., 2008). This was done separately for second, third and fourth graders and the school subjects mathematics, German, and sports. Thereby, we focused on

(1) correlations between students' self-reported grades and corresponding teacher-reported grades,

(2) differences between student- and teacher-reported grades with regard to (a) mean differences, (b) average absolute differences as well as (c) percentages of students' correctly reported grades, under-, and over-reports.

(3) Concerning relations with additional criteria, we focused on correlations with (a) intelligence and (b) corresponding academic self-concepts in mathematics, reading, and sports.

3.1.2 Method

Sample and procedure. The sample under investigation comprised N = 424 elementary school students (grade 2/3/4 n = 86/181/157) attending 38 classes in 10 elementary schools. Data collection took place in June and July 2014 at the end of the school year during regular lessons (two school periods of 45 min each) and was conducted by trained experimenters in small groups. In the first school period, students' non-verbal intelligence was assessed. Data on students' academic self-concepts and their self-reported grades were collected in the second period. Students directly answered on the provided material. During the data assessment, corresponding class teachers were asked to report their students' actual reported grades.

Instruments. Students were asked to report their grades from their last midterm report cards in mathematics, German, and sports. Additionally, corresponding class teachers reported their students' actual reported grades for the same subjects. The German grading system consists of reported grades from *1* (very good) to *6* (insufficient), with numerically lower reported grades indicating better scholastic achievements.

Students' nonverbal intelligence was assessed with three non-verbal intelligence subtests (series, classification, and matrices) of the German Grundintelligenztest Skala 1 – Revision (CFT 1–R; Weiß & Osterland, 2013). Academic self-concepts in mathematics, reading, and sports were assessed with the corresponding eight items of the SDQ I (Marsh, 1992; German translation by Arens, Trautwein et al., 2011), for example, "I am good at mathematics" or "I like reading". Randomized within each class, subject-specific academic self-concepts were administered using either a three-, four-, or five-point Likert-type rating scale; sum scores were *z*-standardized per response format and grade.

Analyses. All analyses were conducted with SPSS 23. Separately for each elementary school grade and school subject, we calculated correlations between student- and teacher-reported grades (research aim 1). Subsequently, we compared these correlations between adjacent grades (effect size q; difference between two Fisher transformed correlation coefficients; small/medium/large effect: $q \geq 0.10/0.30/0.50$; Cohen, 1988, p. 115).

Concerning mean differences between student- and teacher-reported grades (research aim 2a), we computed 3×2 ANOVAs with a three-stage factor 'grade' and a two-stage repeated measure factor 'reporter' (students, teachers) as well as the depended variable 'reported grade' – separately for each subject. Additionally (research aim 2b), we examined the average absolute differences between students- and teacher-reported grades (teacher-reported minus students-reported grade) – separately for each subject and grade.

Regarding research aim (2c), we calculated the percentages of correctly reported grades, under-, and over-reports.

Research aim (3) was examined by calculating correlations of student- and teacher-reported grades with intelligence and corresponding subject-specific academic self-concepts per subject and grade. In addition, the effect size *q* for the correlation difference between students-and teacher-reported grades with the additional criterion was computed.

3.1.3 Results

Regarding research aim (1), correlations between student- and teacher-reported grades were substantially positive in all three school subjects and grades (grade 2/3/4 mathematics: r = .68/.80/.90; German: r = .64/.84/.86; sports: r = .49/.73/.84). Furthermore, correlations were substantially higher in higher grades (all ps < .05; -.46 ≤ q ≤ -.27), except for the comparison between third and fourth graders in German (p = .26; q = -.07).

Concerning the differences between student- and teacher-reported grades (research aim 2a), the ANOVA did neither reveal a significant main effect for ‘reporter’ (F[1,415] = 0.34, p = .56, η^2 = .001), nor a significant interaction of ‘reporter’ and ‘grade’ in mathematics (F[1,415] = 0.37, p = .55, η^2 = .001). In German, the means of students’ self-reported grades were significantly lower than the means of the teacher-reported grades (F[1,413] = 7.14, p < .05, η^2 = .017); again, no significant interaction effect occurred (F[1,413] = 2,83, p = .09, η^2 = .007). In sports, the means of students’ self-reported grades were significantly lower than the means of the teacher-reported grades (F[1,407] = 12.12, p < .05, η^2 = .029). Additionally, a significant interaction effect occurred (F[1,407] = 4.00, p < .05, η^2 = .010): Students in grade 2 and 3 substantially reported better grades compared to their teachers (F[1,253] = 17.11, p < .05, η^2 = .063). This difference between reporters was substantially higher in second compared to third graders (F[1,253] = 11.11, p < .05, η^2 = .042).

Regarding research aim (2b), average absolute differences indicated numerically lower differences between reporters in higher grades (grade 2/3/4 mathematics: $Diff_{abs}$ = 0.50/0.32/0.22; German: $Diff_{abs}$ = 0.42/0.25/0.19; sports: $Diff_{abs}$ = 0.54/0.35/0.24). Correspondingly and regarding research aim (2c), the percentage of correct reports by students was higher in higher grades (mathematics: $\%_{correct}$ = 42/66/78; German: $\%_{correct}$ = 54/73/81; sports: $\%_{correct}$ = 39/64/73), the percentage of misreports was lower in higher grades. In general, the amount of under-reports was lower compared to over-reports.

Concerning the correlations of student- and teacher-reported grades with intelligence (research aim 3a), correlations between teacher-reported grades and intelligence were mostly comparable or numerically higher than the student-reported grade–intelligence correlations (.06 ≤ q ≤ .22). Conversely, the correlations between

teacher-reported grades and corresponding subject-specific academic self-concepts (research aim 3b) were mostly comparable or numerically lower compared with student-reported grade–self-concept correlations ($-.24 \leq q \leq .07$).

3.1.4 Supplementary analyses

The results of study 1 are based on analyses with second to fourth graders from the German federal states of Lower Saxony and Saxony-Anhalt. Data were assessed in June and July 2014, analyses were conducted in summer 2014, the corresponding manuscript was submitted to the journal in December 2014. Supplementary analyses of the full sample comprising second to fourth graders from Lower Saxony, Saxony-Anhalt, and students from Saarland (data assessment in June and July 2015) are presented in this subsection.

Method. The sample under investigation comprised N = 858 second to fourth grade students (grade 2/3/4 n = 253/321/284) attending 59 classes in 16 German elementary schools. Student- and teacher-reported grades, non-verbal intelligence, and academic self-concepts were assessed as described above (subsection 3.1.2). Analyses were conducted analogously to the analyses of the 2014 sample (see above).

Results. Results were mostly comparable to the results presented above: Regarding research aim (1), correlations between student- and teacher-reported grades were substantially positive and mostly of high magnitude in the three school subjects and grades (Tab. 1). Furthermore, correlations were substantially higher in higher grades (all ps < .05; $-.49 \leq q \leq -.23$), except for the comparison between third and fourth graders in German (p = .14; q = -.09).

Regarding differences between student- and teacher-reported grades (research aim 2a), the ANOVA neither revealed a significant main effect for ‘reporter’ (F[1,627] = 0.10, p = .75, η^2 = .000), nor a significant interaction of ‘reporter’ and ‘grade’ in mathematics (F[1,627] = 1.83, p = .18, η^2 = .003). In German, the means of students’ self-reported grades were significantly lower than the means of the teacher-reported grades (F[1,647] = 13.74, p < .05, η^2 = .021). Additionally, a significant interaction effect occurred (F[1,647] = 6.10, p < .05, η^2 = .009): Students in grade 2 and 3 substantially reported better grades compared to their teachers (F[1,408] = 11.31, p < .05, η^2 = .027). This difference between reporters was substantially higher for second compared to third graders (F[1,408] = 6.48, p < .05, η^2 = .016). In sports, the means of students’ self-reported grades were significantly lower than the means of the teacher-reported grades (F[1,610] = 14.68, p < .05, η^2 = .024). Again, a significant interaction effect occurred (F[1,610] = 5.49, p < .05, η^2 = .009): Students in grade 2 and 3 substantially reported better grades compared to their teachers (F[1,368] = 18.68, p < .05, η^2 = .049). This difference between reporters was substantially higher for second compared to third graders (F[1,368] = 12.24, p < .05, η^2 = .032).

Table 1

Correlations between student-reported and teacher-reported grades, supplemented by Cramer's V-values (χ^2- difference test for correlation coefficients), degrees of freedom, and the effect size q

		grade 2	*q*	grade 3	*q*	grade 4	*V*	*df*
	mathematics	.73*	-.23	.82*	-.27	.89*	27.488*	2
r	German	.54*	-.49	.80*	-.09	.83*	47.695*	2
	sports	.38*	-.38	.71*	-.24	.81*	63.937*	2

Note. *$p < .05$

Table 2

Means (M) and standard deviations (SD) for student- and teacher-reported grades, effect sizes for mean differences (d), and average absolute differences ($Diff_{abs}$) between both raters as well as percentages of correctly reported grades (%corr), under- (%under), and over-reports (%over)

			grade 2	grade 3	grade 4
mathematics	teacher	*M* (*SD*)	1.77 (.80)	2.21 (.87)	2.45 (.92)
	student	*M* (*SD*)	1.70 (.89)	2.03 (.96)	2.24 (.88)
		d	0.08	0.20	0.23
		$Diff_{abs}$	0.13	0.15	0.21
		% over	25	19	19
		% corr	61	74	74
		% under	14	7	7
German	teacher	*M* (*SD*)	2.04 (.80)	2.12 (.83)	2.41 (.77)
	student	*M* (*SD*)	1.88 (.88)	2.07 (.86)	2.27 (.88)
		d	0.19	0.06	0.17
		$Diff_{abs}$	0.27	0.09	0.10
		% over	31	18	16
		% corr	57	72	77
		% under	12	10	7
sports	teacher	*M* (*SD*)	1.84 (.72)	1.88 (.68)	1.94 (.67)
	student	*M* (*SD*)	1.47 (.74)	1.79 (.80)	1.80 (.72)
		d	0.51	0.12	0.20
		$Diff_{abs}$	0.35	0.08	0.15
		% over	43	12	24
		% corr	47	68	77
		% under	10	20	19

Concerning research aim (2b), average absolute differences indicated numerically lower differences between reporters in higher grades (Tab. 2). Correspondingly and regarding research aim (2c), the percentage of correct reports by students was of higher magnitude in higher grades, the percentage of misreports was lower in higher grades. In general, the amount of under-reports was lower compared to over-reports.

Regarding correlations of student- and teacher-reported grades with intelligence (research aim 3a), correlations between teacher-reported grades and intelligence were mostly comparable or numerically higher than the student-reported grade–intelligence correlations (Tab. 3; $-0.03 \leq q \leq 0.17$). Conversely, the correlations between teacher-reported grades and corresponding subject-specific academic self-concepts (research aim 3b) were mostly comparable or numerically lower compared to student-reported grade–self-concept correlations ($-0.26 \leq q \leq 0.01$).

Table 3

Correlations between student-reported and teacher-reported grades with intelligence and academic self-concepts, supplemented by the effect size q

		intelligence			academic self-concept		
	reporter	student	teacher	*q*	student	teacher	*q*
grade 2	mathematics	-.31*	-.42*	0.13	-.58*	-.38*	-0.26
	German	-.21*	-.36*	0.16	-.29*	-.23*	-0.06
	sports	-.08	-.24*	0.17	-.51*	-.25*	-0.31
grade 3	mathematics	-.34*	-.44*	0.12	-.44*	-.44*	0.00
	German	-.30*	-.35*	0.06	-.31*	-.32*	0.01
	sports	-.14*	-.11	-0.03	-.50*	-.41*	-0.11
grade 4	mathematics	-.42*	-.47*	0.06	-.53*	-.45*	-0.11
	German	-.26*	-.28*	0.02	-.44*	-.33*	-0.13
	sports	-.04*	-.16*	0.12	-.58*	-.49*	-0.13

Note. $*p < .05$.

3.1.5 Discussion

The results of study 1 indicated that students' self-reported grades were less accurate in grade 2 and 3 compared to those reported by fourth graders. In all examined subjects, correlations of student- and teacher-reported grades were of medium magnitude in grade 2 and of high magnitude in grades 3 and 4. Across grades, correlations were of higher magnitude in higher grades. The means of student-reported grades were lower than the means of the corresponding teacher-reported grades reflecting overestimations in all

grades. Differences between data sources (teacher, students) were lower in higher grades; correspondingly, the percentage of correctly reported grades was of higher magnitude in higher grades. Results also pointed to a bias in self-reported grades towards students' corresponding self-concepts.

Due to the design of the study, we were not able to investigate why students did not report their actual reported grades accurately. Possible reasons might be intentional misrepresentations or memory errors. Participation in the data collection was anonymous, self-reported grades were neither accessible to teachers nor to other students. Therefore, the likelihood of deliberate misrepresentations should have been very small. Concerning memory errors, it seems reasonable that students might remember reported grades less accurate with an increasing time interval between report cards had been handed out and the data collection. Results by Dickhäuser and Plenter (2005) indicated no such time interval effect for seventh and eighth graders (three weeks vs. four and a half months). However, it must be noted that not only the time interval differed, but also the importance of the queried reported grades (reported grade of a specific exam vs. from a report card). The assumption of less accurate reported grades depending on the time interval calls for further research.

As to the results of study 1, valuable information concerning the accuracy of self-reported grades in elementary school students can be derived: Psychometric properties regarding the accuracy of self-reported grades were comparable for fourth grades and German secondary school students (Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008). Therefore, the results of study 1 indicate that the use of students' self-reported grades as indicators of their actual grades seems to be reasonable for fourth graders – provided that the data collection is anonymous. However, self-reported grades were sufficiently less accurate in second and third graders. Correlation differences between student- and teacher-reported grades with additional criteria indicated systematic biases of student-reported grades in terms of comparable or higher correlations with academic self-concepts. Construct-irrelevant variances of student-reported grades artificially increased their correlation with corresponding self-concepts, therefore affecting the validity of such self-reports in these elementary school grades. Building upon the results of study 1, teacher-reported grades were used as indicators for scholastic achievements in studies 2 and 3.

3.2 Excursus: Assessment of academic self-concepts

In this excursus, the potentially differential use of response formats with a varying number of response categories in elementary school students was examined (grades 1 to 4). Because of the focus of this thesis on scholastic achievements and academic self-

concepts, analyses were conducted on students' responses to competence and affect self-concept items of the well-established SDQ I (Marsh, 1992).

3.2.1 Theoretical background and research aims

When assessing data, questionnaires with three- to five-point Likert-type scales are typically used when applying rating scales with children (Mellor & Moore, 2013). However, there is no definitive agreement among researchers on the number of response categories that optimizes the psychometric properties of a Likert-type rating scale. Regarding relative frequencies of students chosen response categories in Likert-type rating scales, first and second graders seem to mostly tick off extreme response categories. In contrast, third and fourth graders seem to use extreme response categories less often and thus, tick off middle categories more frequently (e.g., Baldering, 1993, Chambers & Johnston, 2002). However, elementary school students are considered not to be overtaxed by rating scales with more than two response categories in principle Baldering (1993). In accordance with this assumption, satisfying psychometric properties like at least acceptable reliabilities were repeatedly reported for instruments assessing self-concepts with three-, four- or five-point Likert-type scales (Arens & Hasselhorn, 2015; Arens et al., 2013; Ehm, Lindberg et al., 2014; Lohbeck & Möller, 2017; Marsh et al., 1991; Pinxten et al., 2015; Poloczek et al., 2011).

To examine whether a measurement instrument is operating in a comparable manner for each response format under investigation, measurement invariance across response formats could be taken into account (Byrne, 1996; Christ & Schlüter, 2012; Geiser, 2013). Among others, results of a study by Adelson and McCoach (2010) with third to sixth grade students indicated a comparable factor meaning across four- and five-point Likert-type scales (metric measurement invariance). Unfortunately, analyses were only run for the total sample; the study does not allow drawing conclusions about measurement invariance across response formats separately for each grade. Thus, it was one aim of this thesis to extend this prior finding to (a) elementary school grades 1 to 4 with separate analyses per grade and to (b) three-point Likert-type scales as another frequently used response format with elementary school students.

Given the lack of systematic studies regarding the use of rating scales in elementary school students (grades 1 to 4) with a particular focus on the number of response categories, the following analyses aimed to shed light on

(1) relative frequencies of the chosen response categories for the competence and affect self-concept subscales in mathematics and reading as well as
(2) reliabilities of each self-concept subscale – separately for each elementary school grade and response format (three-, four-, or five-point Likert-type scale).

(3) Additionally, measurement invariance across response formats in an academic self-concept structure model was inspected – separately for each elementary school grade.
(4) Furthermore, students' fun and strain to answer academic self-concept items on the assigned response format were exploratorily examined.

Based on prior studies, it seemed reasonable to assume that first and second graders mostly tick off extreme response categories – regardless of the response format. In contrast, third and fourth graders should use extreme response categories less often and thus, tick off middle categories more frequently – again, regardless of the response format. Reliabilities might be at least acceptable and of comparable magnitude across response formats for the examined self-concept scales. Concerning measurement invariance across response formats, metric invariance across resonse formats was assumed at least for third and fourth graders; for elementary school students in lower grades, corresponding analyses were exploratory.

3.2.2 Method

Sample and procedure. The sample under investigation consisted of N = 1207 first to fourth graders (grade 1/2/3/4 n = 349/253/321/284) attending 83 classes in 16 German elementary schools. Participation was voluntary and anonymous. Students' and their parents' informed consent was obtained prior to testing, the parents of n = 196 students did not allow their child to participate. Furthermore, n = 198 students of the initial sample (N = 1609) did not take part due to illness or other reasons unrelated to the study, n = 8 students were excluded from the analyses due to a unreasonably low intelligence score.

Within each class, students were randomly assigned to one of the three different response formats under investigation and therefore, to different class rooms to respond to the self-concept items either on a three-, four-, or five-point Likert-type rating scale (for a more detailed sample description for each condition per grade see Tab. 4). All self-concept items were read aloud twice; after listening to each item, students marked their answer directly on their provided questionnaire.

Instruments. Competence and affect self-concepts in mathematics and reading were assessed with the corresponding four items per self-concept component of the SDQ I (Marsh, 1992; German translation by Arens, Trautwein et al., 2011). Depending on the response format, students answered the items either on a three-point (*strongly disagree* [1] – *partly agree* [2] – *strongly agree* [3]), on a four-point (*strongly disagree* [1] – *somewhat disagree* [2] – *somewhat agree* [3] – *strongly agree* [4]) or on a five-point (*strongly disagree* [1] – *somewhat disagree* [2] – *partly agree* [3] – *somewhat agree* [4] – *strongly agree* [5]) rating scale. Additionally, students were asked how much fun they

had while answering the SDQ I self-concept items and how strenuous it was to respond to the items (five-point Likert-type rating scale ranging from *a little* [1] to *a lot* [5]).

Table 4

Sample characteristics (sample size, gender, and age), manifest means (M) and standard deviations (SD) for each self-concept subscale as well as for the perceived fun and strain while answering the self-concept items – separately for each response category condition (three-/four-/five-point Likert-type scale) and grade

			number of response format		
grade			3	4	5
1	sample size	*n*	114	121	114
	female	*n*	56	63	52
	age	*M (SD)*	7.05 (0.44)	7.01 (0.40)	7.03 (0.50)
	SC math competence	*M (SD)*	2.53 (0.58)	3.43 (0.75)	4.43 (0.87)
	SC math affect	*M (SD)*	2.55 (0.65)	3.40 (0.88)	4.53 (0.87)
	SC reading competence	*M (SD)*	2.53 (0.56)	3.40 (0.76)	4.44 (0.81)
	SC reading affect	*M (SD)*	2.48 (0.61)	3.40 (0.82)	4.34 (1.02)
	fun	*M (SD)*	4.71 (0.75)	4.47 (1.11)	4.59 (1.08)
	strain	*M (SD)*	1.55 (1.00)	1.88 (1.35)	1.71 (1.28)
2	sample size	*n*	87	85	81
	female	*n*	45	47	41
	age	*M (SD)*	8.09 (0.64)	8.07 (0.43)	8.03 (0.42)
	SC math competence	*M (SD)*	2.48 (0.60)	3.42 (0.81)	4.07 (0.96)
	SC math affect	*M (SD)*	2.40 (0.68)	3.26 (0.97)	3.95 (1.16)
	SC reading competence	*M (SD)*	2.61 (0.48)	3.43 (0.73)	4.29 (0.80)
	SC reading affect	*M (SD)*	2.50 (0.53)	3.36 (0.83)	4.01 (1.08)
	fun	*M (SD)*	4.53 (0.95)	4.63 (0.89)	4.64 (0.83)
	strain	*M (SD)*	1.76 (1.03)	1.44 (0.79)	1.81 (1.32)
3	sample size	*n*	106	108	107
	female	*n*	53	58	62
	age	*M (SD)*	8.99 (0.40)	9.07 (0.49)	9.07 (0.52)
	SC math competence	*M (SD)*	2.33 (0.58)	3.08 (0.77)	3.97 (1.01)
	SC math affect	*M (SD)*	2.28 (0.65)	3.02 (1.00)	3.83(1.21)
	SC reading competence	*M (SD)*	2.50 (0.56)	3.45 (0.66)	4.32 (0.83)
	SC reading affect	*M (SD)*	2.32 (0.59)	3.16 (0.88)	4.09 (1.06)
	fun	*M (SD)*	4.30 (1.08)	4.56 (0.82)	4.53 (0.85)
	strain	*M (SD)*	1.67 (1.07)	1.40 (0.75)	1.58 (1.06)
4	sample size	*n*	92	96	96
	female	*n*	39	51	50
	age	*M (SD)*	10.03 (0.50)	10.08 (0.54)	10.06 (0.46)
	SC math competence	*M (SD)*	2.40 (0.58)	2.98 (0.86)	3.78 (0.99)
	SC math affect	*M (SD)*	2.25 (0.67)	2.83 (1.00)	3.37 (1.38)
	SC reading competence	*M (SD)*	2.51 (0.49)	3.26 (0.72)	3.99 (0.96)
	SC reading affect	*M (SD)*	2.29 (0.57)	2.98 (0.92)	3.88 (1.11)
	fun	*M (SD)*	4.17 (1.07)	3.97 (1.23)	4.28 (0.93)
	strain	*M (SD)*	1.56 (0.83)	1.81 (1.11)	1.55 (0.86)

Note. SC = self-concept.

Intelligence was assessed with the three nonverbal intelligence subtests of the CFT 1–R (Weiß & Osterland, 2013) to ensure sufficient cognitive ability for adequate comprehension processing and answering. Eight students (2nd grade: $n = 2$; 3rd grade: $n = 6$) with a score corresponding to an IQ below 70 were excluded from the sample.

Analyses. Concerning research aims (1) and (2), analyses regarding relative frequencies of the used response options as well as reliabilities were conducted with SPSS 23 – separately for each elementary school grade, response format, and each competence and affect self-concept subscale in mathematics and reading. Reliability values (Cronbach's α) below .60 are deemed unacceptable, .60 to .70 low but sufficient, .70 to .80 acceptable, .80 to .89 very good, and above .90 excellent (DeVellis, 2012, p. 109; Murphy & Davidshofer, 2001, p. 142).

Concerning research aim (3), analyses regarding measurement invariance were conducted with *MPlus* (7.11; Muthén & Muthén, 1998-2013). In order to examine measurement invariance across response formats, correlation models with the latent competence and affect self-concepts in mathematics and reading were specified separately for each grade and response format (Fig. 3). Latent self-concept factors were allowed to correlate as well as self-concept residual terms of items with identical item stems. In all models, the latent competence and affect self-concept factors in mathematics and reading were indicated by the corresponding SDQ I items, all non-standardized loadings were freed, and factor variances restricted to 1. Full-information maximum likelihood (FIML) estimation was used to handle the few missing data (maximum of 3.5% item non-response). The "type = complex" specification was used to control for potential effects due to school class affiliation and clustering of the data (students in classes). To test for metric invariance across response formats – separately for each grade – with the MLR estimator, a multi-group model with all factor loadings and intercepts freed across response formats was compared to a more restricted model with the factor loadings of each item constrained to be equal across response formats. To examine whether factor loading restrictions substantially worsened the model fit, $\Delta\chi^2$ (with Satorra-Bentler correction [test value T; Satorra, 2000]) and ΔCFI (ΔCFI < .01; Chen, 2007) were used as criteria between competing models. To describe model fits in general, χ^2 values with *df* were complemented by CFI, TLI, RMSEA, and SRMR/WRMR. CFI and TLI above .90 (.95), and RMSEA values below .08 (.05) typically signify an acceptable (good) fit (e.g., Little, 2013). Subsequently, corresponding analyses were run with *z*-standardized self-concept data (standardized per response format and grade). Such *z*-transformations per response format led to equivalent means and standard deviations across response formats and brought different response formats to a comparable metric without losing information about the magnitude of self-concepts in, e.g., students who

were assigned to the four- or five-point Likert-type scale compared to students who responded on a three-point scale.

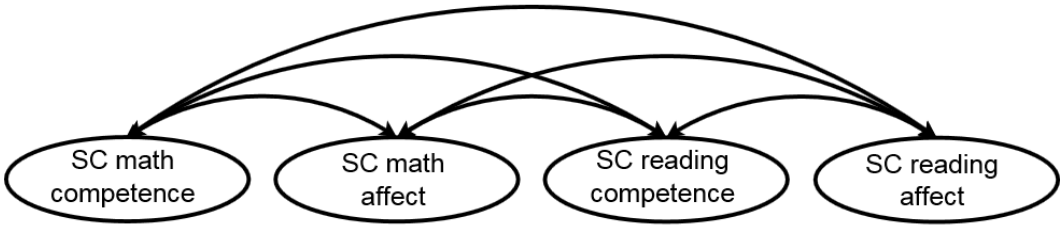


Figure 3. Self-concept correlation model with competence and affect self-concept factors for mathematics and reading. SC = self-concept.

To further underline the robustness of the data, analyses regarding measurement invariance were additionally run with the WLSMV estimator which is especially designed for ordinal data. However, with the WLSMV estimator, measurement invariance tests across response formats can only be run in *Mplus* and other statistical packages like *R* (R Core Team, 2013) if each group (response format: three-, four-, or five-point Likert-type rating scales) has the same categories for ordinal variables. To overcome this obstacle, Muthén (2011; see discussion 'categorical indicators' on www.statmodel.com) suggested to collapse categories for groups. In accordance with this suggestion and the relative frequencies of the used response categories (see Appendix B to E for the competence self-concepts in mathematics and reading – separately for each grade), categories for the four- and five-point scales were collapsed, therefore pretending that all students in all grades responded to all self-concept items on a three-point scale. In all response formats under investigation, the extreme categories were labeled the same (*strongly disagree* and *strongly agree*) and therefore remained as the extreme categories. Middle categories were collapsed for the four- and five-point scale to form one new middle category. Analyses regarding measurement invariance were identical to those with the MLR estimator.

Concerning the exploratory analyses of students' fun and strain to answer academic self-concept items on the assigned response format (research aim 4), means and standard deviations for these one-item measures were calculated with SPSS 23 – separately for each elementary school grade. ANOVAs with a three-stage factor 'response format' and the depended variable 'fun' or 'strain', respectively, were computed (p = .05).

3.2.3 Results

Means and standard deviations for each self-concept subscale are depicted in Table 1 – separately for each elementary school grade, response format, and the competence and affect self-concept subscales in mathematics and reading. Regarding the relative frequencies of the response categories for competence and affect self-concepts in mathematics (research aim 1; see Appendices B and C), first and second graders mostly used the response option indicating strong agreement in all response formats (grade 1: 65% to 73%; grade 2: 50% to 63%). The *strongly disagree* response option had rarely been ticked off (grade 1: 4% to 16%; grade 2: 3% to 15%). Middle categories were used more often compared to the option indicating strong disagreement but less often compared to the option indicating strong agreement. Third and fourth graders used the middle options more often (resulting in a flatter distribution of response frequencies). In the three-point response format, they even used the middle option (grade 3: 42% to 47%; grade 4: 43% to 55%) comparably or more often than the extreme category indicating strong agreement (grade 3: 43%; grade 4: 37% to 38%). Nonetheless, the *strong agreement* option was (almost) always ticked off the most in the four- and five-point response format (grade 3: 38% to 44%; grade 4: 33% to 36%). For the competence and affect self-concepts in reading, comparable result patterns emerged (see Appendices D and E), except that in third grade the *strong agreement* option was still used the most by the students, both for items of the competence as well as the affect self-concept component.

Concerning research aim (2), reliabilities of the competence and affect self-concept subscales in mathematics and reading were all of at least acceptable magnitude ($\alpha \geq .75$) and of comparable magnitude across response formats – regardless of the grade (Tab. 5). Additionally, Cronbach's α values for the self-concept scales were of numerically higher magnitude in higher elementary school grades.

Table 5

Cronbach's α for each self-concept subscale – separately for each grade and response format (three-/four-/five-point Likert scale)

grade	1	2	3	4
SC math competence	.84/.84/.82	.91/.92/.90	.91/.90/.89	.94/.92/.90
SC math affect	.89/.91/.87	.90/.93/.89	.94/.95/.94	.92/.94/.96
SC reading competence	.80/.80/.75	.84/.87/.87	.90/.92/.88	.85/.88/.90
SC reading affect	.85/.86/.83	.86/.87/.86	.90/.90/.91	.86/.91/.93

Note. SC = self-concept.

Table 6

Measurement invariance testing for self-concept correlation models across response format conditions (three-, four-, or five-point Likert-type scales) – separately for each grade with the MLR estimator and raw data in the upper part, with the MLR estimator and z-standardized data in the middle part, and with the WLSMV estimator and collapsed categories in the lower part

grade		χ^2	*df*	CFI	TLI	RMSEA	SRMR/ WRMR	$\Delta\chi^2$	Δdf	*p*	ΔCFI
MLR estimator, raw data											
1	con	515.41*	270	.880	.841	.088	.078	39.08	24	.05	.007
	met	554.49*	294	.873	.845	.087	.100				
2	con	475.29*	270	.903	.870	.095	.073	14.88	24	.70	.004
	met	490.17*	294	.907	.886	.089	.091				
3	con	367.41*	270	.968	.959	.058	.048	35.79	24	.06	.002
	met	402.64*	294	.966	.958	.059	.069				
4	con	415.18*	270	.951	.935	.075	.051	25.49	24	.39	<.001
	met	440.67*	294	.951	.940	.073	.067				
MLR estimator, z-standardized data											
1	con	515.41*	270	.880	.841	.088	.078	36.34	24	.05	.006
	met	552.15*	294	.874	.846	.087	.083				
2	con	475.29*	270	.903	.870	.095	.073	10.81	24	.99	.009
	met	480.28*	294	.912	.892	.087	.076				
3	con	367.41*	270	.968	.959	.058	.048	16.33	24	.88	.003
	met	386.32*	294	.971	.964	.054	.051				
4	con	415.18*	270	.951	.935	.075	.051	11.85	24	.98	.004
	met	427.24*	294	.955	.945	.069	.054				
WLSMV estimator, collapsed categories											
1	con	357.95*	270	.988	.984	.053	1.383	24.39	24	.44	.001
	met	375.00*	294	.989	.986	.049	1.429				
2	con	369.17*	270	.986	.982	.066	1.416	26.97	24	.31	.001
	met	390.58*	294	.987	.984	.062	1.452				
3	con	338.85*	270	.995	.993	.049	1.180	24.55	24	.43	<.001
	met	361.40*	294	.995	.994	.046	1.207				
4	con	342.15*	270	.994	.991	.053	1.220	27.91	24	.26	<.001
	met	364.99*	294	.994	.992	.051	1.263				

Notes. χ^2 = chi-square goodness-of-fit statistic; *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual; SRMR = standardized root mean squared residual; con = configural; met = metric. *p < .05.

Regarding measurement invariance across response formats (research aim 3; separately for each grade) with the MLR estimator and raw (non-standardized) data as well as with the MLR estimator and *z*-standardized data, the models with all factor loadings and intercepts freed across response formats and the more restricted model with

the factor loadings of each item constrained to be equal across response formats revealed unacceptable fit values in grade 1 and 2 (Tab. 6). In contrast, fit values were at least acceptable for both models in grade 3 and 4. However, non-significant nested comparison values ($p \geq .05$) and ΔCFI < .01 further indicated metric factorial invariance across response formats in all grades (Tab. 6).

With the WLSMV estimator and therefore, with collapsed categories, fit indices for the models with all factor loadings and intercepts freed across response formats and for the more restricted model with the factor loadings of each item constrained to be equal across response formats were at least acceptable in all grades (Tab. 6). Due to non-significant nested comparison values ($p \geq .05$) and ΔCFI < .01, measurement invariance across response formats was assumed for each grade.

Concerning students' self-perceived fun and strain while answering academic self-concept items (research aim 4), means were above the midpoint of the scale for fun ($3.97 \leq M \leq 4.71$; Tab. 4) and below the midpoint for experienced strain: ($1.40 \leq M \leq 1.88$) – regardless of the assigned response format and grade. Within each grade, no significant differences across response formats occurred for experienced fun (grade 1: $F[2;325] = 1.59$, $p = .21$, $\eta^2 = .010$; grade 2: $F[2;243] = 0.39$, $p = .68$, $\eta^2 = .003$; grade 3: $F[2;312] = 2.46$, $p = .09$, $\eta^2 = .016$; grade 4: $F[2;275] = 1.95$, $p = .15$, $\eta^2 = .014$) and strain (grade 1: $F(2;314) = 1.81$, $p = .17$, $\eta^2 = .011$; grade 2: $F(2;243) = 2.95$, $p = .06$, $\eta^2 = .024$; grade 3: $F(2;312) = 2.14$, $p = .12$, $\eta^2 = .013$; grade 4: $F(2;273) = 2.23$, $p = .11$, $\eta^2 = .016$).

3.2.4 Discussion

The aim of this excursus was to contribute to the knowledge about the potentially differential use of response formats with a varying number of response categories (three-, four-, or five-point Likert-type scales) in elementary school students. Therefore, relative frequencies of the chosen response categories for the competence and affect self-concept subscales in mathematics and reading as well as reliabilities of each self-concept subscale were analyzed. Additionally, measurement invariance across response formats was inspected.

The results support prior findings regarding response biases (Baldering, 1993, Chambers & Johnston, 2002) in the sense that first and second graders mostly ticked off extreme response categories and that middle categories were rather rarely chosen. In contrast, third and fourth graders used extreme response categories less often and thus, ticked off middle categories more often. The number of response categories available to students did not seem to impact these tendencies. In each grade, reliabilities of the competence and affect self-concept subscales in mathematics and reading were all of at least acceptable magnitude and of comparable magnitude across response formats. Also for each grade, metric measurement invariance across response formats was revealed emphasizing that the measurement part of the self-concept factors was comparable across

response formats and indicating a comparable content-related meaning for the competence and affect self-concepts factors in mathematics and reading across response formats. Additionally, students seemed to have fun and to perceive strain to a small extent while answering the academic self-concept items of the SDQ I.

Despite these encouraging results, the issue whether assessments with different response formats result in equivalent construct measurements, and subsequently, in comparable relations between variables needs further exploration. First, the present sample size per grade–response format combination was quite small ($81 \leq n \leq 121$), especially in grade 2 ($n \leq 87$). This was an inevitable consequence of the strict randomization procedure across all response formats and grades. Second, comparisons could only be made between students' answers on a three-, four-, or five-point Likert-type scale due to the present study design. In future studies, an intraindividual study design would be desirable to investigate whether students might prefer one response format over the other and might provide, for example, less biased ratings in one format. The influence of possible moderators for choosing particular categories of a non-dichotomous rating scale (e.g., intelligence) could be examined.

The results of this excursus indicate that no fundamentally different data regarding self-concepts across response formats were provided in elementary school students. Based on these results, *z*-standardized data (standardized per response format and grade) from different response scales were combined to form one sample per grade. Furthermore, with *z*-standardization the data could be regarded as continuous and therefore, met the criteria for the MLR estimator.

3.3 Study 2: Internal/external-frame-of-reference model

Schneider. R., & Sparfeldt, J. R. (under revision). Twofold multidimensional academic self-concepts in elementary school students: Social and dimensional comparisons. *Contemporary Educational Psychology*.

In study 2, we examined effects of scholastic achievements in mathematics and German on (non-)corresponding competence and affect self-concepts in elementary school students (grades 1 to 4) within the I/E model framework. The sample of this study stemmed from German elementary schools in Lower Saxony, Saxony-Anhalt and Saarland. Because students in these federal states receive reported grades from second grade onwards, reported grade equivalents (see below) were used as achievement indicators in grade 1. In grade 2 to 4, reported grades in mathematics and German indicated corresponding scholastic achievements. Because of the differing grading procedure across grades, analyses regarding or including first grade students can only be considered as supplemental analyses.

3.3.1 Preliminary analyses

In the manuscript of study 2, we focused on social and dimensional comparison processes within the I/E model. Analyses regarding the twofold multidimensional structure of academic self-concepts indicating a competence-affect self-concept distinction in elementary school grades 1 to 4 were included in the manuscript. Furthermore, correlations between competence and affect self-concepts within and across subjects were reported (Tab. 3 in the manuscript). However, correlations were not examined regarding grade-related differences. For the sake of completeness, the preliminary analyses for study 2 shed light on

(a) average correlations between self-concept factors as well as the specific correlations within (e.g., competence and affect self-concept in mathematics) and across subjects (e.g., competence self-concept in mathematics and reading) – separately for each elementary school grade.

(b) Subsequently, the self-concept factor correlations were compared between grades.

Based on prior results of studies, we expected lower average self-concept correlations in higher elementary school grades. More specifically, we expected lower self-concept correlations within the competence or affect component (competence mathematics and competence reading self-concept, affect mathematics and affect reading self-concept) in higher grades. The extent to which competence and affect components are distinguished within the mathematical and verbal domain might not vary between grades.

Method. These preliminary analyses for study 2 were based on the sample described in the excursus (subsection 3.2.2). The assessment procedure is described in detail in subsection 3.1.2.

Competence and affect self-concepts in mathematics and reading were assessed with the corresponding four items per self-concept component of the SDQ I (Marsh, 1992; German translation by Arens, Trautwein et al., 2011). Students responded to these items either on a three-, four-, or five-point Likert-type rating scale. We adjusted students' responses per response format in each grade using *z*-transformation (see section 3.2 for the rationale behind this procedure).

All preliminary analyses were conducted with *MPlus* (7.11; Muthén & Muthén, 1998-2013) with common specifications (FIML estimation, "type = complex" specification, MLR estimator). A self-concept correlation model with competence and affect factors for mathematics and reading was specified separately for each elementary school grade. Self-concept factors were indicated by the corresponding SDQ I items with loadings of the first item per factor restricted to 1. Self-concept residual terms of items with identical item stems were allowed to correlate.

Measurement invariance as a prerequisite for subsequent comparisons of correlation coefficients across grades was inspected (Geiser, 2013, p. 99–116). To test for

metric invariance, a model with all factor loadings and intercepts freed across the four grades was compared to a model with the factor loadings constrained to be equal across grades (using Satorra-Bentler corrected $\Delta\chi^2$ and ΔCFI < .01).

Regarding self-concept correlations in each grade (research aim a), self-concept correlations were averaged after Fisher's transformation of the correlation coefficients and then backtransformed. Concerning research aim (b), the resulting averaged correlations (one per grade) were compared across grades (χ^2-difference test, test value *V*; Bortz, 2005); the effect size *q* was calculated for pairwise comparisons.

Furthermore, the correlation pattern of self-concept factors was inspected separately for each grade. To analyze whether specific self-concept factor correlations differed significantly between grades, a sequential approach was used. For all models used in this approach, we imposed equality constraints on the residuals of self-concept items with identical item stems and constrained self-concept factor variances to 1. In an initial omnibus test ($\alpha < .05$), a multi-group model with all self-concept factor correlations constrained to be equal across grades was compared to a multi-group model without these constraints (multi-group standard model). If the used Wald test (test value T_w) revealed significant test statistics, the equality constraints would worsen the model fit significantly and thus, the compared path coefficients differ statistically significant. In case of self-concept correlation differences across grades, we further analyzed among which respective grades the self-concept factor correlations differed significantly. Therefore, two-group models among two respective grades with again all self-concept factor correlations constrained to be equal were compared to corresponding two-group models with self-concept correlations that were allowed to vary (two-group standard model; one-tailed, 6 comparisons, adjusted $\alpha < .017$). Finally and in order to analyze which specific self-concept factor correlations differed significantly among two respective grades, two-group models with one self-concept factor correlation constrained equal between grades were compared to the two-group standard model (one-tailed, 6 comparisons, adjusted $\alpha < .017$). The effect size *q* was calculated.

Results. Concerning measurement invariance across grades, the differentially restricted models with latent competence and affect mathematics and reading self-concept factors revealed good fit indices (Tab. 7). Due to a non-significant nested comparison value (baseline–metric: $T = 36.36$, $df = 36$, $p = .45$) measurement invariance can be assumed. ΔCFI also confirmed measurement invariance (configural–metric: ΔCFI < .001).

Average self-concept correlations (grade 1/2/3/4: $\bar{r}$ = .70/.52/.44/.49) differed significantly across grades ($V = 29.063$, $df = 3$, $p < .01$). Pairwise comparisons showed significant differences for the average correlations between grade 1 vs. 2 ($p < .01$, $q = .29$), grade 1 vs. 3 ($p < .01$, $q = .40$), and between grade 1 vs. 4 ($p < .01$, $q = .33$). The

remaining mean correlations did not differ statistically significantly (grade 2 vs. 3: p = .11, q = .10; grade 2 vs. 4: p = .32, q = .04; grade 3 vs. 4: p = .22, q = -.06).

Table 7

Fit indices for the measurement invariance models and the models for group comparisons between grades

	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Measurement invariance of the 4-factor model						
configural	633.32*	360	.962	.950	.054	.049
metric	672.44*	396	.962	.954	.052	.052
Fit indices for the multi-group self-concept correlation model without restrictions						
1 to 4	913.97*	552	.964	.960	.047	.047
1 vs. 2	495.28*	260	.940	.930	.055	.053
1 vs. 3	471.62*	260	.958	.950	.049	.047
1 vs. 4	507.47*	260	.949	.940	.055	.049
2 vs. 3	368.98*	260	.979	.975	.038	.044
2 vs. 4	399.95*	260	.971	.966	.045	.046
3 vs. 4	349.19*	260	.985	.983	.034	.037

Notes. χ^2 = chi-square goodness-of-fit statistic; *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean squared residual. *p < .05.

Concerning specific correlations of the self-concept factors in all four elementary school grades (Tab. 8), the competence and affect factors within each self-concept domain were highly correlated in all four grades (mathematics: .75 ≤ r ≤ .95; reading: .71 ≤ r ≤ .89). The competence self-concept in mathematics showed low to medium relations to the competence self-concepts in reading (.27 ≤ r ≤ .49), the affect self-concept in mathematics also showed low to medium relations to the affect self-concepts in reading (.19 ≤ r ≤ .47).

Self-concept factor correlations differed significantly across grades (multi-group model: T_w = 47.55, *df* = 18, p < .01) and more specifically between grade 1 vs. 3 (T_w = 28.20, *df* = 6, p < .01) and grade 1 vs. 4 (T_w = 17.16, *df* = 6, p < .01). No significant differences were found for the remaining comparisons between grades (grade 1 vs. 2: T_w = 11.08, *df* = 6, p = .08; grade 2 vs. 3: T_w = 4.61, *df* = 6, p = .59; grade 2 vs. 4: T_w = 11.32, *df* = 6, p = .08; grade 3 vs. 4: T_w = 12.57, *df* = 6, p = .05). Furthermore, 6 out of the resulting 12 comparisons (2 grade comparisons [grade 1 vs. 3 and 1 vs. 4] × 6 self-concept factor correlations) revealed significant differences (see Tab. 8). Large effect sizes in terms of smaller correlations in higher grades resulted for the mathematics

competence–mathematics affect correlations (grade 1 vs. 3: q = 0.81; grade 1 vs. 4: q = 0.86), medium effect sizes were revealed for the mathematics affect–reading affect (grade 1 vs. 4: q = 0.32), the reading competence–reading affect (grade 1 vs. 3: q = 0.49) and the mathematics affect–reading competence self-concept correlations (grade 1 vs. 3: q = 0.34; grade 1 vs. 4: q = 0.29).

Table 8

Latent correlations of competence self-concepts and affect self-concepts in mathematics and reading – separately for elementary school grades 1 to 4

	math competence				math affect				reading competence			
grade	1	2	3	4	1	2	3	4	1	2	3	4
math affect	.95[a]	.86[a,b]	.77[b]	.75[b]								
reading competence	.49	.38	.33	.27	.45[a]	.28[a,b]	.14[b]	.19[b]				
reading affect	.39	.20	.16	.18	.47[a]	.29[a,b]	.26[a,b]	.19[b]	.88[a]	.76[a,b]	.71[b]	.89[a,b]

Notes. First grade: all correlations $p < .05$; second grade: $r > .20$ $p < .05$; third grade: $r > .16$ $p < .05$; fourth grade: all correlations $p < .05$. [ab]Different indices indicate significant correlation coefficient differences between grades (e.g., the correlations between competence and affect self-concepts in mathematics significantly differ between grade 1 vs. 3 and 1 vs. 4, but not between grade 1 vs. 2, 2 vs. 3, 2 vs. 4, nor between grade 3 vs. 4).

Discussion. Concerning the preliminary analyses of study 2, average self-concept correlations within the self-concept model with competence-affect distinction were substantially lower in second, third, and fourth grades in comparison to first grade students. From second grade on, self-concept correlations remained relatively stable. This result pattern became also evident from an inspection of specific self-concept correlations that, if they varied significantly between grades, were substantially lower in grades 3 and 4 than in grade 1. In all four elementary school grades, the competence and affect factors were highly, but not perfectly correlated within each subject. These correlations were comparable in magnitude with previous findings (e.g., Arens & Hasselhorn, 2015; Arens, Trautwein et al., 2011; Marsh et al., 1999) and indicated, along with factor analytical evidence, a competence-affect self-concept distinction in all four elementary school grades. The results also further emphasize the domain specificity of the competence and the affect components.

Building upon these results, it seemed reasonable to assign the competence-affect self-concept distinction to the I/E model. Consequently, study 2 focused on the relations between reported grades and competence/affect self-concepts within the I/E model.

Thereby, the partially lower academic self-concept correlations in higher grades compared to first grade might be explained by comparison processes postulated within the I/E model and their differences between elementary school grades.

3.3.2 Theoretical background and research aims

Academic self-concepts are important correlates and predictors of successful scholastic learning (e.g., Craven & Marsh, 2008; Guay et al., 2010; Marsh, 2007; Möller et al., 2009). Within the well-established I/E model framework (Marsh, 1986), the formation of domain-specific academic self-concepts is explained by the joint operation and especially the relative weight of social and dimensional comparison processes. Social (internal) comparison processes refer to the comparisons made by students relating their own academic achievement in one subject to the achievements of others in the same subject. Dimensional (internal) comparisons refer to the contrast of one's academic achievement in one subject with one's own achievement in another subject (for more details see subsection 2.3.2).

Across elementary school grades, social comparison processes seem to increase, cognitively more demanding dimensional comparison processes seem to occur earliest in third grade (for theoretical considerations see Guay et al., 2003; Harter, 2006; Skaalvik & Skaalvik, 2002) resulting in lower correlations between subject-specific self-concepts in higher elementary school grades (also see subsection 3.3.1). Ehm, Lindberg et al. (2014) analyzed the I/E model separately for first, second, and third graders. Concerning comparison processes for scholastic achievements in mathematics and reading, all path coefficients indicating social comparison processes were substantial in all grades. Furthermore, these path coefficients were of substantially higher magnitude in higher grades. Path coefficients indicating dimensional comparison processes were non-substantial in first and second grade, but substantial in third grade. Subsequently, the correlation between corresponding self-concept residuals was of smaller magnitude in higher grades. Unfortunately, the interpretation of those results across grades is limited by, for example, the use of different indicators of scholastic achievements across grades.

So far, I/E model-related studies with elementary school students did not take the competence–affect self-concept distinction into account. Due to, for example, the high correlation between competence and affect self-concepts, the I/E model pattern with scholastic achievements and competence self-concepts should be also applicable to scholastic achievements and affect self-concepts. Furthermore, no studies on the formation of subject-specific affect self-concepts (or closely related constructs) with elementary school students exist, yet. Assigning the competence–affect distinction to the I/E model framework would allow the examination of possibly differential relations between scholastic achievement and the two correlated, but still distinct self-concept

components and, thus, deepen the understanding of social and dimensional comparisons on both self-concept components.

Therefore, we extended prior research by assigning the competence-affect-distinction to the I/E model framework to simultaneously examine relations between scholastic achievements and both academic self-concept components in elementary school students. In study 2, we inspected

(1) relations within the I/E model focusing on (a), correlations between scholastic achievements, (b) effects of scholastic achievements on corresponding competence and affect self-concepts (indicating social comparison processes) and on both non-corresponding self-concept components (indicating dimensional comparison processes) as well as (c) self-concept residual correlations – separately for each elementary school grade (1 to 4).

(2) Subsequently, we compared these relations between scholastic achievements and (non-)corresponding competence and affect self-concepts among elementary school grades.

Based on prior results of I/E model studies with elementary school students, we expected stronger social comparison processes in higher grades and substantial dimensional comparison processes emerging earliest in third grade. Concerning the competence-affect distinction, we expected more pronounced relations between reported grades and (non-)corresponding competence self-concept measures compared to the relations between reported grades and affect self-concepts.

3.3.3 Method

Sample and procedure. Study 2 was based on the sample described in the excursus (subsection 3.2.2). The assessment procedure is described in detail in subsection 3.1.2.

Instruments. Scholastic achievement was operationalized by reported grades in mathematics and German. Class teachers copied these reported grades from their student's last midterm report cards. In first grade in Lower Saxony, Saxony-Anhalt, and Saarland, students generally do not get numeric reported grades. Thus, teachers were asked to evaluate their students' achievements on a grade-equivalent scale (reported grade equivalents). Reported grades were inversely scored so that higher numerical values indicated higher scholastic achievements. Competence and affect self-concepts were assessed as described in the preliminary analyses section of this study (subsection 3.3.2).

Analyses. All analyses were conducted with *MPlus* (7.11; Muthén & Muthén, 1998-2013) with common specifications (full-information maximum likelihood (FIML) estimation, "type = complex" specification, MLR estimator). Concerning research aim (1), extended I/E models with manifest reported grades in mathematics and German as achievement indicators as well as competence and affect self-concept factors for

mathematics and reading were specified separately for each elementary school grade (Fig. 4). Within these extended I/E model, (a) correlations between reported grades, (b) unidirectional paths from reported grades on their (non-)corresponding self-concept factors, and (c) the correlation pattern of self-concept factor residuals was inspected.

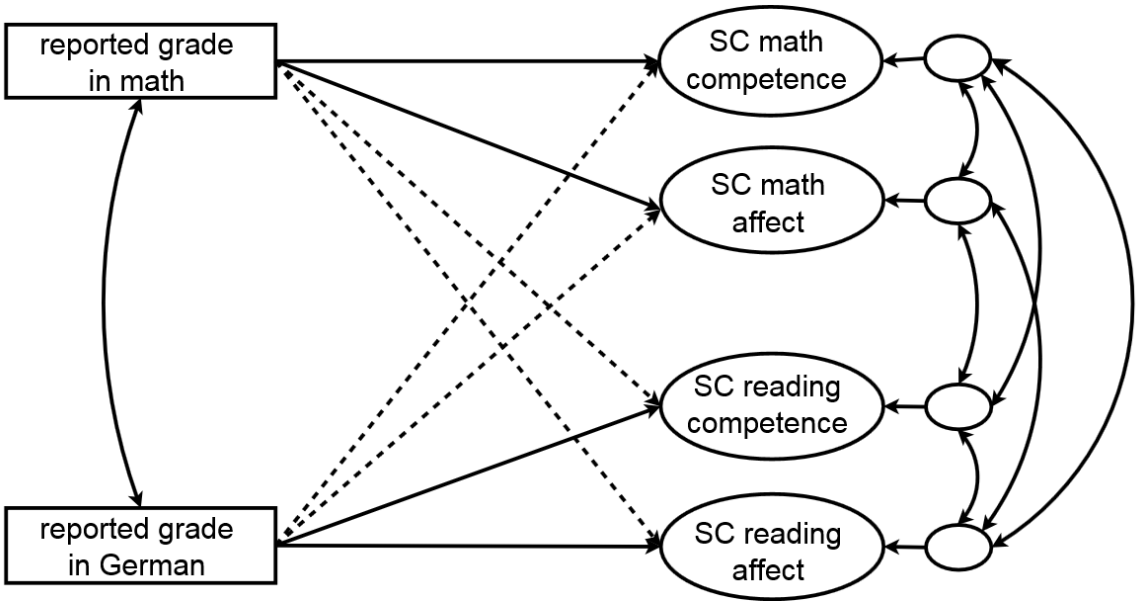


Figure 4. Extended I/E model with manifest mathematics and German reported grades as well as latent competence and affect self-concept (SC) factors for mathematics and reading. Unidirectional solid line = social comparison processes, unidirectional dotted line = dimensional comparison processes.

After measurement invariance was supported, analyses regarding differences in path coefficients between grades were run twice: first, only with students with numerical reported grades (grades 2 to 4) and second with all students also including grade 1. I/E model analyses regarding first grade or including first grade are considered as supplemental analyses because of the different grading procedure in grade 1 compared to grade 2 to 4. To inspect if (and subsequently, which) path coefficients from reported grades on (non-)corresponding competence and affect self-concepts differed significantly between grades, differentially restricted multi-group I/E models were compared to each other (for these models, correlations between reported grades in mathematics and German were constrained to be equal across grades in order to avoid biased results due to covariance shifting). A corresponding analysis strategy was used to compare correlation coefficients of the self-concept factor residuals across grades (for these models, correlations between reported grades in mathematics and German as well as path

coefficients from reported grades on competence and affect self-concepts were constrained to be equal across grades).

3.3.4 Results

Within the extended, grade-specific I/E models of research aim (1), (a) correlations between reported grades in mathematics and German were of large magnitude in all four elementary school grades ($.55 \leq r \leq .74$; Tab. 9). Concerning effects of scholastic achievements on academic self-concepts (research aim 1b), path coefficients from reported grades on corresponding competence self-concepts indicating social comparison processes were mostly of large magnitude for mathematics ($.45 \leq \beta \leq .74$) and mostly of medium magnitude for German ($\beta = .33 \leq \beta \leq .52$). For the affect self-concept factors, path coefficients were of moderate magnitude for mathematics ($.35 \leq \beta \leq .48$) and for German ($.20 \leq \beta \leq .33$).

Table 9

Reported grade correlations, standardized path coefficients relating grades to the corresponding and non-corresponding self-concept factors, and correlations of the self-concept factor residuals within the I/E model framework – separately for elementary school grades 1 to 4

	1	2	3	4
RG math ↔ RG German	.74*[a]	.55*[b]	.65*[b]	.64*[b]
RG math → SC math competence	.45*[a]	.57*[a,b]	.62*[a,b]	.74*[b]
RG math → SC math affect	.42*	.35*	.43*	.48*
RG math → SC reading competence	-.16*	.00	-.02	.06
RG math → SC reading affect	-.07	-.19*	-.16	-.02
RG German → SC math competence	-.04	-.10	-.08	-.18*
RG German → SC math affect	-.15*	-.10	-.21*	-.21*
RG German → SC reading competence	.52*	.31*	.43*	.33*
RG German → SC reading affect	.33*	.20*	.32*	.30*
SC math competence ↔ SC math affect	.95*[a]	.88*[a,b]	.79*[b]	.76*[b]
SC reading competence ↔ SC reading affect	.85*[a]	.81*[a,b,c]	.73*[b,c]	.89*[a,c]
SC math competence ↔ SC reading competence	.43*[a]	.39*[a,b]	.30*[a,b]	.18*[b]
SC math affect ↔ SC reading affect	.46*[a]	.37*[a,b]	.36*[a,b]	.17*[b]
SC math competence ↔ SC reading affect	.34*[a]	.30*[a,b]	.27*[a,b]	.11[b]
SC math affect ↔ SC reading competence	.42*[a]	.32*[a,b]	.15*[b]	.15*[b]

Notes. RG = reported grade, SC = self-concept. *$p < .05$. [abc]Different indices per row indicate significant coefficient differences between grades (e.g., the correlations between reported grades in mathematics and German significantly differ between grade 1 vs. 2, 1 vs. 3, and 1 vs. 4, but not between grade 2 vs. 3, 2 vs. 4, nor between grade 3 vs. 4).

Within the I/E model framework, path coefficients between reported grades and non-corresponding competence and affect self-concepts indicating dimensional comparison processes were mostly negative ($-.21 \leq \beta \leq .06$), but only some path coefficients substantially differed from zero (e.g., the path coefficient from the reported grade in German on the mathematics affect self-concept factor in first, third, and fourth grades; Tab. 9). Regarding correlations of self-concept residuals (research aim 1c), competence and affect self-concept factor residuals were highly correlated within subjects in all grades (mathematics: $.76 \leq r \leq .95$; reading: $.73 \leq r \leq .89$). Across subjects, correlations were of low to medium magnitude for the competence component ($.18 \leq r \leq .43$) as well as the affect component ($.17 \leq r \leq .46$).

Regarding grade-related differences (research aim 2), the correlation between reported grades in mathematics and German was substantially higher in grade 1 compared to the correlation in all other grades (Tab. 9). Furthermore, the relation between reported grades in mathematics and the mathematics competence self-concept was stronger in grade 4 compared with grade 1. All other path coefficients on corresponding self-concepts did not differ substantially between grades. Path coefficients indicating dimensional comparison processes did not differ substantially across grades. Furthermore, significant differences between self-concept residual correlations occurred mostly between grade 1 and 3 or grade 1 and 4 with lower correlations in higher grades (for detailed results see Tab. 9).

3.3.5 Discussion

In study 2, we inspected effects of scholastic achievements in mathematics and German on (non-)corresponding academic competence and affect self-concepts within the I/E model framework. Combining the competence–affect distinction and the traditional I/E model, the extended I/E model of study 2 allowed the simultaneous examination of path coefficients from reported grades on (non-)corresponding competence self-concepts and on (non-)corresponding affect self-concepts. In all elementary school grades, path coefficients on corresponding competence and affect self-concepts were of moderate or high magnitude. Path coefficients on the non-corresponding self-concept components were mostly negative and partially (6 of 16 path coefficients) substantial. No grade-related differences occurred, except for the path coefficient from the reported grade in mathematics on its corresponding competence self-concept factor which was of higher magnitude in grade 4 compared to grade 1. Additionally, significant self-concept residual differences mostly occurred between grade 1 and higher elementary school grades in terms of lower correlations in higher grades.

Concerning path coefficients on competence self-concepts, our results were in line with previous studies for social comparison processes. But against expectations, only some path coefficients indicating dimensional comparison processes were substantial in

our data. One possible reason for these differing results might be the wording of the self-concept items. In contrast to other self-concept instruments, items of the SDQ I did not promote any social compasrison processes (e.g., “In mathematics, I am one of the best students”) or dimensional comparison processes (e.g., “I am better in mathematics than German”), thereby not evoking such comparison processes artificially. Regarding affect self-concepts, path coefficients from reported grades on affect self-concepts were of numerically smaller magnitude compared to path coefficients on competence self-concepts (as expected from research regarding expectancy-value theory). These results widened the knowledge about the interplay of social and dimensional comparison processes on the formation of subject-specific affect self-concepts. Furthermore, conclusions on construct-relevant differences between both self-concept components can be drawn. As expected, scholastic achievement and competence self-concept showed a higher conceptual proximity than scholastic achievement and affect self-concept which was evidenced by stronger bi- and unidirectional relations for the competence than the affect self-concept component with corresponding reported grades.

Furthermore, the specific grade where reported grades instead of verbal evaluations are implemented in elementary school varies between federal states (see section 2.1.2). Our sample was derived from federal states where reported grades are implemented in grade 2. One might intuitively assume that the magnitude of social and dimensional comparison processes in grade 2 might depend on the grade reported grades are implemented. In our sample, the implementation of reported grades did not result in substantial differences in comparison processes between grade 1 and higher elementary school grades (except for the path coefficient from the reported grade in mathematics on its corresponding competence self-concept factor). First graders already showed a result pattern that corresponded with substantial social and some substantial dimensional comparison processes indicating that the formation of domain-specific academic self-concepts might be based on additional feedback sources like oral feedback by teachers and classmates and/or observations in class. For second graders without reported grades, path coefficients of comparable magnitude as in grade 1 (and in our sample also as in grade 3 and 4) can be expected.

3.4 Study 3: Statistical prediction of scholastic achievements

Schneider, R., Lotz, C., & Sparfeldt, J. R. (2018). Smart, confident, interested: Contributions of intelligence, self-concept, and interest to elementary school achievement. *Learning and Individual Differences, 62*, 23–35.

In study 3, we examined the statistical prediction of reported grades by intelligence, their corresponding competence self-concepts, and corresponding affect self-concepts–

separately for the school subjects mathematics and German. Furthermore, we focused on the unique effects that each predictor exhibited on reported grades beyond the other predictors. In the corresponding publication of study 3, affect self-concept was referred to as interest. However, in this thesis, the term "affect self-concept" is used because of the close connection to previous self-concept research (see section 2.2.1 and 2.2.2 for the competence–affect distinction of academic self-concepts) and the operationalization of affect self-concepts by the corresponding SDQ I items.

The sample of this study stemmed from German elementary schools in Lower Saxony, Saxony-Anhalt and Saarland. In these federal states, students receive reported grades from second grade onwards. Therefore, the sample for the following analyses only comprised elementary school students in grades 2 to 4.

3.4.1 Theoretical background and research aims

Although scholastic achievement and therefore, educational success, is largely determined by intelligence (e.g., Deary, Strand, Smith, & Fernandes, 2007; Jensen, 1998; Roth et al., 2015), there is a substantial amount of achievement variance left unexplained. In past research with elementary school students, motivational variables such as academic self-concepts were frequently shown to account for substantial amounts of achievement variance beyond intelligence in fourth graders (e.g., Helmke, 1997b; Schicke & Fagan, 1994; Spinath et al., 2006; Weber et al., 2013).

The only study with intelligence, competence self-concept, and affect self-concept (interest) as separate predictors of scholastic achievements instead of conglomerates of various cognitive or motivational variables revealed that both self-concept components were substantial predictors of scholastic achievement beyond intelligence in prediction models with solely intelligence and competence self-concept or solely intelligence and affect self-concept (Spinath et al., 2006). However, the predictive power of affect self-concept (interest) vanished when the three predictors were simultaneously considered possibly due to the large amount of shared variance between competence and affect self-concepts within one subject. Commonality analyses in the study by Spinath et al. (2006) were only based on two predictors (intelligence and one self-concept component). Therefore, the analysis of specific and common variances within a simultaneous prediction of scholastic achievement by intelligence, competence self-concept, and affect self-concept still needed further examination.

Furthermore, all previous studies examining the prediction of scholastic achievement were conducted with fourth graders. However, one could assume differences in the prediction of scholastic achievements by intelligence and academic self-concepts across elementary school grades due to a progression in cognitive development, self-concept formation processes (see subsection 2.3.2 for the formation of domain-specific academic self-concepts in the I/E model framework), and increasing scholastic

experiences across elementary school grades. Thus, the examination of the prediction pattern of intelligence, competence self-concept, and affect self-concept on scholastic achievement in each elementary school grade and across grades as well as the examination of unique effects of each predictor beyond the others per grade would deepen the understanding of the interplay of cognitive and motivational variables as predictors of scholastic achievement in elementary school students.

Therefore, we aimed to examine the relevance of intelligence, competence self-concept, and affect self-concept on reported grades in the core elementary school subjects mathematics and German in four steps:

(1) Within statistical prediction models (separately for grades 2, 3 and 4 as well as separately for the two school subjects mathematics and German), we statistically tested the differential relevance of intelligence and competence self-concept for reported grades. Unique effects of each predictor beyond the other were examined.
(2) Analogously to (1), we examined the statistical prediction reported grades in mathematics and German by intelligence and affect self-concept.
(3) Considering the substantial amount of shared variance between all variables (especially between the two self-concept components), we statistically tested the differential relevance of all three predictors in concert and examined the unique effects of each predictor beyond the others.
(4) Possible grade-related differences in the predictive power of intelligence, competence self-concept, and affect self-concept across elementary school grades were inspected.

Concerning the simultaneous prediction of the reported grade using only intelligence and one self-concept component, we expected to find unique effects of intelligence beyond the respective self-concept component as well as unique effects of competence or affect self-concept beyond intelligence. Within a model with all three predictors in concert, we expected the unique effects of affect self-concept to vanish, but competence self-concept and intelligence to show unique effects beyond the other two respective predictors. Concerning possible developmental changes in the predictive power of the three predictors across elementary school grades, we assumed comparable path coefficients of intelligence on the reported grade across grades and higher path coefficients from both self-concept components on the reported grade in higher elementary school grades.

3.4.2 Method

Sample and procedure. Study 3 was based on the sample comprising N = 858 second to fourth grade students (grade 2/3/4 n = 253/321/284) attending 59 classes in 16 German elementary schools. The assessment procedure is described in detail in subsection 3.1.2.

Instruments. Non-verbal intelligence as well as teacher-reported grades as indicators of scholastic achievements in mathematics and German were assessed as described in subsection 3.1.2. Reported grades were inversely scored so that higher numerical values indicated higher scholastic achievements.

Competence and affect self-concepts in mathematics and reading were assessed with the corresponding four items per self-concept component of the SDQ I (Marsh, 1992; German translation by Arens, Trautwein et al., 2011). Students responded to these items either on a three-, four-, or five-point Likert-type rating scale. We adjusted students' responses per response format in each grade using *z*-transformation (see section 3.2 for the rationale behind this procedure).

Analyses. All analyses were conducted with *MPlus* (7.11; Muthén & Muthén, 1998-2013) with common specifications (FIML estimation, "type = complex" specification, MLR estimator). Regarding research aim (1), a latent model with intelligence and competence self-concept as predictors and the reported grade as the criterion was established (intelligence–competence self-concept model) – separately for the two subjects mathematics and German as well as separately for elementary school grades 2 to 4. Equality constraints were imposed on path coefficients from the predictors on the reported grade to test whether path coefficients differed significantly from each other. To examine the increments of intelligence and competence self-concept, we employed the Cholesky factoring approach on the prediction model (de Jong, 1999; Loehlin, 1996; also see Lotz et al., 2018) – again, separately for mathematics and German as well as grades 2 to 4. By orthogonally decomposing the explained variances, a hierarchical regression analysis could be performed in only one model without altering the model fit or affecting the measurement part of the model. To examine the intelligence increment, competence self-concept was assigned first priority, second priority was assigned to intelligence. Thus, the intelligence Cholesky factor represented the intelligence increment after the shared variance with competence self-concept was partialled out. Analogously, to examine the competence self-concept increment, we assigned intelligence first and competence self-concept second priority. An inspection of the path coefficients from the Cholesky factors with last priority on the reported grade indicated whether the particular increment substantially predicted the reported grade beyond the other predictor. Moreover, squared regression coefficients can be interpreted as the proportion of uniquely explained variance by each predictor on the reported grade.

Regarding research aim (2), all analyses were conducted analogously for a model with intelligence and affect self-concept as predictors (intelligence–affect self-concept model) and concerning research aim (3), for a model with intelligence, competence self-concept, and affect self-concept as predictors (3-predictor model). Finally and concerning grade-related differences of research aim (4), path coefficients in the prediction model

with intelligence and both self-concept components in concert were compared across grades (multi-group analyses; after measurement invariance across grades was supported).

3.4.3 Results

All results are first presented for mathematics and second for German.

Mathematics. Concerning research aim (1), path coefficients from intelligence and competence self-concept (intelligence–competence self-concept model) on the reported grade were substantially positive and of comparable magnitude in all elementary school grades ($p \geq .73$; Fig. 5). Correspondingly, increments of both predictors were of comparable magnitude in all grades (intelligence: grade 2/3/4 $\Delta R^2 = .20/.18/.12$, ps $< .05$; competence self-concept: $\Delta R^2 = .18/.18/.14$, ps $< .05$).

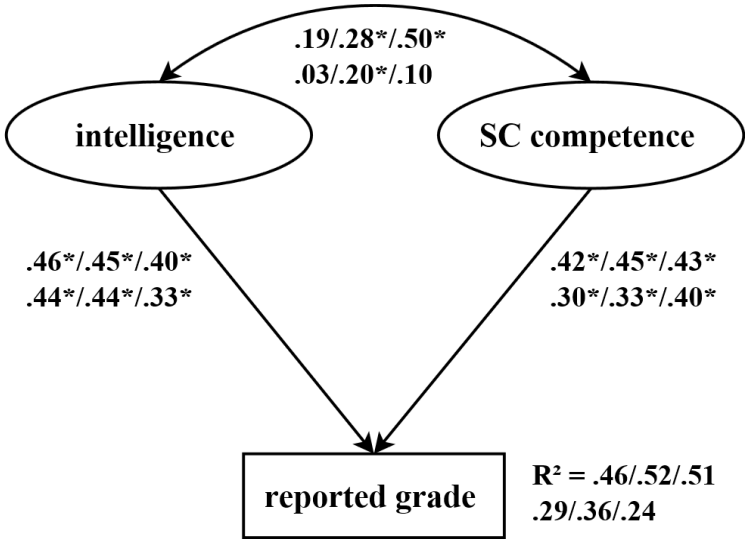


Figure 5. Intelligence–competence self-concept model with path coefficients separately for each grade (2/3/4) and the two school subjects mathematics (above) and German (below). SC = Self-concept. *$p < .05$.

Within the intelligence–affect self-concept model of research aim (2), the path coefficients from intelligence were substantially positive in all grades in mathematics whereas for affect self-concept, only the path coefficients in grades 2 and 3 were substantially positive. Regarding the differential relevance of both predictors, intelligence revealed significantly higher path coefficients on the reported grade in mathematics than affect self-concept in all grades (ps $< .01$; Fig. 6). In this model, the intelligence increments accounted for $\Delta R^2 = .26/.30/.30$ (ps $< .05$), affect self-concept only revealed substantial increments in grades 2 and 3 ($\Delta R^2 = .06/.04$, ps $< .05$).

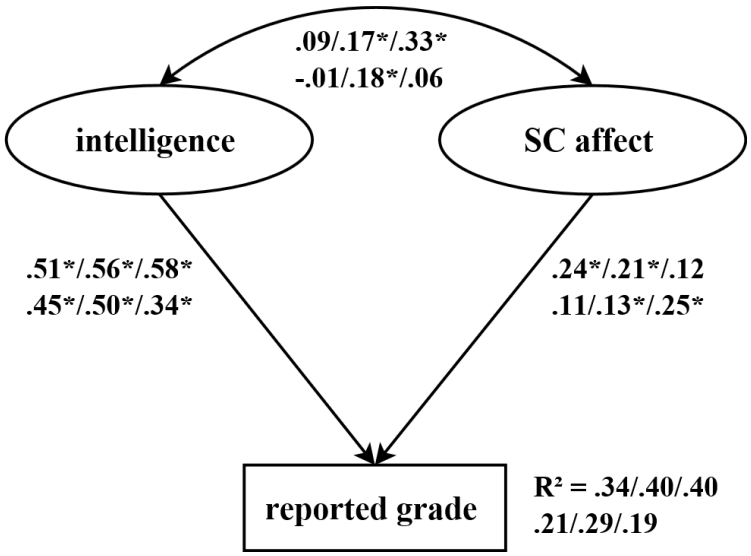


Figure 6. Intelligence–affect self-concept model with path coefficients separately for each grade (2/3/4) and the two school subjects mathematics (above) and German (below). SC = Self-concept. *p < .05.

When all three predictors were simultaneously taken into consideration (research aim 3), the path coefficients from intelligence and competence self-concept on the reported grade were substantially positive whereas the path coefficients from affect self-concept on the reported grade were substantially negative in all grades (Fig. 7; ps < .05). Concerning increments of the three predictors beyond the other two predictors, increments of intelligence and competence self-concept were of comparable magnitude in all grades (intelligence: ΔR^2 = .17/.17/.11, ps < .05; competence self-concept: ΔR^2 = .19/.18/.14, ps < .05). The affect self-concept increments explained only a very small amount of reported grade variance (ΔR^2 = .07/.04/.03, ps < .05) beyond the other two predictors. A total amount of 54% to 55% reported grade variance was explained by all three predictors, common variances accounted for ΔR^2 = .11/.16/.26.

Regarding grade-related differences (research aim 4), no differential relevance of the three predictors occurred across elementary school grades 2 to 4 (p = .43).

German. For German, result patterns were mostly comparable to the ones for mathematics. Path coefficients from intelligence and competence self-concept (intelligence–competence self-concept model; research aim 1) on the reported grade in German were substantially positive and of comparable magnitude in all elementary school grades ($p \geq .13$; Fig. 5). Correspondingly, increments of both predictors were of comparable magnitude in all grades (intelligence: grade 2/3/4 ΔR^2 = .19/.18/.10, ps < .05; competence self-concept: ΔR^2 = .09/.10/.12, ps < .05).

Within the intelligence–affect self-concept model of research aim (2), the path coefficients from intelligence were substantially positive in all grades whereas for affect self-concept, only the path coefficients in grades 3 and 4 were substantially positive. Regarding the differential relevance of both predictors, intelligence revealed significantly higher path coefficients on the reported grade in German than affect self-concept in grades 2 and 3 (*p*s < .01; Fig. 6) but not grade 4 (*p* = .43). In this model, the intelligence increments accounted for ΔR^2 = .20/.24/.12, *p*s < .05, affect self-concept only revealed substantial increments in grades 3 and 4 (ΔR^2 = .02/.06, *p*s < .05).

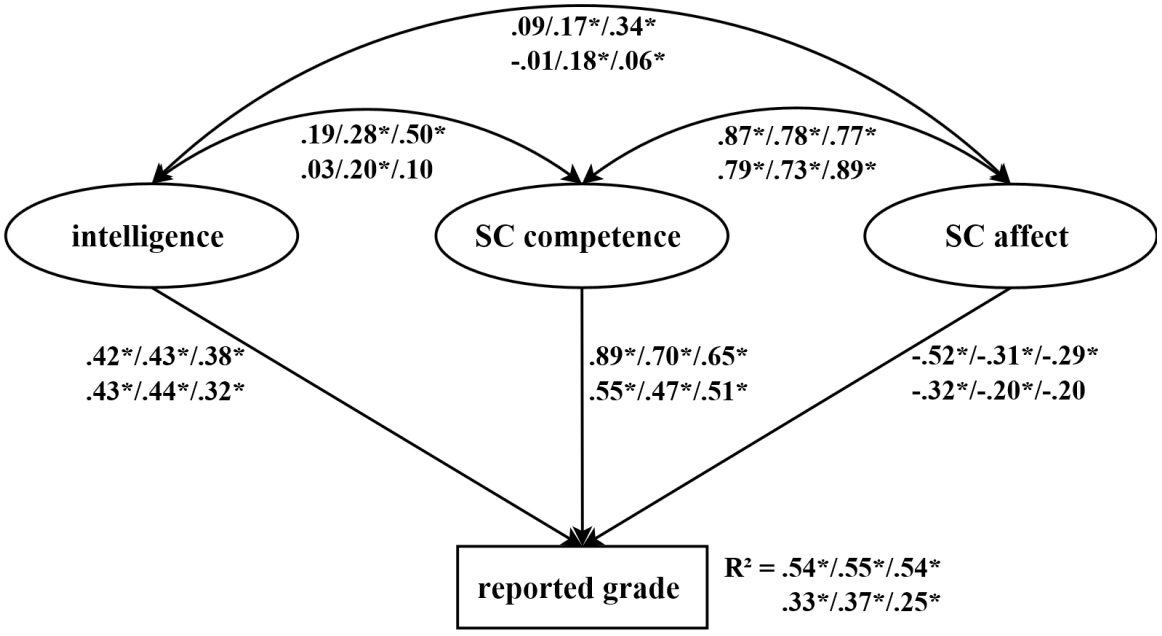


Figure 7. 3-predictor model with path coefficients separately for each grade (2/3/4) and the two school subjects mathematics (above) and German (below). SC = Self-concept. **p* < .05.

When all three predictors were simultaneously taken into consideration (research aim 3), the path coefficients from intelligence and competence self-concept on the reported grade in German were substantially positive whereas the path coefficients from affect self-concept on the reported grade were substantially negative in grades 2 and 3 (Fig. 7; *p*s < .05). Concerning increments of the three predictors, increments of intelligence and competence self-concept were of comparable magnitude in all grades (intelligence: ΔR^2 = .18/.18/.10, *p*s < .05; competence self-concept: ΔR^2 = .12/.10/.05, *p*s < .05). The affect self-concept increments explained only a very small substantial amount of reported grade variance in grades 2 and 3 (ΔR^2 = .04/.02, *p*s < .05) beyond the other

two predictors. A total amount of 25% to 37% reported grade variance was explained by all three predictors. In grade 2, predictors did not share common variances; for grades 3 and 4, the amount of shared variances was small, $\Delta R^2 = .07/.09$.

Regarding grade-related differences (research aim 4), again no differential relevance of the three predictors occurred across elementary school grades ($p = .75$).

3.4.4 Discussion

In study 3, the statistical prediction of reported grades in mathematics and German by intelligence, corresponding competence self-concept, and corresponding affect self-concept was examined for elementary school grades 2, 3, and 4. Simultaneously considering intelligence and one self-concept component as predictors, both self-concept components explained unique reported grade variance beyond intelligence (with the exception of the affect self-concept of fourth graders in mathematics and of second graders in German). Whereas intelligence and competence self-concept were comparable relevant for reported grades in both school subjects and all examined grades, intelligence was mostly more relevant than affect self-concept for the corresponding reported grade. When all three predictors were jointly considered, competence and affect self-concept contributed substantially to the prediction of reported grades in mathematics and German beyond intelligence in all three examined grades, with the exception of the affect self-concept of fourth graders in German. No evidence was provided for significant grade-related differences.

Concerning both self-concept components, competence self-concept was the stronger motivational predictor. Affect self-concept only explained very small amounts of reported grade variance in all examined grades and both school subjects. These small increments of affect self-concept on reported grades are in line with expectancy-value theory and might be the result of the highly structured school system in elementary school that lacks opportunities like course choices (Kunter, Baumert, & Köller, 2007). In contrast to Geman elementary schools, affect self-concepts might have a stronger effect on corresponding scholastic achievements in higher grades when students have more possibilities to engage in interest-based learning activities.

When examining path coefficients from intelligence and competence self-concept as well as path coefficients from intelligence and affect-self-concept on corresponding reported grades, all path-coefficients were of positive magnitude. Surprisingly, when controlling for the shared variance between intelligence, competence self-concept, and affect self-concept (Fig. 7), path coefficients from competence-self-concept on the reported grade increased in both subjects and all grades in comparison to the reported grade–competence self-concept correlation and the path coefficients from competence self-concept on the reported grade in the intelligence–competence self-concept model (Fig. 5). Simultaneously, the path coefficients from affect-self-concept on the reported

grade decreased in comparison to the reported grade–affect self-concept correlation and the path coefficients from affect self-concept on the reported grade of the intelligence–affect self-concept model (Fig. 6) and became even negative in both subjects and all grades. This result pattern indicates a suppression phenomenon (see Conger, 1974). Therefore, a content-related interpretation of path coefficients might result in incorrect conclusions (e.g., that higher affect self-concepts lead to worse reported grades). One reason for the occurrence of this suppressor effect could be that the competence self-concept items might have contained not only competence components, but also affective components. These affective aspects were extracted from the competence self-concept factor and a purer estimate of self-perceived ability remained in the competence self-concept factor which might have resulted in the increased relation between competence self-concept and reported grades. However, the affective components that might have been extracted from the competence self-concept factor are theoretically no criteria-irrelevant variance and should also be positively related to the criteria. Thus, this explanatory approach merely provides an explanation for the increased competence self-concept–reported grade-relation, but not for the decreased affect self-concept–reported grade-relation. Therefore, this unexpected suppressor effect should be interpreted with caution and awaits its replication.

4 General discussion

The aim of this thesis was to shed light on the interplay of scholastic achievements and academic self-concepts in elementary school students and to detect possible substantial differences in these relations across grades. In this chapter, the main findings of this thesis are summarized first. Second, implications for further research as well as practical implications are pointed out. Third, the design of the dissertation project, key features of the presented analyses, and methodological issues are critically discussed.

4.1 Summary of main findings

Prior to examining the relations between subject-specific scholastic achievements (operationalized by reported grades) and academic self-concepts, study 1 and the excursus dealt with the assessment of reported grades and academic self-concepts in elementary school students. The results of study 1 indicated that students' self-reported grades were less reliable and valid indicators of their actual reported grades in grades 2 and 3 than in grade 4 or as reported in prior studies with secondary school students (cf. Dickhäuser & Plenter, 2005; Sparfeldt et al., 2008). Concerning the assessment of competence and affect self-concepts, elementary school students in grades 1 to 4 did not seem to be overtaxed by academic self-concept rating scales with three, four, or five response categories. Psychometric properties were comparable across response formats. Moreover, metric measurement invariance indicated a comparable content-related meaning for competence and affect self-concept factors in mathematics and reading across response formats.

In study 2, we examined effects of reported grades in mathematics and German on corresponding competence and affect self-concepts (indicating social comparison processes) and on non-corresponding competence and affect self-concepts (indicating dimensional comparison processes) within the well-established I/E model framework (Marsh, 1986). These unidirectional relations were inspected separately for elementary school grades 1 to 4 and subsequently compared across grades. Path coefficients indicating social comparison processes were positive and of moderate to high magnitude in all four elementary school grades. Path coefficients indicating dimensional comparisons were mostly negative, but only 6 out of 16 path coefficients were substantial

without revealing a systematical pattern. The relations between reported grades and corresponding competence self-concepts were more pronounced compared to those between reported grades and affect self-concepts. Regarding grade-related differences, only the relation between the reported grade in mathematics and the mathematics competence self-concept was significantly stronger in grade 4 compared to grade 1.

In study 3, attention was shifted towards the statistical prediction of reported grades in mathematics and German by intelligence, corresponding competence self-concepts, and corresponding affect self-concepts. The incremental validity of intelligence, competence self-concept, and affect self-concepts for reported grades was examined – separately for the two core subjects mathematics and German as well as elementary school grades 2 to 4. When considering intelligence in combination with solely competence self-concept or affect self-concept, each of them was a substantial predictor of reported grades in both subjects and all examined elementary school grades (with the exception of the affect self-concept of fourth graders in mathematics and of second graders in German). Whereas intelligence and competence self-concept were comparably important predictors of reported grades, intelligence was a stronger predictor than affect self-concept. When controlling for the shared variance between competence and affect self-concept, both self-concept components contributed substantially to the prediction of reported grades above and beyond intelligence (with the exception of affect self-concept of fourth graders in German). Moreover, competence self-concept was the stronger predictor compared to affect self-concept. Regarding grade-related differences in the prediction of reported grades, no such differences were evidenced for mathematics or German.

4.2 Implications

The following section focuses on implications for further educational and psychological research on the assessment of scholastic achievements, on the assessment of competence and affect self-concepts as well as on the interplay of scholastic achievements and academic self-concepts in elementary school students. Furthermore, practical implications are highlighted in each subsection.

4.2.1 Implications regarding the assessment of scholastic achievements and academic self-concepts

The results of study 1 concerning the accuracy of students' self-reported grades in mathematics, German, and sports indicated that self-reported grades were less accurate for second and third grades than secondary school students (results for secondary school students were reported by Dickhäuser and Plenter [2005] as well as Sparfeldt et al. [2008]). Even though correlations between students' self-reported grades and their actual

grades were of comparable magnitude with typically found test-retest-reliabilities, substantial over-reports raise doubts whether students' self-reported grades are adequate indicators of scholastic achievements in grades 2 and 3. If students' self-reported grades are nevertheless used in these grades, one might expect increased measurement errors and corresponding limitations regarding the interpretation of study results. For fourth graders, self-reported grades were quite accurate; accuracy values were of comparable magnitude with values for secondary school students. Therefore, the use of students' self-reported grades as indicators of students' actual grades seems to be reasonable for fourth graders – provided that the data collection is anonymous. Building upon the results of study 1, we used teacher-reported grades as indicators of scholastic achievements in studies 2 and 3.

However, students in all examined elementary school grades showed a consistent tendency to portray their own academic achievements as slightly better than their actual performances what might represent a self-enhancement mechanism (Buckelew, Byrd, Key, Thornton, & Merwin, 2013; Dickhäuser & Plenter, 2005; Schwarz & Beaver, 2014; Sticca, Goetz, Nett, Hubbard, & Haag, 2017; Willard & Gramzow, 2008). Positive short-term effects of over-reporting grades in terms of a self-enhancement strategy might boost students' mood, competence self-concept, and self-esteem. But, such strategies might also go along with long-term risks like maladaptive learning strategies as well as less learning effort and, consequently, lower academic achievements (e.g., Sticca, Goetz, Nett et al., 2017). Robins and Beer (2001) hypothesized that if self-enhancement (e.g., through over-reported grades) leads to higher competence self-concepts and greater learning effort, scholastic achievements should remain stable or increase over time. In contrast, if self-enhancement leads to higher competence self-concepts but also lower learning effort (students might underestimate the level of effort needed for certain levels of achievement and, therefore, might invest an insufficient amount of effort), scholastic achievements should decrease over time. Furthermore, research has shown that self-enhancement might go along with a typical pattern of causal attributions: Whereas success is typically attributed to internal and stable causes like one's own ability, failure is attributed to external and unstable causes (Buckelew et al., 2013; Robins & Beer, 2001). Such attributions might again lead to lower scholastic achievements: Students might not realize the reason(s) for their failure (e.g., poor reported grades), therefore, the likelihood of adjusting one's learning behavior might be reduced (Buckelew et al., 2013). To sum up, it seems to be "crucial to educate students on the importance of accurate self-evaluations and realistic self-expectations" (Sticca, Goetz, Nett et al., 2017, p. 852). This might be especially relevant for elementary school students in order to reduce the likelihood of insufficient learning behavior and, therefore, the likelihood of subsequent lower scholastic achievements.

Regarding the assessment of competence and affect self-concepts, elementary school students did not seem to be overtaxed by academic self-concept rating scales with three, four, or five response categories. Because no fundamentally different data regarding self-concepts across response formats per grade were provided, researchers could either choose three-, four-, or five-point Likert-type rating scales in future studies when assessing academic self-concept data with elementary school students (using the SDQ I). However, based on the results presented in the excursus, it seems to be sufficient to assess competence and affect self-concepts with an economic three-point rating scale for cross-sectional as well as longitudinal studies with only elementary school students. Rating scales with at least four response categories are typically used with secondary school students allowing a reliable and valid assessment of academic self-concepts (e.g., Brunner et al., 2010; Möller, Zimmermann, & Köller, 2014; Niepel et al., 2014, Rost et al., 2007; Schilling et al., 2006). For studies with samples comprising elementary as well as secondary school students, self-concept rating scales with four or five response categories could be used (e.g., if grade-related mean differences in academic self-concepts shall be examined) allowing to keep the number of response categories constant across age groups.

However, future studies concerning the comparison of response formats with regard to the number of presented response categories should take further aspects into account, for example, correlations with additional criteria. Regarding criterion validity, self-concept correlations with additional criteria should not vary substantially or systematically between response formats within each grade. Additionally, further formats like the four-point structure-alternative format (e.g., SPPC; Harter, 1985; also see Byrne, 1996) should be compared with the traditional rating scale format. Applying this format, students first have to respond to each item by indicating which of two logically opposed statements best describes them (e.g., “Some kids often forget what they learn” vs. “Other kids remember things easily”). Second, students are asked whether the chosen statement was “really true for me” or “sort of true for me”. The SPPC is validated for children aged eight to 15 years and its initial development demonstrated strong psychometric properties (Harter, 2012). However, it can be assumed that providing only two response options at a time, especially younger elementary school students might be able to provide less biased ratings. Moreover, the availability of appropriate visual aids for each response category (e.g., representing the response categories with different sized circles in the SPPC) might enhance the ability to use rating scales in elementary school students (Chambers & Johnston, 2002).

The present analyses evaluating the use of rating scales in elementary school students with a particular focus on the number of response categories only took competence and affect self-concepts in mathematics and reading of the SDQ I into

account. Further studies need to inspect whether the findings can be generalized to other instruments to assess academic self-concepts or even to other constructs (see, e.g., Sparfeldt, Lotz, Sapp, & Rost, 2016, for social self-concepts). For example, Lohaus (1989) emphasized that the optimal number of response categories used with rating scales might vary depending on the complexity of the construct under investigation. At low complexity, younger students might be able to respond to corresponding items on a rating scale with a high number of categories. However, younger students might be overtaxed with too many response categories if construct complexity increases (also see Diersch & Walther, 2010).

4.2.2 Implications regarding the interplay of scholastic achievements and academic self-concepts

Studies 2 and 3 were conducted to examine the interplay of scholastic achievements and academic self-concepts in elementary students. In both studies, competence and affect self-concepts revealed a (very) high construct overlap in mathematics (grade 1/2/3/4: r = .94/.87/.79/.78) and reading (r = .86/.79/.73/.90) pointing towards the problem of multicollinearity between variables (Field, 2009). However, the variance inflation factors did not indicate severe multicollinearity for mathematics and reading in all grades (mathematics grade 1/2/3/4: VIF = 3.79/2.06/1.49/1.64; reading grade 1/2/3/4: VIF = 2.78/1.41/1.43/1.76; VIF < 10 suggests non-collinearity, see Myers, 1990). Furthermore, analyses regarding the structure of academic self-concepts in mathematics and reading indicated a preference for structure models with separate competence and affect self-concept factors for mathematics and reading compared to models with mathematics and reading self-concept factors but without a competence–affect distinction in all elementary school grades (fit indices for academic self-concept structure model with and without competence-affect distinction are reported in the manuscript of study 2). As mentioned before, reasons for the very close relation between competence and affect self-concepts within subjects might be that (a) the competence and affect component are both facets of the underlying construct of motivation, (b) both components are reciprocally related, and/or (c) both components are based on self-reported data and, therefore, share method variance. The results of a cross-lagged panel study with seventh graders gave strong hints for a reciprocal connectedness of competence self-concept and affect self-concept (Marsh et al., 2005). Prior competence self-concept significantly affected subsequent affect self-concept (interest) in mathematics, whereas prior mathematics affect self-concept (interest) had only a small effect on the subsequent competence self-concept in mathematics. Reciprocal effect models with competence and affect self-concepts with elementary school students should be examined in future research.

Furthermore, relations between reported grades and competence self-concepts were more pronounced than reported grade–affect self-concept relations. These results

underlined the higher conceptual proximity between scholastic achievement and self-perceived scholastic competence (competence self-concept) compared to intrinsic motivational-affective perceptions (affect self-concept) in all elementary school grades. These results were in line with expectancy-value theory (Wigfield & Eccles, 2000). According to this theory, competence self-concepts are based on students' actual scholastic achievements and corresponding interpretations, judgments by significant others as well as internalized cultural values and norms. By contrast, affect self-concepts are supposed to be determined by individual goals, general self-schemas, and the social background of a student (e.g., parents, culture, norms). Therefore, scholastic achievements should be less relevant for affect self-concepts.

Internal/external-frame-of-reference model. The extension of the traditional I/E model with regard to competence and affect self-concepts (study 2) offered new insights into the effects of social and dimensional comparison processes on the formation of subject-specific competence and affect self-concepts. The extension allowed the inspection of path coefficients of reported grades on the two highly, but not perfectly correlated self-concept components within one model. So far, prior studies either included only self-concept items to assess the competence or affect component, or integrated both self-concept components into one factor (with the exception of the study by Kadir et al., 2017, with secondary school students). The latter might have masked potentially differential relations of scholastic achievement measures on competence and affect self-concepts. As it is difficult to compare I/E model studies with the same constructs under investigation across studies (due to study-specific confounding variables, e.g., differing samples or instruments), it is even more difficult to compare I/E model studies investigating different constructs (competence self-concepts vs. affect self-concepts). Even if path coefficients on competence or affect self-concepts differ between models, no firm conclusions on construct-relevant differences between both self-concept components can be drawn. To overcome that obstacle, the extended I/E model with separate competence and affect self-concept factors was introduced in study 2. As expected, relations between reported grades and their corresponding competence self-concepts were more pronounced than the reported grade–affect self-concept relations. For dimensional comparison processes, path coefficients on competence or affect self-concepts, respectively, were of comparable magnitude.

In accordance to self-determination theory (e.g., Ryan & Deci, 2000), one might assume that the affect self-concept component is a causal consequence of the competence component. Within the I/E model framework, this line of reasoning would result in an I/E model structure where only the competence component is postulated to be dependent on social and dimensional comparison processes. In such a model structure, competence self-concepts act as mediators between reported grades and affect self-concepts. Prior research

with fifth to tenth graders supported such mediating effects: For an I/E model with reported grades in mathematics and German as well as enjoyment ratings in mathematics and language classes (as indicators of affect self-concept), Goetz et al. (2008) reported substantial social (mathematics/German: β = .40/.40) and dimensional comparison processes (mathematics on German/German on mathematics: β = -.21/-.20).

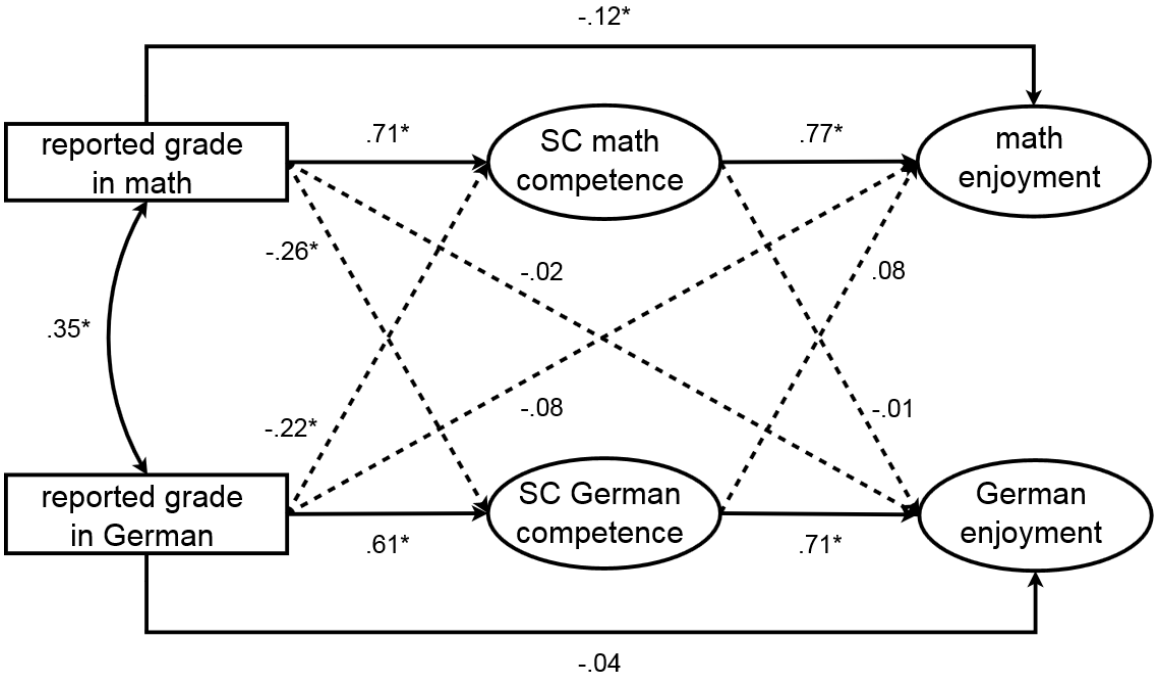


Figure 8. Mediation model by Goetz et al. (2008, p. 21). SC = self-concept, unidirectional solid line = social comparison processes, unidirectional dotted line = dimensional comparison processes.*$p < .05$.

Second and within the mediation model (Fig. 8), direct paths from reported grades on corresponding enjoyment ratings (mathematics/German: β = -.12/.-.04) as well as non-corresponding enjoyment ratings (mathematics on German/German on mathematics: β = -.02/-.08) were non-significant or of negligible effect size. Indirect effects of reported grades in mathematics/German on corresponding enjoyment ratings through corresponding competence self-concepts were substantial (mathematics/German: β = .53/.44). Examining such mediation effects in elementary school students should be part of future research to further deepen the understanding about the interplay of scholastic achievements and motivational-affective self-perceptions. However, longitudinal data should be preferred for examining such mediation effects due to substantially biased estimates of longitudinal parameters in cross-sectional analyses (Maxwell & Cole, 2007).

The results of study 2 indicated that students in grades 1 to 4 carried out social comparison processes. Dimensional comparison processes played only a minor role. Regarding the magnitude of dimensional compassion processes in the I/E model framework, substantial negative path coefficients from scholastic achievements on non-corresponding academic self-concepts were repeatedly reported for third and fourth graders in prior studies (e.g., Ehm, Lindberg et al., 2014; Ehm, Nagler et al., 2014; Pinxten et al., 2015; Skaalvik & Valås, 2011). In our data, only some path coefficients indicating dimensional comparison processes were substantial. As mentioned, one possible reason for these differences might be that the wording of self-concept items in other studies explicitly provoked comparison processes. Overall, there is evidence for the ecological validity of especially dimensional comparison processes. Diary studies (e.g., Möller & Husemann, 2006; Wheeler & Miyake, 1992) have demonstrated that high school and college students made social and dimensional comparison processes with regard to academic matters. However, comparison processes might have been triggered by a specific meeting in preparation for the diary study and by task instructions (Möller & Husemann, 2006). For elementary school students, no such evidence from experimental studies exists, so far. Furthermore, further research is needed to study possible long-term effects of dimensional comparison processes on the formation of academic self-concepts in elementary school students. These effects might explain the partially lower academic self-concept correlations in higher grades compared to first grade.

Concerning psychological and pedagogical consequences of comparison processes, students might socially compare their own achievements with others to establish class rank orders in a reference group and, therefore, to normatively evaluate their own performances. One has to keep in mind that equally able students may evaluate their performance differently according to the achievement level of their class/reference group (social comparison process). According to the big-fish-little-pond-effect (Marsh, 1987; Marsh et al., 2008; Trautwein & Lüdtke, 2005), equally capable students have higher self-concepts in a less capable class/reference group than in a more capable class/reference group. Dimensional comparisons might be useful for students to gain a more differentiated picture of individual strengths and weaknesses. This might be especially relevant for secondary school students who have to, for example, select an educational track with an emphasis on foreign languages or sciences (Köller, Daniels, Schnabel, & Baumert, 2000; Nagy et al., 2006, 2008). Furthermore, it is assumed that internally comparing achievements might regulate affective states and enhance self-worth in the worse-off domain (Möller & Marsh, 2013) by switching their attention to the better off-domain (Möller, Helm, Müller-Kalthoff, Nagy, & Marsh, 2015).

Beyond such positive aspects, the link between scholastic achievements and academic self-concepts should not only be treated as a positive relation. Destructive

effects of poor scholastic achievements and negative feedback on the development of competence and affect self-concepts must be taken into consideration, as well. Dimensional comparison processes might result in an over- or underestimation of one's own abilities and, connected therewith, competence self-concepts in the better-off or worse-off domain, respectively (Möller et al., 2015). Long-term risks like maladaptive learning strategies or learning effort for the better-off domain may subsequently cause lower academic achievements. For teachers, knowledge about the joint operation of social and dimensional comparison processes for the formation of academic self-concepts is important in order to support their students in the development of positive, but realistic self-concepts in different subjects as well as an adequate learning behavior. Wolff, Nagy, Helm, and Möller (2018) pointed out that teachers should, for example, "stress common aspects of math and verbal subjects, such as the dependence of achievement in both domains on effort, and the fact that math and verbal abilities are not mutually exclusive" (p. 65). This might also be effective for affect self-concepts due to the reciprocal connectedness of both self-concept components (Marsh et al., 2005).

Statistical prediction of scholastic achievements. One major aim of educational researchers and practitioners is the enhancement of scholastic achievements. Study 3 underlined the importance of intelligence and competence self-concepts for the statistical prediction of reported grades in mathematics and German. Contributions of affect self-concepts beyond the other predictors were of (very) small magnitude. Current research repeatedly attempted to enhance reported grades or other indicators of scholastic achievement by trainings with a focus on students' cognitive abilities (for an overview see Karbach, Schmiedek, & Hasselhorn, 2018). Especially working memory and executive functions (components of process-based trainings) are considered to be important prerequisites for knowledge acquisition and were shown to be highly correlated with scholastic achievements (Titz & Karbach, 2014). However, the trainability of intelligence or intelligence components like working memory is seen controversially (e.g., Schwaighofer, Fischer, & Bühner, 2015; Zigler & Styfco, 2004). Competence self-concept interventions might be a more fruitful alternative to enhance scholastic achievements. A meta-analytic review demonstrated that competence self-concept interventions were very effective when targeting domain-specific self-concepts (d =1.16; O'Mara, Marsh, Craven, & Debus, 2006). Furthermore, affect self-concept interventions might influence scholastic achievement via the large amounts of shared variance with the corresponding competence self-concept component. According to von Maurice et al. (2014), subject-specific affect self-concepts might be quite easily fostered by curricular or instructional interventions. It will be a task of future intervention studies to examine whether targeting students' self-concept components might be more efficient than targeting students' intelligence to improve students' reported grades. One might also

assume that combined trainings might reveal the highest effect on scholastic achievements.

The decision for a specific secondary school track has to be made after grade 4 (in most German federal states including those from which the sample in the present studies originated) which is usually done on the basis of reported grades on the report cards at this time (Baumert et al., 2003; Bos et al., 2004). According to the conference of ministers of education (Kultusministerkonferenz, 2006) teachers should also include students' tendency and willingness to work hard (working behavior and motivation) for school in their school track recommendation. Because the transition from elementary to secondary school involves numerous structural, curricular, didactic, and social changes (Tobbell & O'Donnell, 2013), scholastic performance (e.g., Ball, Lohaus, & Miebach 2006), competence self-concept, affect self-concept, and intrinsic motivation typically decrease (for an overview see Wigfield, Eccles, & Pintrich, 1996). However, cognitive abilities and motivation should be considered jointly to predict scholastic success in secondary school to avoid wrong school-related decisions at the expense of, for example, student's motivation besides the effects of school transition itself. Overall, Guay et al. (2003) emphasized that "teachers should strive to simultaneously improve both academic self-concept and achievement in order to produce positive changes in both constructs" (p. 134).

Concerning further research on the prediction of reported grades, longitudinal data have the advantage of being able to integrate subject-specific prior knowledge (in terms of prior reported grades) into the prediction models. Although it is easier to achieve higher levels of prior knowledge for more intelligent students, prior knowledge and intelligence cannot be treated as one variable (Heller, 1997; Helmke & Weinert, 1997). Subject-specific knowledge is often organized hierarchically; in these cases, prior knowledge is a mandatory requirement for the acquisition of new knowledge and, therefore, to be successful in the educational system (Klauer & Sparfeldt, 2018). Evidence for the importance of prior knowledge for subsequent scholastic achievement in elementary school students has been found repeatedly (e.g., Ditton & Krüsken, 2006, 2009b; Schneider, Stefanek, & Dotzler, 1997; Stern, 2003; Storch & Whitehurst, 2002; Weinert & Helmke, 1995). Moreover, report card grades are a conglomerate of several single reported grades, containing classroom examinations, homework assignments, or oral examinations. Thus, teachers might already indirectly assess effects of motivation via reported grades, because students have to work constantly throughout the school year to receive a good report card grade which, in fact, needs higher levels of motivation. Therefore, increments of competence self-concepts beyond prior knowledge (in terms of reported grades that already include motivational aspects) might be of small magnitude. For example, the study by Helmke and van Aken (1995) regarding the causal ordering of

prior knowledge (in terms of prior reported grades or test scores) and competence self-concepts in mathematics indicated significant, small effects of competence self-concepts on scholastic achievement beyond prior knowledge. Guay et al. (2003) and Muijs (1997) also reported significant, small effects of a global academic self-concept on a global scholastic achievement measure beyond prior knowledge (prior scholastic achievement) in elementary school students. Concerning affect self-concepts, cross-legged studies did not evidence substantial effects of affect self-concepts (or closely related constructs) on scholastic achievements beyond prior knowledge, so far (von Maurice et al., 2014; Skaalvik and Valås, 1999; Weidinger et al., 2015).

4.3 Critical acknowledgements

This subsection critically reflects upon potential limitations of the studies at hand. It is deivided into three subsections pertaining to the design of the dissertation project and the chosen instruments to assess scholastic achievements, academic self-concepts, and intelligence. Moreover, methodological issues of the employed analysis strategies are considered.

4.3.1 Design of the dissertation project

Sample. The sample of this dissertation project consisted of N = 1207 first to fourth graders (grade 1/2/3/4: n = 349/253/321/284). Non-participation (n = 394 of an initial sample of N = 1609) spread evenly across schools and grades. The participation rate (77%) was in line with or even exceeded the participation rate of other studies with elementary school students. Eight students were excluded from the analyses due to an unreasonably low intelligence score. The mean age per grade (grade 1/2/3/4: M[SD] = 7.03[.44] / 8.06[.51] / 9.04[.47] / 10.06[.50] years) and the gender distribution (n = 617 females, n = 588 males, n = 2 without gender specification) were as expected in the German school system and also comparable to other national and international studies with elementary school children (e.g., Ehm, Lindberg et al., 2014; Lohbeck & Möller, 2017; Marsh & Ayotte, 2003).

The data collection took place in two consecutive years because a sufficient sample size could not be realized within the short time slot at the end of the school year in 2014 (Lower Saxony, Saxony-Anhalt). To ensure that data of all students were comparable regarding the time of the school year, further schools were recruited and tested in June and July 2015 (Saarland). Both subsamples (sample 2014 vs. sample 2015) were comparable in terms of gender distribution (χ^2[1,1205] = .02, p = .90, $|h|$ = 0.01; Cohen's h is a measure of distance between two proportions, small/medium/large effect: $|h| \geq$ 0.20/0.50/0.80; Cohen, 1988, p. 184–185), age (t[1191] = 0.09, p = .93, d = -0.01), their actual reported grades (mathematics: t[1205] = 1.23, p = .24, d = -0.09; German:

t[1204] = 0.57, p = .57, d = -0.05), academic self-concepts (mathematics competence self-concept: t[1110] = -0.75, p = .46, d = 0.05; mathematics affect self-concept: t[1147] = -0.08, p = .94, d = 0.01; reading competence self-concept: t[1120] = -1.42, p = .16, d = 0.09; reading affect self-concept: t[1143] = 0.28, p = .78, d = -0.02), and non-verbal intelligence (t[1207] = 0.49, p = .62, d = -0.03).

Regarding sample size, power analyses with conventional programs like G*Power would not have been adequate for the described analyses due to the hierarchical data structure (students in classes) and the nature of some research aims (testing path coefficients against each other instead of against zero). A simulation study on structural equation models (Wolf, Harrington, Clark, & Miller, 2013) for models with three latent factors (indicated by three to four items), factor loadings of $\lambda = .65$, and a power of $\beta \geq .80$ (for all parameters of interest: factor loadings, correlations, regression paths; statistical comparison of parameters against zero with $\alpha = .05$; a model in which any one of the parameters fell below 80% power would be rejected) reported a minimum sample size of $N \approx 225$ per grade. Higher factor loadings go in hand with lower minimum sample sizes. For example, for a model with three latent factors (indicated by three to four items) and factor loadings of $\lambda = .80$, a minimum sample size of $N = 150$ per grade level would be required. Although the hierarchical data structure was not considered in the simulation study, the results can be used as rough proxies for comparisons of structural equation model parameters against zero. The study by Wolf et al. (2013) did not focus on the comparison of regression path coefficients within one model. To the best of our knowledge, no such (simulation) study exists so far. However, a corresponding sample size per grade was accomplished when considering students' z-standardized responses (studies 2 and 3).

Cross-sectional data. When interpreting the results of this thesis, the cross-sectional data structure must be considered. The data only provided information about the relations between variables and the amount of explained variances at a certain point of time. The results of studies 2 and 3 cannot be interpreted as evidence for causal ordering of reported grades and academic self-concepts. Moreover, path coefficients cannot be compared between models based on longitudinal versus cross-sectional data. With longitudinal data, relations will probably be of smaller magnitude due to possible predictor-unrelated changes in the criterion.

Gathering longitudinal data would have been desirable to provide further insight in the causal ordering of scholastic achievement and academic self-concepts. For example, the reciprocal internal/external-frame-of-reference model (RI/EM; see Möller et al., 2011; Möller et al., 2014) might provide information about longitudinal comparison processes by combining the I/E model and the reciprocal effects model. The latter suggests that subject-specific scholastic achievements and corresponding academic self-

concepts are reciprocally related and allows stronger interpretations about the ordering of variables in comparison to cross-sectional data (see subsection 2.3.1). However, these previous studies did not consider the important competence–affect self-concept distinction. Correspondingly, extended RI/EM results with elementary school students might extent the findings of study 2 and help to explain especially the role of dimensional comparison processes in that age. Furthermore, temporal comparison processes could be examined in longitudinal data (e.g., Albert, 1977; Dickhäuser & Galfe, 2004; Lüdtke & Köller, 2002; Müller-Kalthoff, Helm, & Möller, 2017).

For this dissertation project, a cross-sectional approach was chosen to provide first evidence for developmental processes across elementary school grades regarding relations between scholastic achievement and academic self-concepts and, thereby, to define a more precise period of time in which these developmental processes might be expected. The results of studies 2 and 3 did not indicate statistically significant grade-related differences in the examined relations between scholastic achievements and academic self-concepts (except for the path coefficient from the reported grade in mathematics on the corresponding competence self-concept in study 2).

4.3.2 Instruments

Scholastic achievements, competence self-concepts, affect self-concepts, and intelligence were assessed with the same instruments in all examined elementary school grades to ensure meaningful comparisons of uni- and bidirectional relations between constructs among grades (construct and measurement invariance). Therefore, analyses across grades have been interpreted in a straightforward manner. Otherwise, differences between respective grades in the relation between constructs on the one hand and differences of the operationalizations and constructs on the other hand could have been confounded. However, the adequacy of the administered instruments needs to be discussed.

Assessment of scholastic achievement. Scholastic achievement (and therefore, scholastic success) is typically operationalized by reported grades or standardized competence tests (Steinmayr et al., 2014). Both indicators correlate substantially ($.40 \leq r \leq .72$; e.g., Helmke & van Aken, 1995; Krüsken, 2007; Lorenz & Artelt, 2009; Marsh, 2007; Marsh et al., 2005), but form distinct measures. In this thesis, scholastic achievement was indicated by reported grade equivalents in grade 1 and reported grades in grades 2 to 4. In contrast to standardized competence tests, reported grades are the main criterion to evaluate scholastic success in German (elementary) schools and operate in manifold educational functions (e.g., Beutel et al., 1999; Birkel & Tarnai, 2018; Heine et al., 2006; Willingham et al., 2002). For example, reported grades (and not the results of standardized competence tests) at the end of grade 4 are relevant for the decision for a secondary school track. In contrast, competence test results are of less importance in the

German educational system (Füssel & Leschinsky, 2008). The own test results might not even be known by most of the students.

Regarding results of studies 2 and 3 that focused on the relations between scholastic achievements and academic self-concepts, different results could have been expected for standardized competence tests as indicators of scholastic achievement. Prior research indicated that tests scores were less related to corresponding self-concepts compared to reported grades (Marsh, 2007; Marsh et al., 2014; Marsh & Hau, 2004; Möller et al., 2009). Therefore, less pronounced path coefficients from test scores on competence and affect self-concepts and vice versa could be assumed: Concerning scholastic achievement–academic self-concept relations within the I/E model framework (study 2), students might often not be informed about their competence test results or the test results of their classmates. Correspondingly, the opportunity for social and dimensional comparison processes on the basis of test results might be limited (Marsh, 2007). In their meta-analysis, Möller et al. (2009) found that the type of achievement indicator (reported grades vs. standardized competence tests) was a substantial moderator for the relations between scholastic achievements and self-concepts within the I/E model framework. As expected, path coefficients from test scores on corresponding self-concepts were less pronounced compared to path coefficients relating reported grades and self-concepts. No differences between achievement indicators were found for path coefficients on non-corresponding self-concepts. However, according to the authors, the moderating effect of the achievement indicators was not caused by differing path coefficients on self-concepts, but by the higher correlation between test scores in mathematics and the native language compared to the correlation between corresponding reported grades. Subsequently, they assumed that competence tests tend to measure more global ability components and are, therefore, less domain-specific than reported grades.

As mentioned, students might often not know their own competence test results or the test results of their classmates. Nevertheless, substantial path coefficients from test scores in mathematics and the native language on corresponding competence self-concepts indicating social comparison processes were repeatedly reported in elementary school students (e.g., Ehm, Lindberg et al., 2014; Ehm, Nagler et al., 2014; Lohbeck & Möller, 2017; Marsh, Smith, Barnes, & Butler, 1983; Pinxten et al., 2015; Poloczek et al., 2011; Skaalvik & Valås, 2011). Various feedback sources like reported grades, oral feedback by teachers and classmates as well as observations in class might be used for social comparison processes (thereby forming corresponding competence and affect self-concepts). Constant comparisons with others should result in stable perceptions of one's own abilities and in the perception of stable class rank order in a school subject. This rank order is likely to emerge in competence tests as well (e.g., due to the substantial

correlations with reported grades) and should result in substantial path coefficients from test scores on corresponding self-concepts.

For the prediction of scholastic achievements by cognitive and non-cognitive variables (study 3), standardized competence tests seem to be a purer measure of students' competences and less influenced by non-cognitive aspects than reported grades (Baumert, Lüdtke, Trautwein, & Brunner, 2009; Harlen, 2005). Consequently, higher path coefficients from cognitive variables on corresponding test scores and lower path coefficients from, for example, motivational variables on test scores in comparison to reported grades were reported for secondary school students (e.g., Jansen et al., 2016; Jansen, Schroeders, & Lüdtke, 2014; Lotz et al., 2018; Marsh et al., 2005; Steinmayr & Meißner, 2013). The proportion of overall explained variance by cognitive and non-cognitive variables in test scores was of smaller magnitude compared to reported grades. Comparable results can be expected for elementary school students. However, with regard to the relevance of such anticipated research focusing on the prediction of standardized competence test results, competence test scores are of very small relevance in the German educational system for educational decisions (Füssel & Leschinsky, 2008). Hence, research regarding the prediction of reported grades and corresponding interventions to improve reported grades might be more important.

Assessment of academic self-concepts. For assessing competence and affect self-concepts, the well-established SDQ I (Marsh, 1992) was employed. Originally, this questionnaire was designed for students between 8 and 12 years of age (typically corresponding to grades 3 to 7). Nevertheless, research results indicated that the SDQ I or the short version of the SDQ I can also be used with children from five years of age upwards (e.g., Lohbeck & Möller, 2017; Marsh, Barnes, Cairns & Tidman, 1984; Marsh et al., 1991). Reliability estimates for the competence and affect self-concept scales in mathematics and reading in this thesis were at least acceptable ($\alpha \geq .75$, see Tab. 5 in this thesis; section 3.2) – independently of the response format (three-, four-, or five-point Likert-type scale) or elementary school grade.

Furthermore, the SDQ I comprised items to assess competence and affect self-concepts in reading instead of a more global native language self-concept. Therefore, analyses in studies 2 and 3 were based on competence and affect self-concepts in reading on the one hand side and reported grades in German on the other hand side. Both do not align perfectly with each other. Research with German fifth and sixth graders emphasized that reading and native language self-concept items should not be treated as equivalent indicators for students' verbal self-concepts (Arens et al., 2014). Whereas reading self-concept items seem to reflect students' self-perceptions in the specific domain of reading, the self-concept items for German seem to reflect a more global level of the verbal domain. Results of the study by Arens et al. (2014) can be aligned with the "specificity

matching principle" (Swann, Chang-Schneider, & McClarty, 2007) or the "principle of symmetry" (Wittmann, 1988). These principles assume that an appropriate match between self-concept and other measures (e.g., reported grade) regarding their content domains and specificity level would result in relatively higher relations between such variables compared to non-matching variables. Contrarily, non-matching measures should weaken the relations between variables. However, Arens et al. (2014) noted that reading instruction might play a more important role in the curriculum in the school subject German for elementary school students than secondary school students. In elementary school, learning to read is an essential educational goal whereas in secondary school, reading itself is a prerequisite for carrying out more complex tasks like poetry analyses. Correspondingly, they assumed that the self-concept structure for the native language might be characterized by a more prominent reading self-concept in comparison to an accompanying less important global German self-concept in elementary school students.

Overall, it needs to be a goal of future research to examine to what extent the assessment of reading versus German competence and affect self-concepts might affect the relations between scholastic achievements and corresponding self-concept measures in elementary school students. For example, the results of study 3 revealed that the proportion of totally explained reported grade variance by intelligence, corresponding competence self-concepts, and affect self-concepts was higher in mathematics ($.54 \leq R^2 \leq .55$) compared to German ($.25 \leq R^2 \leq .37$) – independently of the respective grade. However, the increment of intelligence explained a comparable amount of reported grade variance in both subjects across all grades. Therefore, differences in the totally explained variances between subjects might be assignable to the lower symmetry for German/reading in contrast to mathematics. Especially the competence self-concept increment could have been of higher magnitude for German (assumingly of the same magnitude as the competence increment of mathematics) compared to reading. For the affect self-concept, no such differential results for both subjects might be expected because affect self-concept increments were already of (very) small magnitude (e.g., $.03 \leq \Delta R^2 \leq .07$ for mathematics).

Assessment of intelligence. In this dissertation project, intelligence was assessed with tasks of the German adaptation of the Culture Fair Intelligence Test (CFT 1–R; Weiß & Osterland, 2013) in all elementary school grades. Research pointed towards the relevance of a broad intelligence operationalisation, for example, with verbal, numerical, and figural subtests when examining relations between intelligence and other variables (e.g., Jensen & Weng, 1994; Johnson, Nijenhuis, & Bouchard, 2008; Reeve, 2004, Roth et al., 2015). However, because of time restrictions for the data assessment, it was not possible to administer the complete CFT 1-R. Therefore, only the three subtests covering figural reasoning (series, classification, and matrices) were administered. According to

Lohmann & Lakin (2011), figural reasoning tests are good markers of intelligence. Furthermore, reasoning tests provide an effective and efficient instrument for assessing intelligence in groups of elementary school children.

4.3.3 Methodological issues

In this thesis, a variety of advanced methods like measurement invariance testing or the Cholesky factoring approach were applied. All strategies were specifically tailored to the nature of the corresponding research aims. Within the following subsection, some key features of the employed analyses strategies are outlined.

Standardization of responses regarding academic self-concepts. One major issue of this dissertation project was the assessment of academic self-concepts with different response formats regarding the number of response categories. For this thesis with a focus on the relations between scholastic achievements and academic self-concepts, analyses regarding response formats played a subordinate role. However, metric measurement invariance across response formats indicated a comparable content-related meaning for competence and affect self-concepts in mathematics and reading across response formats in all grades (section 3.2). Thus, no fundamentally different data regarding self-concepts across response formats were provided. Based on these results, *z*-standardized data (standardized per response format and grade) from different response scales were combined to form one sample per grade. Furthermore, with *z*-standardization the data could be regarded as continuous and therefore, met the criteria for the MLR estimator.

Measurement invariance testing across grades. In both studies examining the relations between scholastic achievements and academic self-concepts, grade-related differences were inspected. An important strength of this dissertation project was the identical assessment of scholastic achievements, academic self-concepts, and intelligence across all elementary school grades. Due to construct and measurement equivalence, relations between variables can be compared meaningfully across elementary school grades. As a further prerequisite for comparisons across grades, the presence of measurement invariance across grades was examined. Because no mean-related research questions were pursued, metric measurement invariance was deemed to be a sufficient measurement invariance level for our analyses (see Christ & Schlüter, 2012; Geiser, 2013). As decision criteria, we relied on ΔCFI as the main criterion for inspecting measurement invariance and $\Delta\chi^2$ (with Δdf). ΔCFI < .01 (Chen, 2007) as well as non-significant χ^2-difference tests between two differentially restricted models (e.g., configural vs. metric) would indicate the presence of measurement invariance. We preferred ΔCFI over $\Delta\chi^2$ (with Δdf), because χ^2-difference tests seem to be very sensitive to large sample sizes and violations of the normality assumption. As a consequence, trivial discrepancies might lead to the rejection of a model comparison (see Chen, 2007).

Examination of unique variances. Regarding the prediction of reported grades in mathematics and German (study 3), unique variance proportions of intelligence, competence self-concept, and affect self-concept were examined using the Cholesky factoring approach (de Jong, 1999; Loehlin, 1996). One advantage of hierarchical regression analyses with more than two predictors with the Cholesky approach was the efficient examination of increments within only one model for each increment. Moreover, the latent analyses of unique effects could be conducted without altering the model fit or affecting the measurement part of the model. Furthermore, squared path coefficients from the Cholesky factors on reported grades could be straightforwardly interpreted as explained variances. In study 3, the Cholesky approach was also utilized in models with only two predictors instead of established residual models. We chose this approach to be consistent with regard to statistical methods within one study. However, both the Cholesky models and the corresponding residual models would yield identical results regarding the prediction patterns (e.g., path coefficients of predictors on reported grades as well as increments) in study 3 within models with two predictors.

4.4 Final conclusion

The aim of this thesis was the examination of relations between subject-specific scholastic achievements and academic self-concepts. To achieve this aim, attention was firstly paid to the assessments of reported grades and academic self-concepts in elementary school students. Regarding the assessment of reported grades (study 1), student-reported grades were less accurate indicators of their actual reported grades in elementary school grades 2 and 3 compared to older students, but reliable and valid indicators in grade 4. Concerning the assessment of competence and affect self-concepts (in the excursus), rating scales with three, four, or five response categories revealed equivalent data in all elementary school grades.

Subsequently, two empirical studies revealed new insights into the interplay of the examined variables separately for elementary school grades and across grades. Special was paid to the well-established competence–affect distinction of subject-specific academic self-concepts. Competence and affect self-concepts showed very large construct overlaps within school subjects in all elementary school grades. As expected, relations between reported grades and corresponding competence self-concepts were more pronounced compared to those between reported grades and affect self-concepts in both studies. This underlines the higher conceptual proximity between scholastic achievement and self-perceived scholastic competence (competence self-concept) compared to intrinsic motivational-affective perceptions (affect self-concept).

In study 2, we examined effects of scholastic achievements in mathematics and German on (non-)corresponding competence and affect self-concepts within the I/E model framework. Substantial social comparison processes of moderate to high magnitude were found in all four elementary school grades. Path coefficients between reported grades and non-corresponding competence and affect self-concepts indicating dimensional comparisons were mostly negative, only some path coefficients were substantial. In study 3, the focus was shifted towards subject-specific competence and affect self-concepts as statistical predictors of scholastic achievement beyond intelligence. Competence and affect self-concept contributed substantially to the prediction of reported grades in mathematics and German beyond intelligence in the examined elementary school grades (with the exception of affect self-concept of fourth graders in German). Among the two self-concept components, competence self-concept was the stronger predictor.

Because of the identical assessment of scholastic achievements and academic self-concepts (as well as intelligence as a further variable in study 3) in all elementary school grades, grade-related differences in the relation between variables could be examined. Although academic self-concepts seem to become more realistic across elementary school grades (Helmke, 1999; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002) and, therefore, more highly correlated with scholastic achievements, mostly no substantial differences in the relations between reported grades and academic self-concepts were found across elementary school grades in studies 2 and 3.

To sum up, the results of present thesis deepened the understanding about the interplay of scholastic achievements and academic self-concepts in elementary school students. Thereby, the extraordinary importance of distinguishing between competence and affect self-concepts was emphasized when examining relations between subject-specific scholastic achievements and (non-)corresponding academic self-concepts. Strong support for substantial relations between reported grades and both self-concept components within school subjects was provided – regardless of the examined elementary school grade. Teachers should strive to improve students' fundamental competencies like basic arithmetic operations, reading, and writing as well as to simultaneously improve students' competence and affect self-concepts in order to increase the possibility for good scholastic achievements.

5 References

Adelson, J. L., & McCoach, D. B. (2010). Measuring the mathematical attitudes of elementary students: the effects of a 4-point or 5-point likert-type scale. *Educational and Psychological Measurement*, *70*, 796–807.

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*, 485–503.

Arens, A. K., & Hasselhorn, M. (2015). Differentiation of competence and affect self-perceptions in elementary school students: Extending empirical evidence. *European Journal of Psychology of Education*, *30*, 405–419.

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*, 621–634.

Arens, A. K., Trautwein, U., & Hasselhorn, M. (2011). Erfassung des Selbstkonzepts im mittleren Kindesalter: Validierung einer deutschen Version des SDQ I [Self-concept acquisition for middle aged children: Validation of a German version of the SDQ I]. *Zeitschrift für Pädagogische Psychologie*, *25*, 131–144.

Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, *103*, 970–981.

Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2013). A short German version of the Self-Description Questionnaire I: Theoretical and empirical comparability. *International Journal of Research & Method in Education*, *36*, 415–438.

Arens, A. K., Yeung, A. S., & Hasselhorn, M. (2014). Native language self-concept and reading self-concept: Same or different? *The Journal of Experimental Education*, *82*, 229–252.

Baldering, D. (1993). *Selbstkonzepte von Kindern im Grundschulalter* [Self-concepts of elementary school-aged children]. Frankfurt a. M., Germany: Lang.

Ball, J., Lohaus, A. & Miebach, C. (2006). Psychische Anpassung und schulische Leistungen beim Wechsel von der Grundschule zur weiterführenden Schule [Psychological adjustment and school achievement during transition from

elementary to secondary school]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*, 101–109.

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measre the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review, 4*, 165–176.

Baumert, J., Schnabel, K., & Lehrke, M. (1998). Learning math in school: Does interest really matter? In L. Hoffmann, A. Krapp., K. A. Renninger, & J. Baumert (Eds.), *Interest and learning, proceedings of the Seeon conference on interest and gender* (pp. 327–336). Kiel, Germany: Institut für die Pädagogik der Naturwissenschaften.

Baumert, J., Trautwein, U., & Artelt, C. (2003). Schulumwelten–institutionelle Bedingungen des Lehrens und Lernens [School environments–Institutional conditions for learning and instruction]. In J. Baumert, C. Artelt, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele et al. (Hrsg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* [PISA 2000] (pp. 261 – 331). Opladen, Germany: Leske und Budrich.

Bayrisches Staatsministerium für Bildung und Kultus, Wissenschaft und Kunst (2017). *Die bayrische Grundschule*. Retrieved from https://www.km.bayern.de /download/11862_stmbw_die_bayerische_grundschule_2018_web_bf.pdf

Beutel, S.-I., Lütgert, W., Tillmann, K.-J., & Vollstädt, W. (1999). *Ermittlung und Bewertung schulischer Leistungen* [Assessment and evaluation of scholastic achievement]. Hamburg, Germany: Behörde für Schule, Jugend und Berufsbildung.

Birkel, P., & Tarnai, C. (2018). Zensuren und verbale Schulleistungsbeurteilung [Grades and verbal evaluations of scholastic achievement]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., pp. 904–917). Weinheim, Germany: Beltz.

Black, R. (2012). The dissertation marathon. *Contemporary Issues in Educational Research, 5*, 97–104.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*, 1–40.

Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* [Statistics for human and social sciences] (6th ed.). Heidelberg: German: Springer.

Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Eds.). (2004). *IGLU. Einige Länder der Bundesrepublik Deutschlands in nationalen und internationalen Vergleich* [IGLU. National and international comparison of some federal states in Germany]. Münster, Germany: Waxmann.

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102,* 964–981.

Buckelew, S. P., Byrd, N., Key, C. W., Thornton, J., & Merwin, M. M. (2013). Illusions of a good grade: Effort or luck? *Teaching of Psychology, 40,* 131–138.

Byrne, B. M. (1996). *Measuring self-concept across life span: Issues and Instrumentation*. Washington, DC: American Psychological Association.

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36.

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology, 67*, 279–291.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.

Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: a multicohort–multioccasion study. *Learning and Individual Differences, 23*, 172–178.

Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung* [Structural equation models with Mplus. A practical introduction]. München, Germany: Oldenbourg.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement, 34*, 35-46.

Craven, R. G., & Marsh, H. W. (2008). The centrality of the self-concept construct for psychological well-being and unlocking human potential: Implications for child and educational psychologists. *Educational and Child Psychology, 25*, 104–118.

de Jong. P. E. (1999). Hierarchical regression analysis in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 198–211.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Los Angeles, CA: Sage.

Dickhäuser, O., & Galfe, E. (2004). Besser als..., schlechter als... Leistungsbezogene Vergleichsprozesse in der Grundschule [Better …, worse … Achievement related comparison processes in elementary school]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36*, 1–9.

Dickhäuser, O., & Plenter, I. (2005). “Letztes Jahr stand ich zwei.“ Zur Akkuratheit selbstberichteter Zensuren [On the accuracy of self-reported school marks]. *Zeitschrift für Pädagogische Psychologie, 19*, 219–224.

Diersch, N., & Walther, E. (2010). Umfrageforschung mit Kindern und Jugendlichen [Survey research with children and adolscents]. In E. Walther, F. Preckel, & S. Mecklenbräuker (Eds.), *Befragung von Kindern und Jugendlichen* [Survey of children and preadolescents] (pp. 297–318). Göttingen, Germany: Hogrefe.

Ditton, H., & Krüsken, J. (2006). Der Übergang von der Grundschule in die Sekundarstufe I [The transition from primary to secondary schools]. *Zeitschrift für Erziehunsgwissenschaft, 9*, 348–372.

Ditton, H., & Krüsken, J. (2009a). Bildungslaufbahnen m differenzierten Schulsystem – Entwicklungsverläufe von Laufbahnempfehlungen und Bildungsaspirationen in der Grundschulzeit [Educational careers in a tracked school system – Development of teacher recommendations and educational aspirations over the elementary school years]. *Zeitschrift für Erziehungswissenschaften*, *12*, 74–102.

Ditton, H., & Krüsken, J. (2009b). Denn wer hat, dem wird gegeben werden? Eine Längsschnittstudie zur Entwicklung schulischer Leistungen und den Effekten der sozialen Herkunft in der Grundschulzeit [To those who have, will more be given? A longitudinal study concerning the development of school achievement and the effects of social background during primary school]. *Journal für Bildungsforschung Online*, *1*, 33–61.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75–146). San Francisco, CA: Freeman.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents’ achievement task value and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*, 215–225.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children’s self- and task perceptions during elementary school. *Child Development*, *64*, 830–847.

Ehm, J.-H., Lindberg, S., & Hasselhorn, M. (2014). Reading, writing, and mathematics self-concept in elementary school children: Influence of dimensional comparison processes. *European Journal of Psychological Education*, *29*, 277–294.

Ehm, J.-H., Nagler, T., Lindberg, S., & Hasselhorn, M. (2014). Dimensionale Vergleichseffekte zwischen Lesen, Rechtschreiben und Rechnen. Eine Erweiterung des I/E-Modells für die Grundschule [Dimensional comparison effects between

reading, spelling and mathematics. An extension of the I/E-model for elementary school]. *Zeitschrift für Pädagogische Psychologie*, *28*, 51–56.

Faber, G. (1992). Bereichsspezifische Beziehungen zwischen leistungsthematischen Schülerselbstkonzepten und Schulleistungen [Domain-specific relations between academic self-concepts and academic achievements]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *24*, 66–82.

Faber, G. (2007). Academic self-beliefs in the spelling domain: empirical research findings on elementary school students' subject-specific self-concept, causal attributions, and test anxiety. In E. M. Vargios (Ed.), *Educational Psychology Research Focus* (pp. 65–120). New York, NJ: Nova Science Publishers.

Feng, X., & Rost, D. H. (2015). Selbstberichtete Zeugnisdaten: Weitere Evidenz für eine (partielle) Brauchbarkeit [Self-reported grades: Further evidence for their (partial) usability]. *Psychologie in Erziehung und Unterricht*, *19*, 219–224.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage.

Füssel, H.-P., & Leschinsky, A. (2008). Der institutionelle Rahmen des Bildungswesens [The institutional framework of the education system]. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer, & L. Trommer (Eds.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* [The German education system. An overview of structures and developments] (pp. 131–203). Reinbeck, Germany: Rowohlt.

Geiser, C. (2013). *Data analysis with Mplus*. New York, NY: The Guilford Press.

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33*, 9–33.

Gölitz, D., Roick, T., & Hasselhorn, M. (2004). *DEMAT 4+: Deutscher Mathematiktest für vierte Klassen* [DEMAT 4+: German test of mathematics for fourth grade]. Weinheim, Germany: Beltz.

Guay, F., Larose, S., & Boivin, M. (2004). Academic self-concept and educational attainment level: a ten-year longitudinal study. *Self and Identity*, *3*, 53–68.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, *95*, 124–136.

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences*, *20*, 644–653.

Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education, 20,* 245–270.

Harter S. (1985). *Manual for the Self-Perception Profile for Children*. Denver; CO: University of Denver.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York, NY: Guilford.

Harter, S. (2006). The self. In N. Eisenberg (Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, amd personality development* (pp. 505–570). New York, NY: Wiley.

Harter S. (2012). *Self-Perception Profile for Children: Manual and questionnaires (Grades 3–8)*. Denver, CO: University of Denver.

Hasselhorn, M., & Gold, A. (2009). *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren* [Educational Psychology: Successful learning and teaching]. Stuttgart, Germany: Kohlhammer.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: NY: Routledge.

Heine, C., Briedis, K., Didi, H.-J., Hasse, K., & Trost, G. (2006). *Bestandsaufnahme von Auswahl- und Eignungsfeststellungsverfahren beim Hochschulzugang in Deutschland und ausgewählten Ländern* [Student admission in Germany and selected other countries]. Hannover, Germany: HIS–Kurzinformation A 3/2006.

Heller, K. A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literatur-überblick [Individual determinantes of academic performance: Literature review]. In F. E Weinert, & A. Helmke (Eds.), *Entwicklung im Grundschulalter* [Development in elementary school age] (pp. 183–201). Weinheim, Germany: Beltz.

Heller, K. A. & Hany, E. A. (2001). Standardisierte Schulleistungsmessungen [Standardized measurement of academic achievment]. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* [Academic achievement measurement in schools] (pp. 87–101). Weinheim, Germany: Beltz.

Helmke, A. (1997a). Entwicklung lern- und leistungsbezogener Motive und Einstellungen: Ergebnisse aus dem SCHOLASTIK-Projekt [Development of learning- and achievement-related motives and attitudes]. In F. E. Weinert, & A. Helmke (Eds.). *Entwicklung im Grundschulalter* [Development in elementary school age] (pp. 59–76). Weinheim, Germany: Beltz.

Helmke, A. (1997b). Individuelle Bedingungsfaktoren der Schulleistung: Ergebnisse aus dem SCHOLASTIK-Projekt [Individual conditions of school achievement]. In F. E.

Weinert, & A. Helmke (Eds.). *Entwicklung im Grundschulalter* [Development in elementary school age] (pp. 203–216). Weinheim, Germany: Beltz.

Helmke, A. (1999). From optimism to realism? Development of children's academic self-concept from kindergarten to grade six. In W. Schneider, & F. E. Weinert (Eds.), *Individual development from 3 to 12. Findings from the Munich Longitudinal Study* (pp. 198–221). Cambridge, UK: Cambridge University Press.

Helmke, A., & Schrader, F.-W. (2018). Determinanten der Schulleistung. [Determinantes of scholastic achievement]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., pp. 90–102). Weinheim, Germany: Beltz.

Helmke, A., & van Aken, M. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, *87*, 624–637.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen [Conditions of scholastic achievement]. In F. E. Weinert, C. F. Graumann, & N. Birbaumer (Eds.), *Psychologie des Unterrichts und der Schule* [Psychology of teaching and school] (pp. 71–176). Göttingen, Germany: Hogrefe.

Ingenkamp, K. (1971). *Die Fragwürdigkeit der Zensurengebung* [The dubiousness of school grades] (1st ed.). Weinheim, Germany: Beltz.

Ingenkamp, K. (1995). *Die Fragwürdigkeit der Zensurengebung* [The dubiousness of school grades] (9th ed.). Weinheim, Germany: Beltz.

Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der Pädagogischen Diagnostik* [Textbook of pedagogical assessment]. Weinheim, Germany: Beltz.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: gender and domain differences across grades one through twelve. *Child Development*, *73*, 509–527.

Jäger, R. S., & Lissmann, U. (2004). *Von der Beobachtung zur Notengebung. Ein Lehrbuch* [From observarion to grades. A textbook]. Landau, Germany: Verlag Empirische Pädagogik.

Jansen, M., Lüdtke, O., & Schroeders, U. (2016). Evidence for a positive relation between interest and achievement: Examining between-person and within-person variation in five domains. *Contemporary Educational Psychology*, *46*, 116–127.

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, *30*, 11–21.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jensen, A. R., & Weng, L.-J. (1994). What is a good g? *Intelligence*, *18*, 231–258.

Johnson, W., Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 *g*: Consistent results

from five test batteries. *Intelligence, 36*, 81–95.

Kadir, M. S., Yeung, A. S., & Diallo, T. M. O. (2017). Simultaneous testing of four decades of academic self-concept models. *Contemporary Educational Psychology, 51*, 429–446.

Karbach, J., Schmiedek, F., & Hasselhorn, M. (2018) Kognitives Training [Cognitive training]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., 345–353). Weinheim, German: Beltz.

Klauer, K. J., & Sparfeldt, J. R. (2018). Intelligenz und Begabung [Intelligence and giftedness]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., 278–286). Weinheim, German: Beltz.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interst and achievement in mathematics. *Journal of Research in Mathematics Education*, *32*, 448–470.

Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse [Course selection of girls and boys in mathematics: The role of academic self-concept and interest]. *Zeitschrift für Pädagogische Psychologie, 14*, 26–37.

Köller, O., Zeinz, H., & Trautwein, U. (2008). Class-average achievement, marks, and academic self-concept in German primary schools. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self-processes, learning, and enabeling human potential* (p. 331–352). Charlotte, NC: Age Publishing.

Krajewski, K., Küspert, P., & Schneider, W. (2002). *DEMAT 1+: Deutscher Mathematiktest für erste Klassen* [DEMAT 1+: German test of mathematics for first grade]. Weinheim, Germany: Beltz.

Krajewski, K., Liehm, S., & Schneider, W. (2004). *DEMAT 2+: Deutscher Mathematiktest für zweite Klassen* [DEMAT 2+: German test of mathematics for second grade]. Weinheim, Germany: Beltz.

Krapp, A. (2000). Interest and human development during adolescence: An educational-psychological approach. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 109–128). London, UK: Elsevier.

Kriegbaum, K., Jansen, M., & Spinath, B. (2015). Motivation: A predictor of PISA's mathematical competence beyond intelligence and prior test achievement. *Learning and Individual Differences, 43*, 140–148.

Krüsken, J. (2007). Entwicklung von Schülerleistungen und Zensuren in der Grundschule [Development of students' scholastic achievements and grades in elementary

school]. In H. Ditton (Ed.), *Kompetenzaufbau und Laufbahnen im Schulsystem: Ergebnisse einer Längsschnittuntersuchung an Grundschulen* [Competence development and career in the educational system: Results of a longitudinal study at elementary schools] (pp. 41–62). Münster, Germany: Waxmann.

Kultusministerkonferenz (2006). *Übergang von der Grundschule in die Schulen des Sekundarstufenbereichs I. Informationsunterlage* [Transition from elementary school to secondary school] Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2010/2010_10_18-Uebergang-GrundschuleS_eI1-Orientierungsstufe.pdf

Kultusministerkonferenz (2015). *Empfehlungen zur Arbeit in der Grundschule (Beschluss der Kultusministerkonferenz vom 02.07.1970 i. d. F. vom 11.06.2015)* [Recommendations for working in elementary schools]. Retrieved from https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/1970/1970_07_02_Empfehlungen_Grundschule.pdf

Kuncel, N. R., Credé, M. & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63–82.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86,* 148−161.

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, *17*, 494–509.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary school. *Personality and Individual Differences, 42,* 441–451.

Lenhard W., & Schneider, W. (2006). *ELFE 1–6: Ein Leseverständnistest für Erst- bis Sechstklässler* [ELFE 1–6: Reading comprehension test for grades 1–6]. Göttingen, Germany: Hogrefe.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45,* 79–122.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: Guilford.

Loehlin, J. C. (1996). The Cholesky approach: A cautionary note. *Behavior Genetics, 26,* 65–69.

Lohaus, A. (1989). *Datenerhebung in der Entwicklungspsychologie* [Data collection in developmental psychology]. Bern, Switzerland: Huber.

Lohbeck, A. & Möller, J. (2017). Social and dimensional comparison effects on math and

reading self-concepts of elementary school children. *Learning and Individual Differences, 54*, 73–81.

Lohmann, D. F., & Lakin, J. M. (2011). Intelligence and reasoning. In R. J. Sternberg, & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 419–441). Cambridge, UK: Cambridge University Press.

Lorenz, C., & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik [Domain specificity and stability of diagnostic competence among elementary school teachers in the school subjects of German and mathematics]. *Zeitschrift für Pädagogische Psychologie, 23*, 211–222.

Lotz, C., Schneider, R., & Sparfeldt, J. R. (2018). Differential relevance of intelligence and motivation for grades and competence tests in mathematics. *Learning and Individual Differences*, *65*, 30–40.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*, 73–79.

Lüdtke, O., & Köller, O. (2002). Individuelle Bezugsnormorientierung und soziale Vergleiche im Mathematikunterricht: Einfluss unterschiedlicher Referenzrahmen auf das fachspezifische Selbstkonzept der Begabung [Individual reference norm and social comparisons in mathematics classes: The impact of different frames of reference on the domain-specific self-concept of ability]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 34*, 156–166.

Mackintosh, N. J. (2011). *IQ and Human Intelligence* (2nd ed.). Oxford, UK: Oxford University Press.

Marsh, H. W. (1986). Verbal and mathematics self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*, 129–149.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*, 280–295.

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623–636.

Marsh, H. W. (1992). *Self-Description Questionnaire (SDQ) I: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept. A test manual and research monograph.* Sydney, Australia: University of Western Sydney Macarthur, Faculty of Education.

Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society.

Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology, 95*, 687–706.

Marsh, H. W., Barnes, J., Cairns, L., & Tidman, M. (1984). Self-Description Questionnaire: Age and sex effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology, 76*, 940–956.

Marsh, H. W., Byrne, B., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*, 366–380.

Marsh, H. W., & Craven, R. G. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131–198). Orlando, FL: Academic Press.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133–163.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*, 377–392.

Marsh, H. W., Craven, R. G., & Debus, R. (1999). Separation of competency and affect components of multiple dimensions of academic self-concept: A developmental perspective. *Merrill-Palmer Quarterly, 45*, 567–601.

Marsh, H. W., & Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56–67.

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries [Special issue]. *International Journal of Testing, 6*, 311–360.

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: An extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology, 39*, 326–341.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction, 35*, 16–32.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, *16*, 439–476.

Marsh, H. W., & O'Mara, A. J. (2008). Self-concept is as multidisciplinary as it is multidimensional. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 87–115). Charlotte, NC: Information Age.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*, 319–350.

Marsh, H. W., Smith, I. D., & Barnes, J. (1983). Multitrait-multimethod analyses of the self-description questionnaire: Student-teacher agreement of multidimensional ratings of student self-concept. *American Educational Research Journal*, *20*, 333–357.

Marsh, H. W., Smith, I. D., Barnes, J., & Butler, S. (1983). Self-concept: Reliability, stability, dimensionality, validity, and the measurement of change. *Journal of Educational Psychology*, *75*, 772–790.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects model of causal ordering. *Child Development*, *76*, 397–416.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, *89*, 41–54.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychologial Methods*, *12*, 23–44.

Mellor. D. & Moore, K. A. (2013). The use of likert scales with children. *Journal of Pediatric Psychology*, *39*, 369–379.

Ministerium für Bildung, Wissenschaft und Kultur Schleswig-Holstein (2015). *Zeugnisse in der Grundschule und Entwicklungsbericht zum Übergang an die weiterführenden allgemein bildenden Schulen* [Report cards in elementary school and development report for the transition to secondary school]. Retrieved from https://www.schleswig-holstein.de/DE/Fachinhalte/S/schulsystem/noten.html

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional comparisons and their consequences for self-concept, motivation, and emotion. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 430–436). Oxford, UK: Elsevier.

Möller, J., & Husemann, N. (2006). Internal comparisons in everyday life. *Journal of Educational Psychology*, *98*, 342–353.

Möller, J., Köller, O. (2000). Spontaneous and reactive attributions following academic achievement. *Social Psychology of Education*, *4*, 67–86.

Möller, J., Kuska, S. K., & Zaunbauer, A. M. C. (2011). Einflussfaktoren auf die Entwicklung des Selbstkonzepts im Grundschulalter [Factors influencing self-concept development in elementary school]. In F. Hellmich (Ed.), *Selbstkonzepte im Grundschulalter. Modelle, empirische Ergebnisse, pädagogische Konsequenzen* [Self-concepts in elementary school. Models, empirical results, educational consequences] (pp. 159–172). Stuttgart, Germany: Kohlhammer.

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*, 544–560.

Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N. & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research*, *4*, 1–11.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*, 1129–1167.

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal I/E model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal*, *48*, 1315–1346.

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology*, *84*, 591–611.

Moschner, B., & Dickhäuser, O. (2018). Selbstkonzept [Self-concept]. In D.H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds)., *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., pp. 750–756). Weinheim, German: Beltz.

Muijs, R. D. (1997). Predictors of academic achievement and academic self-concept: A longitudinal perspective. *British Journal of Educational Psychology*, *67*, 263–277.

Müller-Kalthoff, H., Helm, F., & Möller, J. (2017). The Big three of comparative judgment: On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Social Psychology of Education*, *20*, 840–973.

Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Muthén, L. K. (2011, November 16). Re: categorical indicators [Web log comment]. Retrieved from http://www.statmodel2.com/discussion/messages/23/106.html?1366853593

Muthén, L. K., & Muthén, B. O. (1998-2013). *MPlus User's Guide.* Los Angeles, CA: Muthén & Muthén.

Myers, R. (1990). *Classical and modern regression with applications* (2nd ed.). Boston, MA: Duxbury.

Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. S. (2008). Gendered high school course selection as a precursor of gendered careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt, & J. S. Eccles (Eds.), *Gender and occupational outcomes. Longitudinal assessments of individual, social, and cultural influences* (pp. 115–143). Washington, DC: American Psychological Association.

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper seondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*, 323–345.

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology, 106,* 1170–1191

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41,* 181–206.

Ostrop, G., Schmude, C., & Valtin, R. (2002). Was denken Kinder über ihre Zeugnisse? [What do children think about their report cards?]. In R. Valtin (Hrsg.), *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand* [What is a good report card? Grades and verbal evaluations on trial] (pp. 49–59). Weinheim, Germany: Juventa.

Pekrun, R. (2000). A social-cognitive, control-value theory of achievement emotions. In J. Heckhausen (Ed.), *Motivational psychology of human development* (143–163). Oxford, UK: Elsevier.

Pinxten, M., De Fraine, B., Van Damme, J., & D'Haenens, E. (2010). Causal ordering of academic self-concept and achievement: effects of type of achievement measure. *British Journal of Educational Psychology, 80,* 689–709.

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived mathematics effort expenditure. *British Journal of Educational Psychology, 84,* 152–174.

Pinxten, M., Wouters, S., Preckel, F., Niepel, C., De Fraine, B., & Verschueren, K. (2015). The formation of academic self-concept in elementary education: A

unifying model for external and internal comparisons. *Contemporary Educational Psychology*, *41*, 124–132.

Pohlmann, B. (2005). Konsequenzen dimensionaler Vergleiche [Consequences of dimensional comparisons]. Münster: Waxmann.

Pohlmann, B., & Möller, J. (2009). On the benefit of dimensional comparisons. *Journal of Educational Psychology*, *101*, 248–258.

Poloczek, S., Karst, K, Praetorius, A.-K., & Lipowsky, F. (2011). Generalisten oder Spezialisten? Bereichsspezifität und leistungsbezogene Zusammenhänge des schulischen Selbstkonzepts von Schulanfängern [Generalists or specialists? Domain specificity of the academic self-concept of school beginners and relationships with achievement]. *Zeitschrift für Pädagogische Psychologie*, *25*, 173–183.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reeve, C. L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than g. *Intelligence*, *32*, 621–652.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, *43*, 73-97.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*, 340–352.

Roick, T., Gölitz, D., & Hasselhorn, M. (2004). *DEMAT 3+: Deutscher Mathematiktest für dritte Klassen* [DEMAT 3+: German test of mathematics for third grade]. Weinheim, Germany: Beltz.

Rost, D. H. (2013). *Handbuch Intelligenz* [Handbook of intelligence]. Weinheim, Germany: Beltz.

Rost, D. H., & Sparfeldt, J. R. (2002). Facetten des schulischen Selbstkonzepts: Ein Verfahren zur Messung des differentiellen Selbstkonzepts schulischer Leistungen und Fähigkeiten (DISK-Gitter) [Academic self-concept facets: A method to assess differential self-concepts of scholastic achievements and abilities (DISC grid)]. *Diagnostica*, *48*, 130–140.

Rost, D. H., Sparfeldt, J. R., & Schilling, S. R. (2007). *DISK-Gitter mit SKSLF-8. Differentielles Schulisches Selbstkonzept mit Skalen zur Erfassung des Selbstkonzepts schulischer Leistungen und Fähigkeiten (Manual)* [DISC grid with SKSLF-8. Differentiated School Self-Concept grid including academic and ability self-concept scales]. Göttingen, Germany: Hogrefe.

Rost, D. H., Sparfeldt, J. R., Dickhäuser, O., & Schilling, S. R. (2005). Dimensional comparisons in subject-specific academic self-concepts and achievements: A quasi-experimental approach. *Learning and Instruction*, *15*, 557–570.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68–78.

Sächsisches Staatsministerium für Kultus (2017). *Verordnung des Sächsischen Staatsministeriums für Kultus über Grundschulen im Freistaat Sachsen. Schulordnung Grundschulen – SOGS* [Order of the Saxonian state ministery for culture and education about elementary schools in the federal state of Saxony]. Retrieved from https://www.gew-sachsen.de/fileadmin/media/publikationen/sn/Rechtsvorschriften_Schule/A_Schulgesetz_Schulordnungen_und_Vorschriften_fuer_alle_Schularten/A_2_a__Schulordnung_Grundschulen__Stand_15-02-17_.pdf

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233–247). London, UK: Kluwer Academic Publishers.

Sauer, J., & Gamsjäger, E. (1996). *Ist Schulerfolg vorhersagbar? Die Determinanten der Grundschulleistung und ihr prognostischer Wert für den Sekundarschulerfolg* [Is educational success predictable? The determinants of elementary school success and their prognostic validity for the success in secondary school]. Göttingen, Germany: Hogrefe.

Schicke, M. C. & Fagan, T. K. (1994). Contributions of self-concept and intelligence to the prediction of academic achievement among grade 4, 6, and 8 students. *Canadian Journal of School Psychology*, *10*, 62–68.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183–212). Hillsdale, NJ: Erlbaum.

Schilling, S. R., Sparfeldt, J. R., & Rost, D. H. (2004). Wie generell ist das Modell? Analysen zum Geltungsbereich des „Internal/External Frame of Reference"-Modells [How general is the model? Analyzing the applicability of the internal/external frame of reference model to more than two school subjects]. *Zeitschrift für Pädagogische Psychologie*, *18*, 221–230.

Schilling, S. R., Sparfeldt, J. R., & Rost, D. H. (2006). Facetten schulischen Selbstkonzepts: Welchen Unterschied macht das Geschlecht? [Gender differences

in subject-specific academic self-concepts]. *Zeitschrift für Pädagogische Psychologie*, *20*, 9–18.

Schmidt, I., Brunner, M., Keller, L., Scherrer, V., Wollschläger, R., Baudson, T. G., & Preckel, F. (2017). Profile information of academic self-concept in elementary school students in grades 1 to 4. *PLoS ONE*, *12*, e0177854.

Schneider, R., Lotz, C., & Sparfeldt, J. R. (2018). Smart, confident, interested: Contributions of intelligence, self-concept, and interest to elementary school achievement. *Learning and Individual Differences*, *62*, 23–35.

Schneider, R., & Sparfeldt, J. R. (under revision). Twofold multidimensional academic self-concepts in elementary school students: Social and dimensional comparisons. *Contemporary Educational Psychology*.

Schneider, R., & Sparfeldt, J. R. (2016). Zur (Un-)Genauigkeit selbstberichteter Zensuren bei Grundschulkindern [The accurarcy of self-reported grades in elementary school]. *Psychologie in Erziehung und Unterricht*, *63*, 48–59.

Schneider, W., & Pressley, M. (1989). *Memory development between 2 and 20*. Heidelberg, Germany: Springer.

Schneider, W., Stefanek, J., & Dotzel, H. (1997). Erwerb des Lesens und des Rechtschreibens. Ergebnisse aus dem SCHOLASTIK-Projekt [Learning to read and to write: Results of the SCHOLASTIK project]. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* [Development in elementary school age] (pp. 113–129). Weinheim, Germany: Beltz.

Schrader, F.-W., & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer [Day-to-day performance evaluation by teachers]. In F. E. Weinert (Ed.)., *Leistungsmessungen in Schulen* [Performance measurement in schools] (pp. 45 – 58). Weinheim, Germany: Beltz.

Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, *59*, 84–95.

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*, 138–166.

Schwarz, J. A., & Beaver, K. M. (2014). Making (up) the grade? Estimating the genetic and environmental influences of discrepancies between self-reported grades and official GPA scores. *Journal of Youth and Adolescence*, *44*, 1125–1138.

Schwarzer, R., Jerusalem, M. (1982). Selbstwertdienliche Attributionen nach Leistungsrückmeldungen. [Self-serving attributions after performance feedback]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 14*, 47–57.

Seitz, W., & Rausche, A. (2004). *Persönlichkeitsfragebogen für Kinder zwischen 9 und 14 Jahren* [Personality inventory for children between 9 and 14 years]. Göttingen: Hogrefe.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441.

Skaalvik, E. M., & Havget, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58*, 292–307.

Skaalvik, E. M., & Skaalvik, S. (2002). Internal and external frame of reference for academic self-concept. *Educational Psychologist, 37*, 233–244.

Skaalvik, E. M., & Valås, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *The Journal of Experimental Education*, 67, 135–149.

Skaalvik, E. M., & Valås, H. (2001). Achievement and self-concept in mathematics and verbal arts: A study of relations. In R. J. Riding & S. G. Rayner (Eds.), *International perspectives on individual differences, Vol. 2: Self perception* (pp. 221–238). Westport, CT: Ablex.

Sparfeldt, J. R., Buch, S. R., Rost, D. H. & Lehmann, G. (2008). Akkuratesse selbstberichteter Zensuren [The accuracy of self-reported grades in school]. *Psychologie in Erziehung und Unterricht, 55*, 68–75.

Sparfeldt, J. R., Lotz, C., Sapp, S., & Rost, D. H. (2016). Psychometrische Prüfung des „Bildertest zum sozialen Selbstkonzept“ (BSSK) [„Solitary with others“ – An evaluation oft he BSSK (Pictorial scale of social self-concept)]. *Diagnostica, 62*, 255–265.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*, 363–374.

Statistisches Bundesamt (2017). *Allgemeinbildende und berufliche Schulen* [general and vocational schooling]. Retrieved from https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Schulen/Tabellen/AllgemeinBildendeBeruflicheSchulenSchulartenSchueler.html;jsessionid=00C7F032B96D85DA70AEDE2EE2D7E640.InternetLive2

Steinmayr, R., & Meißner, A. (2013). Zur Bedeutung der Intelligenz und des Fähigkeitsselbstkonzeptes bei der Vorhersage von Leistungstests und Noten in Mathematik [The importance of intelligence and ability self-concept for the

prediction of standardized achievement tests and grades in mathematics]. *Zeitschrift für Pädagogische Psychologie, 27,* 273–282.

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. In L. H. Meyer (Ed.), *Oxford bibliographies online: Education*. New York, NY: Oxford University Press.

Steinmayr, R., Sauer, J., & Gamsjäger, E. (2018). Prognose von Schulerfolg [Predicting educational success]. In D. H. Rost, J. R. Sparfeldt, & S. R. Buch (Eds.), *Handwörterbuch Pädagogische Psychologie* [Concise dictionary of educational psychology] (5th ed., pp. 653-665). Weinheim, Germany: Beltz.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*, 80–90.

Stern, E. (2003). Lernen ist der mächtigste Mechanismus der kognitiven Entwicklung. Der Erwerb mathematischer Kompetenzen [Learning as the most powerful meachanism of cognitive development. The acquisition of mathematical competencies]. In W. Schneider, & M. Knopf (Eds.), *Entwicklung, Lehren und Lernen* [Development, teaching, and learning] (pp. 207–217). Göttingen, Germany: Hogrefe.

Sticca, F., Goetz, T., Bieg, M., Hall, N. C., Eberle, F., & Haag, L. (2017). Examining the accuracy of students' self-reported academic grades from a correlational and a discrepancy perspective: Evidence from a longitudinal study. *PloS ONE, 12*, e0187367.

Sticca, F., Goetz, T., Nett, U. E., Hubbard, K., & Haag, L. (2017). Short- and long-term effects of over-reporting of grades on academic self-concept and achievement. *Journal of Educational Psychology, 109*, 842–854.

Stock, C., & Schneider, W. (2008a). *DERET 1–2+: Deutscher Rechtschreibtest für das erste und zweite Schuljahr* [DERET 1–2+: German spelling test for first and second graders]. Göttingen, Germany: Hogrefe.

Stock, C., & Schneider, W. (2008b). *DERET 3–4+: Deutscher Rechtschreibtest für das dritte und vierte Schuljahr* [DERET 3–4+: German spelling test for third and fourth graders]. Göttingen, Germany: Hogrefe.

Storch, S. A., & Whitehurst, G. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947.

Swann,W. B., Jr., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*, 84–94.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. *Psychological Research*, *78*, 852–868.

Tobbell, J., & O'Donnell, V. L. (2013). The formation of interpersonal and learning relationships in the transition from primary to secondary school: Students, teachers and school context. *International Journal of Educational Research, 59,* 11–23.

Trautwein, U., & Lüdtke, O. (2005). The big-fish-little-pond effect: Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie, 19,* 137–140.

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104,* 763–777.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefes and academic achievement: A meta-analytic review. *Educational Psychologist, 39,* 111–133.

Valtin, R., Wagner, C. & Schwippert, K. (2005). Schülerinnen und Schüler am Ende der vierten Klasse – schulische Leistungen, lernbezogene Einstellungen und außerschulische Lernbedingungen [Students at the end of grade four – scholastic achievements, learning-related attitudes, and non-scholastic learning conditions]. In W. Bos, Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Eds.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* [IGLU. In-depth analyses of reading comprehension, framework conditions, and additional studies] (pp. 187–238). Münster, Germany: Waxmann.

van Aken, M. A. G., Helmke, A., & Schneider, W. (1997). Selbstkonzept und Leistung – Dynamik eines Zusammenspiels: Ergebnisse aus dem SCHOLASTIK-Projekt [Self-concept and achievment – Dynamics of an interplay]. In F. E. Weinert, & A. Helmke (Eds.). *Entwicklung im Grundschulalter* [Development in elementary school age] (pp. 341–350). Weinheim, Germany: Beltz.

von Maurice, J., Dörfler, T., & Artelt, C. (2014). The relation between interests and grades: Path analyses in primary school age. *International Journal of Educational Research, 64,* 1–11.

Weber, H. S., Lu, L. Shi, J., & Spinath, F. M. (2013). The roles of cognitive and motivational predictors in explaining school achievement in elementary school. *Learning and Individual Differences, 25,* 85–92.

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2015). Zur Bedeutung von Grundschulnoten für die Veränderung von Intrinsischer Motivation und Fähigkeitsselbstkonzept in Deutsch [The role of grades for changes in intrinsic motivation and ability self-concept in German]. *Zeitschrift für Pädagogische Psychologie, 29,* 193–204.

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *British Journal of Educational Psychology*, *87*, 187–204.

Weinert, F. E., & Helmke, A. (1995). Inter-classroom differences in instructional quality and interindividual differences in cognitive development. *Educational Psychologist*, *30*, 15–20.

Weiß, R. H., & Osterland, J. (2013). *CFT 1-R. Grundintelligenzskala 1 – Revision* [CFT1-R. Intelligence scale]. Göttingen, Germany: Hogrefe.

Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, *64*, 956–972.

Wetzel, E., & Greiff, S. (2018). The world beyong rating scales: Why we should think more carefully about the response format in questionnaires. *European Journal of Psychological Assessment*, *34*, 1–5.

Wheeler, L., & Miyake, K. (1992). Social comparison in everyday life. *Journal of personality and social psychology*, *62*, 760–773.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324).Washington, DC: American Psychological Association.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81.

Wigfield, A., Eccles, J. S. & Pintrich, P. (1996). Development between the ages of 11 and 25. In D. C. Berliner & R. C. Calfee (Hrsg.), *Handbook of educational psychology* (pp. 148–185). New York, NY: Macmillan.

Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement solutions. *Educational Psychologist*, *26*, 233–261.

Wild, K.-P. (1991). *Identifikation hochbegabter Schüler: Lehrer und Schüler als Datenquelle* [Identification of giftet pupils. Teachers and pupils as data resources]. Heidelberg, Germany: Asanger.

Willard, G., & Gramzow, R. H. (2008). Exaggeration in memory: Systematic distortion of self-evaluative information under reduced accessibility. *Journal of Experimental Social Psychology*, *44*, 246–259.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*, 1–37.

Wittmann, W. W. (1988). Multivariante reliability theory. Principles of symmetry and successful validation strategies. In R. B. Catell & J. R. Nesselroade (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). New York, NY: Plenum Press.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational Psychological Measurement, 76*, 913–934.

Wolff, F., Nagy, N., Helm, F., & Möller, J. (2018). Testing the internal/external frame of reference model of academic achievement and academic self-concept with open self-concept reports. *Learning and Instruction*, *55*, 58–66.

Zigler, E., & Styfco, S. J. (2004). *The Head Start Debates*. Baltimore, MD: Brookes Publishing Company.

6 Appendix

Appendix A

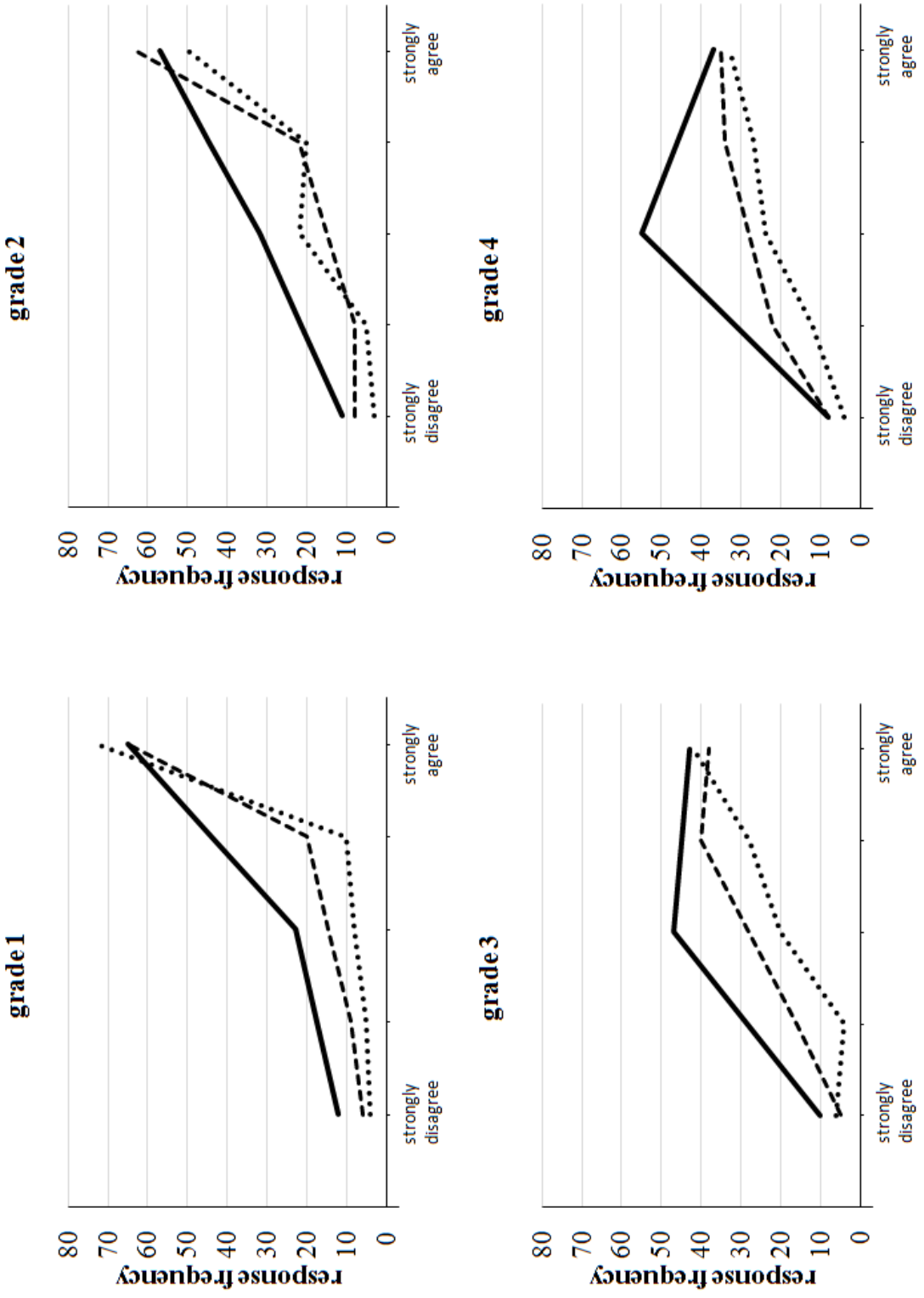
Latent correlations of reported grades in mathematics and German, competence and affect self-concepts in mathematics and reading as well as intelligence – separately for elementary school grades 1 to 4

grade		1	2	3	4	5	6
1	1 RG math						
	2 RG German	.74*					
	3 SC math competence	.42*	.29*				
	4 SC math affect	.31*	.26*	.94*			
	5 SC reading competence	.22*	.40*	.44*	.39*		
	6 SC reading affect	.18*	.28*	.36*	.45*	.86*	
	7 intelligence	.21*	.60*	.38*	.25*	.33*	.29*
2	2 RG German	.55*					
	3 SC math competence	.52*	.22*				
	4 SC math affect	.29*	.09	.87*			
	5 SC reading competence	.17	.31*	.38*	.31*		
	6 SC reading affect	.08	.10	.20	.31*	.79*	
	7 intelligence	.54*	.45*	.19	.09	.03	-.01
3	2 RG German	.65*					
	3 SC math competence	.56*	.32*				
	4 SC math affect	.29*	.06	.79*			
	5 SC reading competence	.25*	.41*	.35*	.15*		
	6 SC reading affect	.05	.22*	.23*	.30*	.73*	
	7 intelligence	.57*	.50*	.28*	.17	.20*	.18*
4	2 RG German	.64*					
	3 SC math competence	.62*	.29*				
	4 SC math affect	.34*	.09	.78*			
	5 SC reading competence	.27*	.37*	.26*	.18*		
	6 SC reading affect	.17*	.29*	.16*	.17*	.90*	
	7 intelligence	.61*	.36*	.50*	.33*	.10	.06

Notes. RG = reported grade, SC = self-concept. *$p < .05$.

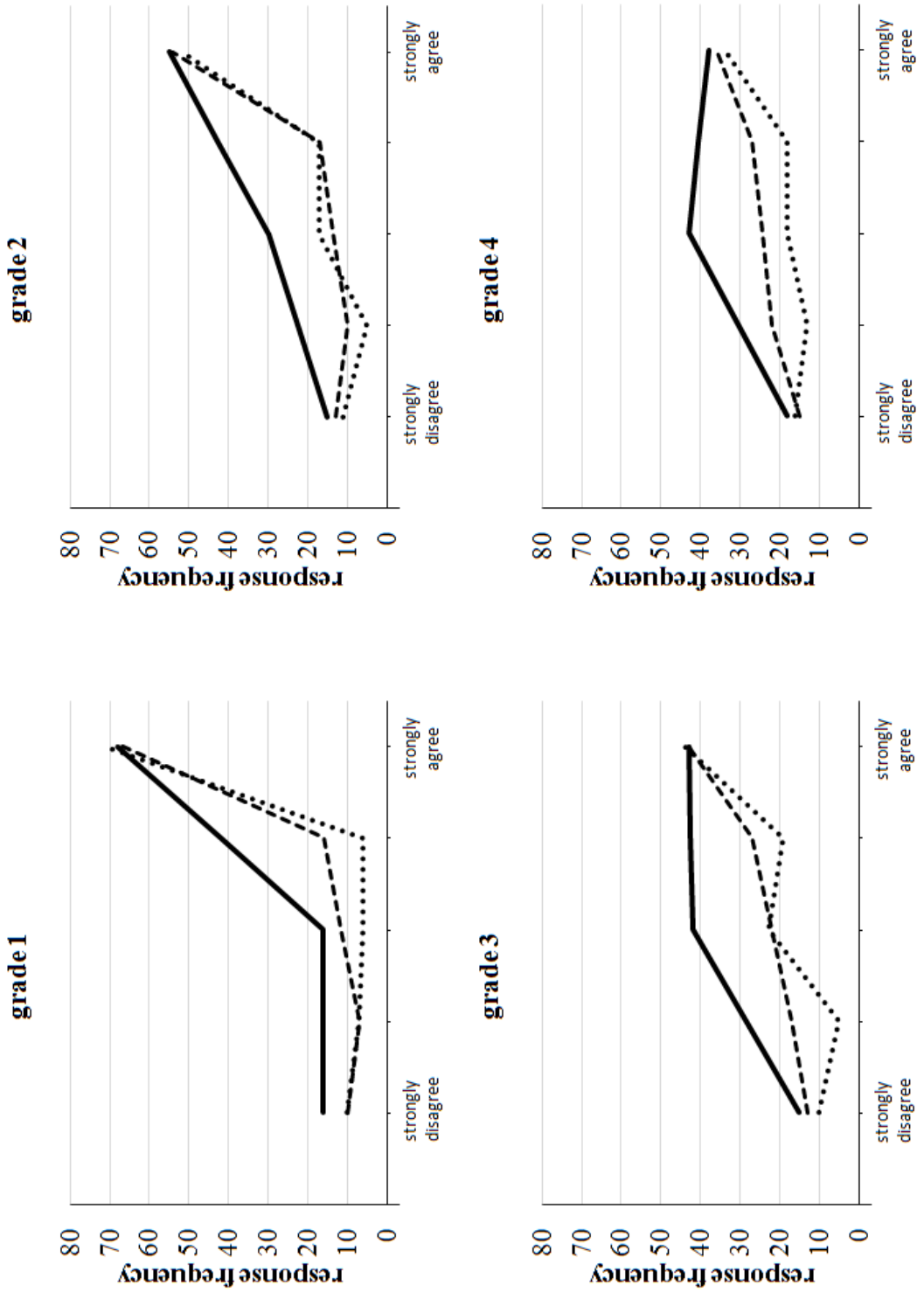
Appendix B

Relative frequencies (%) of response categories for the competence self-concept in mathematics – separately for each response format. Solid line = three-point scale, dashed line = four-point scale, dotted line = five-point scale.



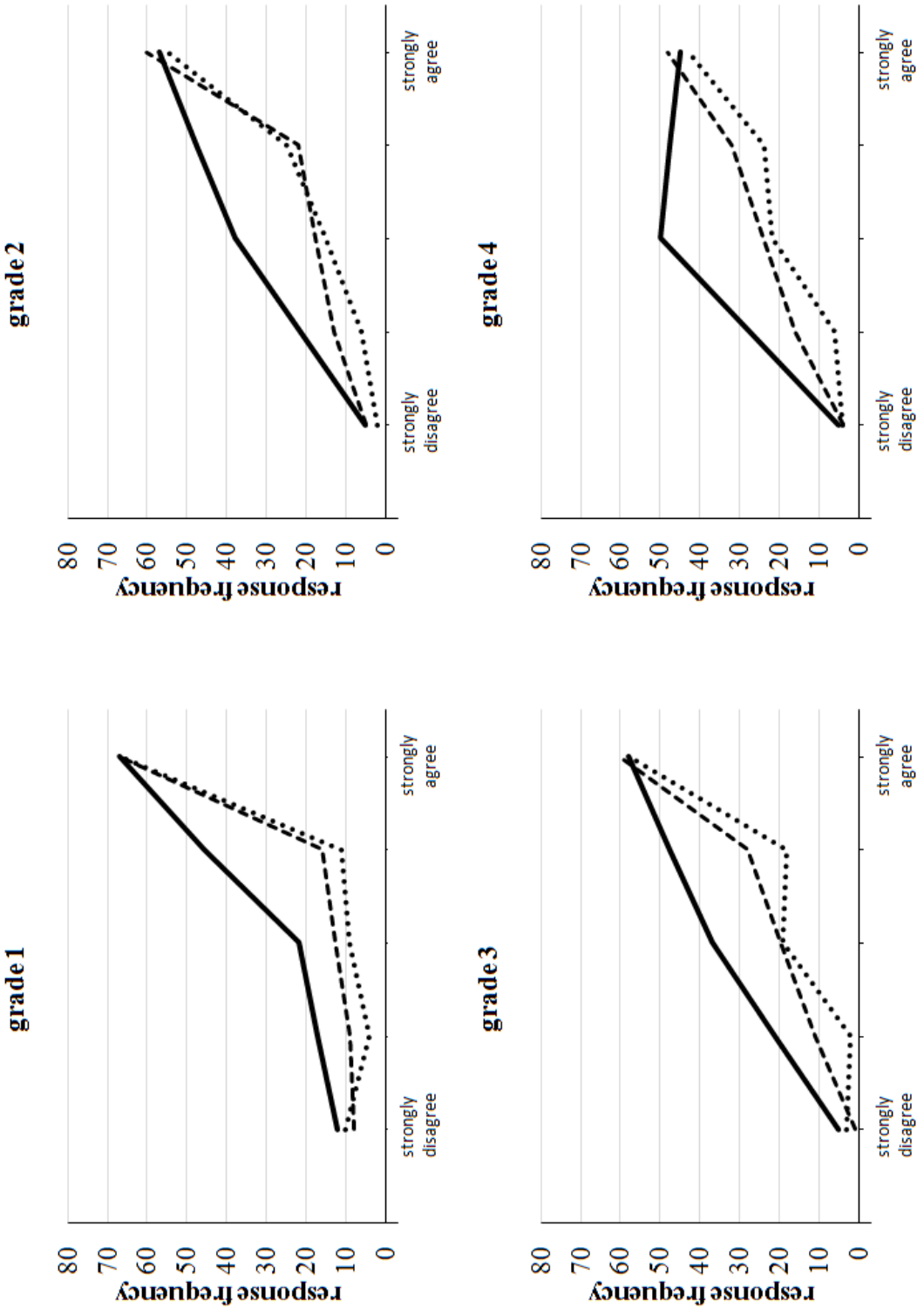
Appendix C

Relative frequencies (%) of response categories for the affect self-concept in mathematics – separately for each response format. Solid line = three-point scale, dashed line = four-point scale, dotted line = five-point scale.



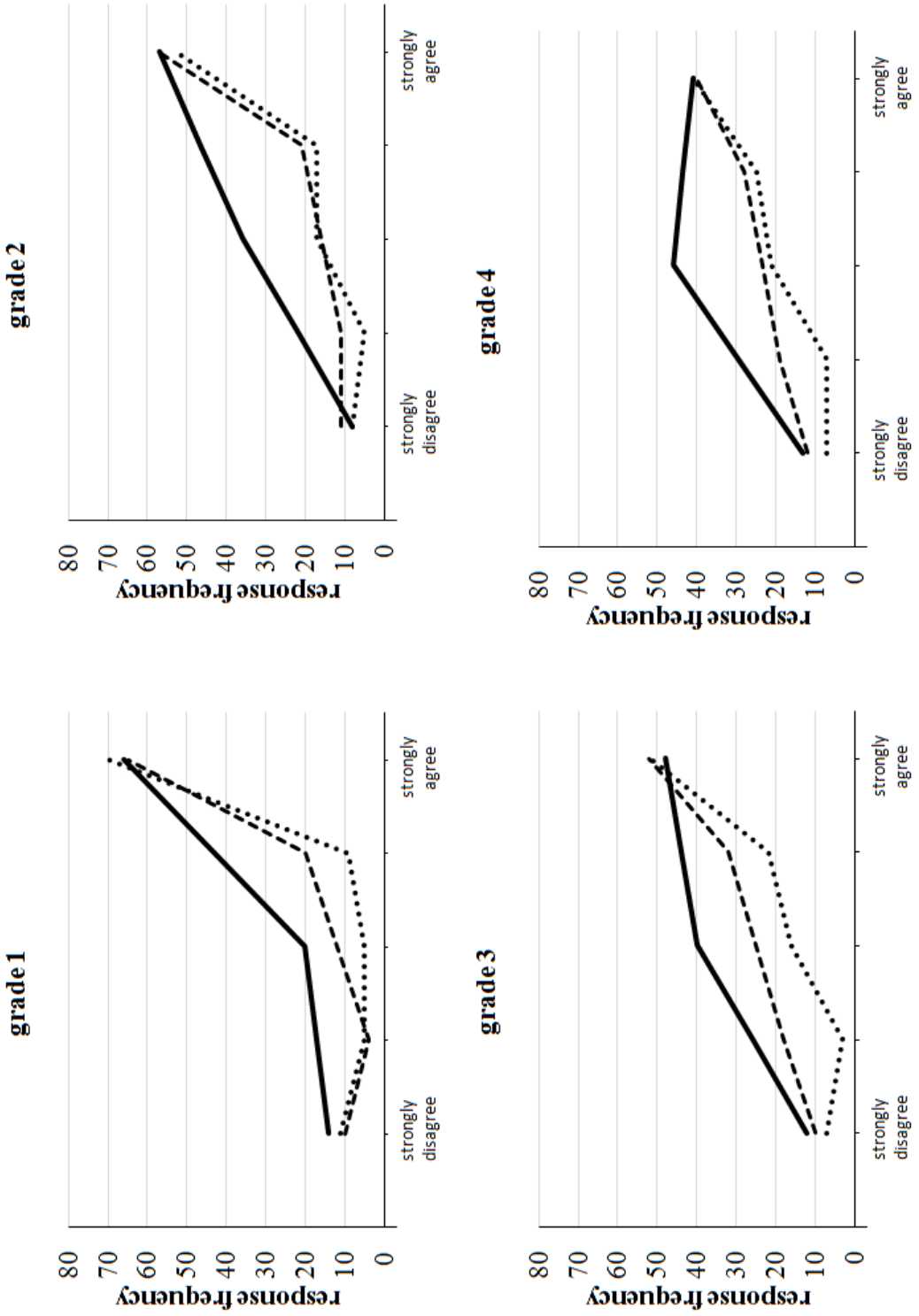
Appendix D

Relative frequencies (%) of response categories for the competence self-concept in reading – separately for each response format. Solid line = three-point scale, dashed line = four-point scale, dotted line = five-point scale.



Appendix E

Relative frequencies (%) of response categories for the affect self-concept in reading – separately for each response format. Solid line = three-point scale, dashed line = four-point scale, dotted line = five-point scale.



7 Publications

Study 1

Schneider, R., & Sparfeldt, J. R. (2016). Zur (Un-)Genauigkeit selbstberichteter Zensuren bei Grundschulkindern [The accurarcy of self-reported grades in elementary school]. *Psychologie in Erziehung und Unterricht*, *63*, 48–59.

Study 2

Schneider. R., & Sparfeldt, J. R. (under revision). Twofold multidimensional academic self-concepts in elementary school students: Social and dimensional comparisons. *Contemporary Educational Psychology*.

Study 3

Schneider, R., Lotz, C., & Sparfeldt, J. R. (2018). Smart, confident, interested: Contributions of intelligence, self-concept, and interest to elementary school achievement. *Learning and Individual Differences*, *62*, 23–35.

Lebenslauf

Name	Rebecca Schneider
E-Mail	rebecca.schneider@posteo.de
Geburtsdatum und -ort	10.07.1987 in Wolmirstedt

Beruflicher Werdegang

10/2018 – heute	**Institut zur Qualitätsentwicklung im Bildungswesen** Wissenschaftliche Mitarbeiterin
03/2014 – 09/2018	**Universität des Saarlandes** Bildungswissenschaften Lehrstuhl für Diagnostik, Beratung & Intervention Wissenschaftliche Mitarbeiterin davon 02/2016 – 02/2017 Beratungsstelle „Forschende Lehre mit komplexen Datensätzen“ im „Bund-Länder-Programm für bessere Studienbedingungen und mehr Qualität in der Lehre (‘Qualitätspakt Lehre’)“

Schulische Ausbildung und Studium

04/2014 – 12/2018	**Universität des Saarlandes** Bildungswissenschaften Lehrstuhl für Diagnostik, Beratung & Intervention Doktorandin der Psychologie
10/2011 – 01/2014	**Otto-Friedrich-Universität Bamberg** Psychologie, Master of Science
10/2008 – 09/2011	**Otto-von-Guericke-Universität Magdeburg** Psychologie, Bachelor of Science
08/2004 – 06/2007	**Diesterweg-Gymnasium Tangermünde** Allgemeine Hochschulreife