Technische Universität Dortmund

Fakultät Erziehungswissenschaft, Psychologie und Bildungsforschung

Selbsteinschätzungsbiaseffekte und Selbsteinschätzungseffekte auf akademische Leistung operationalisiert über Schulnoten

Dissertation zur Erlangung des akademischen Grades Doktor der Philosophie

(Dr. phil.)

vorgelegt von Patrick Paschke

geboren am 28.01.1991 in Menden

Matrikelnummer 212788

Vorgeschlagene Erstgutachterin: Prof. Dr. Ricarda Steinmayr

Vorgeschlagene Zweitgutachterin: Prof. Dr. Fani Lauermann

Dortmund, Juli 2023

Danksagung

Mein vorrangiger Dank gilt Ricarda Steinmayr, die meine Promotion von der frühesten Planung bis zur Abgabe begleitet hat. Neben der fachlichen Betreuung hat ihre offene Art und stetige Bereitschaft sich mit Vorschlägen und auch neuen Ideen auseinanderzusetzen viel dazu beigetragen, dass meine Promotionszeit nicht nur inhaltlich interessant war, sondern auch Spaß gemacht hat. Die Unterstützung beim Einsatz in diesem Forschungsfeld auch unkoventioneller Verfahren war für mich ein wichtiger motivierender Faktor.

Ebenso danke ich Anne Weidinger, die mich bis zu ihrem Weggang von der TU Mitte 2021 bei der Dissertation unterstützt hat und auch darüber hinaus als Co-Autorin meines zweiten empirischen Artikels tätig war. Sie hatte stets ein offenes Ohr für Fragen jeglicher Art und hat mich mit ihrem Engagement inspiriert, meine eigene Arbeit ebenfalls stetig zu verbessern.

Zudem danke ich meinen Kolleginnen und Kollegen für die entspannte und freundliche Arbeitsatmosphäre in den letzten Jahren. Speziell nennen möchte ich hier Josi Michels und Lisa Tometten, mit denen ich mich regelmäßig über unsere jeweiligen Promotionsvorhaben austauschen konnte, sowie Sebastian Bergold, der als mein erster Büronachbar geholfen hat, mich in die Arbeitsabläufe an der TU einzuarbeiten. Stellvertretend für die ansonsten viel zu lange Liste netter Kolleginnen und Kollegen seien hier zudem Linda Wirthwein, Helene Eckert und Anke Heyder genannt, mit denen ich an unterschiedlichen Stellen und in verschiedenen Projekten zusammenarbeiten konnte.

Zuletzt möchte ich meiner Familie und meinen Freunden danken, die den notwendigen Ausgleich zur Arbeit an der Promotion geschaffen haben. Ein herzliches Dankeschön geht an meine Eltern Heike und Andreas Paschke, meine Schwester Tanja Paschke sowie Michael Kockmeyer, Patrick Behr, Kerst Mattner und Berfin Yekta.

Inhaltsverzeichnis

Zusammenfassung

Die Selbsteinschätzung eigener Kompetenzen ist von Bedeutung für die akademische Leistung von Schüler*innen, beispielsweise für bessere Noten, erreichte Abschlüsse und Ergebnisse in standardisierten Schulleistungstests (z.B. Marsh et al., 2022; Talsma et al., 2018; Wu et al., 2021). Weniger ist allerdings darüber bekannt, ob die Diskrepanz zwischen selbsteingeschätzten und tatsächlichen Kompetenzen (self-evaluation Bias; SE Bias; Bonneville-Roussy et al., 2017) ebenfalls einen Einfluss auf die akademische Leistung von Schüler*innen hat (Trautwein & Möller, 2016). Der SE Bias ist konzeptionalisiert als die Diskrepanz, nicht die absolute Diskrepanz, zwischen Selbsteinschätzung und tatsächlicher Kompetenz. Mit anderen Worten, der SE Bias kann positive wie negative Werte annehmen, wobei positive Werte einer Selbstüberschätzung und negative Werte einer Selbstunterschätzung entsprechen (Humberg et al., 2018). In der Literatur werden unterschiedliche Hypothesen über SE Bias Effekte auf akademische Leistung vertreten: Am häufigsten ist die Hypothese, dass positive SE Bias Effekte auf akademische Leistung bestehen (z.B. Bonneville-Roussy et al., 2017; Leduc & Bouffard, 2017; Lee, 2021). Demnach wäre ein größerer SE Bias förderlich für akademische Leistung, was impliziert, dass Selbstüberschätzung förderlich, Selbstunterschätzung hingegen hinderlich ist. Andere Studien legen allerdings auch andere Effekte nahe. Beispielsweise zeigen Studien zum selbstregulierten Lernen, dass akkurate Selbsteinschätzungen nützlich sind, um Lernverhalten effektiv zu motivieren und steuern (z.B. Dunslosky & Rawson, 2012; Hacker & Bol, 2019; van Loon & Oeri, 2023), sodass ein SE Bias von null (also eine akkurate Selbsteinschätzung) optimal sein sollte. Demnach bestünde also ein umgekehrt u-förmiger Effekt des SE Bias auf akademische Leistung und Selbstüberschätzung wie Selbstunterschätzung wären beide hinderlich. Zudem legen methodisch-theoretische Arbeiten (z.B. Edwards & Parry, 1993; Humberg et al., 2018; 2019a) nahe, dass die weitaus meisten Studien zu SE Bias Effekten von erheblichen theoretischen wie methodischen Mängeln betroffen sind. Der Hauptkritikpunkt besteht in der mangelnden Differenzierung zwischen Effekten des SE Bias und Effekten der Selbsteinschätzung an sich. Dadurch konnte in bisherigen Studien nicht auseinander gehalten werden, ob tatsächlich Effekte eines SE Bias oder schlicht Effekte einer hohen Selbsteinschätzung per se vorliegen. In der vorliegenden Arbeit sollen SE Bias Effekte auf akademische Leistung theoretisch wie empirisch von Effekten der Selbsteinschätzung abgegrenzt und in Hinblick auf ihre Wirkmechanismen untersucht werden. Die Arbeit beinhaltet drei empirische Beiträge – zwei publizierte Studien, sowie eine weiterführende Analyse, welche sich in Vorbereitung zur Publikation befindet und im Diskussionsteil der Dissertation präsentiert wird.

In *Beitrag I* wurden lineare SE Bias Effekte auf Schulnoten in Mathematik und Deutsch untersucht. Die Stichprobe bestand aus 284 Gymnasiast*innen zu Beginn der zehnten Klassenstufe. Wir erfassten das Fähigkeitsselbstkonzept in Mathematik und Deutsch, sowie die jeweiligen Kompetenzen mit standardisierten Kompetenztests. Als abhängige Variable zur Operationalisierung der akademischen Leistung wurden die Zeugnisnoten in Mathematik und Deutsch am Ende der zehnten Klassenstufe erfasst. Zusätzliche Kontrollvariablen waren die Ausgangswerte der Noten, der sozioökonomische Status (SÖS) und das Geschlecht. Eine neue Methode, die condition-based regression analysis (CRA; Humberg et al., 2018), wurde verwendet, um die Effekte von SE Bias und Selbsteinschätzung empirisch zu trennen und auf Signifikanz zu prüfen. Die Ergebnisse wiesen darauf hin, dass keine SE Bias Effekte auf die Noten bestanden, sondern lediglich positive Haupteffekte des Fähigkeitsselbstkonzepts (in beiden Domänen) und der Kompetenz (in Mathematik).

Beitrag II stellt eine Erweiterung von Beitrag I dar, da neben linearen SE Bias Effekten auch nonlineare Effekte untersucht wurden. Wie bereits erwähnt, legen verschiedene Studien die Existenz nonlinearer (z.B. umgekehrt u-förmiger) SE Bias Effekte auf akademische Leistung nahe. Auch andere nonlineare Effekte, beispielsweise ein positiver kurvilinearer Effekt, welcher mit zunehmendem SE Bias geringer und ab einem gewissen

Wert negativ wird, wurde diskutiert (optimal margin hypothesis; siehe z.B. Baumeister et al., 1989; Helmke, 1998; Praetorius et al., 2016). Untersucht wurden 504 Gymnast*innen zu Beginn der zehnten Klassenstufe. Erfasst wurden analog zu Beitrag I das Fähigkeitsselbstkonzept in Mathematik, die Mathematikkompetenz, die Mathematikzeugnisnoten am Ende der zehnten sowie am Ende der neunten Klassenstufe und weitere Kontrollvariablen. Die Daten wurden mit der response surface analysis (RSA; Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2019a) untersucht, welche die Modellierung nonlinearer SE Bias Effekte und ihre Abgrenzung von Selbsteinschätzungseffekten erlaubt. Die Modelle, welche unterschiedliche Hypothesen über SE Bias Effekte repräsentieren (positiver linearer Effekt, negativer linearer Effekt, umgekehrt u-förmiger Effekt, optimal margin Effekt, kein Effekt), wurden durch informationstheoretische Modellvergleiche anhand von Akaike-Gewichten vergleichend evaluiert (Burnham & Anderson, 2002; Humberg et al., 2019a; Wagenmakers & Farrell, 2004). Die Ergebnisse wiesen auf eine deutliche Überlegenheit des sogenannten beneficial self-evaluation and competence models hin, welches lediglich positive lineare Effekte des Fähigkeitsselbstkonzepts und der Kompetenz postuliert, aber keine nonlinearen Effekte oder SE Bias Effekte. Somit bestätigen und erweitern die Befunde aus Beitrag II jene aus Beitrag I, da in Beitrag II auch nonlineare SE Bias Effekte zurückgewiesen werden konnten.

In der *weiterführenden Analyse (Beitrag III)* wurde untersucht, inwieweit die Effekte des Fähigkeitsselbstkonzepts, der Kompetenz und potentielle SE Bias Effekte auf akademische Leistung von Erfolgserwartungen und subjektiven Werten mediiert werden. Eine entsprechende Mediation von Effekten des Fähigkeitsselbstkonzepts wird im Erwartungs-Wert-Modell (Eccles & Wigfield; 2020; 2023; Wigfield et al., 2020) angenommen. Zudem gehen verschiedene Autor*innen davon aus, dass motivationale Variablen SE Bias Effekte auf akademische Leistung mediieren (z.B. Bonneville-Roussy et al., 2017; Lee, 2021; Taylor & Brown, 1988). Wir untersuchten dieselbe Stichprobe wie in Beitrag II unter Hinzunahme zusätzlicher Messzeitpunkte und Variablen. Als Prädiktoren wurden weiterhin das Fähigkeitsselbstkonzept sowie die Kompetenz in Mathematik zu Beginn der zehnten Klassenstufe genutzt. Als Mediatoren dienten die Erfolgserwartungen und subjektiven Werte in Mathematik zu Beginn der elften Klasse und als abhängige Variable die Mathematikzeugnisnote am Ende der elften Klasse. Der Ausgangswert der Noten und weitere Kontrollvariablen wurden berücksichtigt. Die Ergebnisse wiesen darauf hin, dass lediglich die Kompetenz, nicht aber das Fähigkeitsselbstkonzept oder der SE Bias die späteren Noten vorhersagte. Der Effekt der Kompetenz wurde partiell von den Erfolgserwartungen, nicht aber von den subjektiven Werten mediiert.

Insgesamt zeigen die Beiträge eindeutig, dass keine Hinweise auf SE Bias Effekte auf akademische Leistung bestehen. Für die akademische Leistung von Schüler*innen operationalisiert über Schulnoten scheint somit lediglich die absolute Ausprägung von Fähigkeitsselbstkonzept und Kompetenz, nicht aber ihre Diskrepanz, relevant zu sein.

Abstract

Self-evaluations of own competences are important for students' academic achievement, e.g., for better grades, attained degrees, and results of standardized school achievement tests (e.g., Marsh et al., 2022; Talsma et al., 2018; Wu et al., 2021). However, less is known about potential effects of the discrepancy between self-evaluated and actual competences (self-evaluation bias; SE Bias; Bonneville-Roussy et al., 2017) on academic achievement (Trautwein & Möller, 2016). The SE bias is conceptualized as the discrepancy, not the absolute discrepancy, between self-evaluated and actual competences. In other words, the SE bias can have positive and negative values. Positive values represent a self-overestimation and negative values represent a self-underestimation. Different hypotheses about SE bias effects on academic achievement have been proposed. Most common is the hypothesis of positive SE bias effects on academic achievement (e.g., Bonneville-Roussy et al., 2017; Leduc & Bouffard, 2017; Lee, 2021). Thus, a positive discrepancy between self-evaluated and actual competence (self-overestimation) would be beneficial for academic achievement, while a negative discrepancy (self-underestimation) would be detrimental. However, other studies suggest different types of effects. For example, studies on self-regulated learning show that students use accurate self-evaluations to motivate and guide learning behavior (e.g., Dunslosky & Rawson, 2012; Hacker & Bol, 2019; van Loon & Oeri, 2023). Therefore, an SE bias of zero (accurate self-estimation) should be most beneficial. In other words, there would be a reverse u-shaped effect of SE bias on academic achievement. Moreover, theoretical and methodological studies (e.g., Edwards & Parry, 1993; Humberg et al., 2018; 2019a) suggest that the vast majority of studies on SE bias effects suffer from major theoretical and methodological problems. The main point of criticism is the missing distinction between effects of the bias and effects of the self-evaluation per se. Therefore, prior studies could not distinguish whether there are actual SE bias effects or only positive main effects of the self-evaluation. The goal of the present work is to theoretically and empirically disentangle SE bias effects on academic achievement from self-evaluation effects on academic achievement.

In *Study I*, we analyzed linear SE bias effects on report card grades in mathematics and German. The sample consisted of 284 students at the beginning of Grade 10 in academic track schools ("Gymnasium"). We assessed students' ability self-concepts in mathematics and German as well as the respective competences with standardized competence tests. The dependent variables were the students' report card grades in mathematics and German at the end of Grade 10 as a measure of academic achievement. Additionally, we controled for prior grades, socioeconomic status (SES), and gender. We used a novel method, the condition-based regression analysis (CRA; Humberg et al., 2018), to empirically separate SE bias effects from self-evaluation effects and test them for significance. The results indicated no SE bias effects, but only positive main effects of ability self-concept (both domains) and competence (mathematics).

Study II is an extension of Study I since we analyzed not only linear, but also nonlinear SE bias effects. As pointed out, some studies suggest that there might be nonlinear (e.g., reverse u-shaped) SE bias effects. Other nonlinear effects have been discussed as well, for example, a positive curvilinear effect which becomes less positive and at some point negative with increasing SE bias values (optimal margin hypothesis; e.g., Baumeister et al., 1989; Helmke, 1998; Praetorius et al., 2016). We analyzed data from 504 academic track students at the beginning of Grade 10. As in Study I, we assessed the ability self-concept and competence in mathematics at the beginning of Grade 10, the report card grades in mathematics at the end of Grade 10 as well as at the end of Grade 9, and additional control variables. Data were analyzed with response surface analysis (RSA; Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2019), which allows modelling nonlinear SE bias effects and separating them from self-evaluation effects. We compared the different models which represented different hypotheses about SE bias effects (beneficial linear effect, detrimental

linear effect, reverse u-shaped effect, optimal margin effect, no effect) with information-theoretic model comparison using Akaike weights as the criterion (Burnham & Anderson, 2002; Humberg et al., 2019a; Wagenmakers & Farrell, 2004). The so-called beneficial self-evaluation and competence model, which posits only positive linear effects of ability self-concept and competence but no nonlinear effects or SE bias effects, was by far the best model according to Akaike weights. Therefore, the results of Study II replicate and expand the results from Study I, since nonlinear SE bias effects could be rejected, as well.

In a *supplementary analysis (Study III)*, we analyzed whether the effects of ability self-concept and competence as well as hypothetical SE bias effects on academic achievement were mediated by expectancy of success and subjective task values. A mediation of the effect of ability self-concept on academic achievement by expectancy of success and subjective task values is posited by expectancy-value theory (Eccles & Wigfield; 2020; 2023; Wigfield et al., 2020). Furthermore, different authors suggest that motivational variables mediate SE bias effects on academic achievement (e.g., Bonneville-Roussy et al., 2017; Lee, 2021; Taylor & Brown, 1988). We analyzed data from the same sample as in Study II, but with additional measurement time points and variables. As in Study II, we assessed ability self-concept and competence in mathematics in the first half of Grade 10 as predictors. As mediators, we assessed students' expectancies of success and subjective task values in mathematics in the first half of Grade 11. As the dependent variable, we assessed students' grades in mathematics on the report card from the end of Grade 11. We controled for prior grades and other control variables. Only competence but not ability self-concept or SE bias predicted subsequent grades. The effect of competence was partially mediated by expectancy of success, but not by subjective task values.

Overall, the two empirical studies and the supplementary analysis show that there are no indications of SE bias effects on academic achievement. Thus, only the absolute values of

ability self-concept and competence, but not their discrepancy, appear to influence students' academic achievement.

Einleitung

Es erscheint intuitiv plausibel, dass hohe Selbsteinschätzungen eigener Fähigkeiten und Kompetenzen für die akademischen Leistungen von Schüler*innen von Vorteil sein können. Schließlich ist anzunehmen, dass Schüler*innen, die davon ausgehen, nicht über die zum Erreichen guter akademischer Leistungen notwendigen Kompetenzen zu verfügen, weniger Zeit und Anstrengung in das Erbringen dieser Leistungen investieren. Diese Annahme ist auch im Erwartungs-Wert-Modell (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020) repräsentiert. Demnach wirkt sich eine hohe subjektive Einschätzung eigener Fähigkeiten und Kompetenzen (Fähigkeitsselbstkonzept), vermittelt unter anderem über höhere Erfolgserwartungen, positiv auf akademische Leistungen aus (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020). Sicher werden nicht alle Schüler*innen mit geringen Fähigkeitsselbstkonzepten auch geringe Leistungen erbringen, ebenso wenig wie ein hohes Fähigkeitsselbstkonzept in allen Fällen zu hohen Leistungen führt. Im Mittel sollte ein hohes Fähigkeitsselbstkonzept sich aber förderlich auf akademische Leistungen auswirken. Diese Annahme wurde inzwischen in zahlreichen empirischen Studien bestätigt (z.B. Marsh, 2022; Marsh et al., 2022; Wu et al., 2021). Somit kann der Befund, dass sich hohe Selbsteinschätzungen eigener Fähigkeiten und Kompetenzen im Mittel positiv auf akademische Leistungen auswirken, inzwischen als gesichert gelten. Allerdings ist nach wie vor unklar, ob darüber hinaus auch die Diskrepanz zwischen den selbst eingeschätzten und den tatsächlichen Kompetenzen eine Rolle spielt (Trautwein & Möller, 2016). Mit anderen Worten, es ist wenig darüber bekannt, ob es für akademische Leistungen förderlich ist, seine eigenen Fähigkeiten und Kompetenzen zu überschätzen, sie akkurat einzuschätzen, oder sie gar zu unterschätzen. Auch komplexere Zusammenhänge sind denkbar. Beispielsweise wäre es im Sinne der Theorie einer *optimal margin of illusion* (Baumeister, 1989) möglich, dass eine moderate Selbstüberschätzung für akademische Leistungen förderlich ist, eine extreme Selbstüberschätzung hingegen hinderlich. Natürlich ist es ebenso möglich, dass die

Diskrepanz zwischen selbsteingeschätzten und tatsächlichen Kompetenzen über die absolute Ausprägung der Selbsteinschätzung hinaus keinen bedeutsamen Einfluss auf akademische Leistungen hat. Zwar gibt es eine beachtliche Anzahl empirischer Studien zu dieser Fragestellung (z.B. Bonneville-Roussy et al., 2017; Lee, 2021; Leduc & Bouffard, 2017), allerdings ist die Befundlage insgesamt weniger eindeutig als zu den Effekten der Selbsteinschätzung per se. Ein noch bedeutsameres Problem sind allerdings die theoretischen und methodischen Mängel in der deutlichen Mehrzahl der Studien zu diesem Thema. In nahezu allen entsprechenden Studien wurden Effekte der Diskrepanz zwischen Selbsteinschätzung und tatsächlicher Kompetenz weder theoretisch noch methodisch hinreichend von Effekten der Selbsteinschätzung getrennt (für eine Diskussion siehe Humberg et al., 2018; 2019a). Um zwischen diesen beiden Effekten zu differenzieren, verwende ich den Begriff *Selbsteinschätzungseffekt* für einen Effekt der Höhe der Selbsteinschätzung per se und den Begriff *Selbsteinschätzungsbiaseffekt (SE Bias Effekt)* für einen Effekt der Diskrepanz zwischen Selbsteinschätzung und tatsächlicher Kompetenz (*SE Bias*).

Das primäre Ziel der vorliegenden Dissertation ist es, einen Beitrag zur Beantwortung der Frage zu leisten, ob über Selbsteinschätzungseffekte hinaus SE Bias Effekte auf akademische Leistung bestehen. Als sekundäres Ziel soll zudem überprüft werden, ob etwaigen SE Bias Effekten dieselben im Erwartungs-Wert-Modell beschriebenen Prozesse zugrundeliegenden, wie den Effekten der Selbsteinschätzung.

Die Arbeit ist in drei Abschnitte gegliedert. Im ersten Abschnitt werden die theoretischen Hintergründe und methodischen Besonderheiten der Forschung zu Selbsteinschätzungseffekten und SE Bias Effekten auf akademische Leistungen besprochen. Im ersten Teil dieses Abschnitts (Kapitel 1.1) werden theoretische Grundlagen sowie empirische Befunde aus Studien zu Selbsteinschätzungseffekten beschrieben und zusammengefasst. Im zweiten Teil (Kapitel 1.2) werden theoretische und methodische

Grundlagen zu SE Bias Effekten erläutert. Hierbei liegt der Fokus auf einer eindeutigen Abgrenzung der SE Bias Effekte von Selbsteinschätzungseffekten. Im dritten Teil (Kapitel 1.3) werden die Inhalte der vorausgehenden Kapitel integriert und daraus die Fragestellungen und Hypothesen der vorliegenden Arbeit abgeleitet. Im zweiten Abschnitt werden zwei im Rahmen der vorliegenden Dissertation erstellte und publizierte empirische Arbeiten vorgestellt (Kapitel 2.1 und 2.2). Im dritten Abschnitt wird zunächst eine dritte, bisher noch nicht publizierte, empirische Arbeit als weiterführende Analyse vorgestellt (Kapitel 3.1). Anschließend werden alle empirischen Arbeiten inklusive der weiterführenden Analyse in einer abschließenden Gesamtdiskussion integriert, kritisch diskutiert und in den bisherigen Forschungskontext eingebettet (Kapitel 3.2 bis 3.5).

1 Theoretischer Hintergrund

1.1 Selbsteinschätzungseffekte auf akademische Leistung

1.1.1 Konzept und empirische Befunde

Die Effekte von Selbsteinschätzungen verschiedener Fähigkeiten und Kompetenzen auf die eigene akademische Leistung wurde bereits in einer großen Zahl empirischer Studien untersucht (z.B. Arens & Niepel, 2023; Bakadorova & Raufelder, 2020; Fu et al., 2020; Marsh, 2022; Marsh et al., 2022; Preckel et al., 2017; Sewasew & Koester, 2019; Sewasew & Schroeders, 2019; Weidinger et al., 2018). Zur Verständlichkeit sei an dieser Stelle erwähnt, dass *Selbsteinschätzung* kein einheitliches, wohldefiniertes psychologisches Konstrukt darstellt. Vielmehr handelt es sich um einen Oberbegriff, der verschiedene spezifische Konstrukte, wie das Fähigkeitsselbstkonzept und die Selbstwirksamkeit, umfasst. In der größtenteils englischsprachigen Literatur werden teils uneinheitliche Oberbegriffe genutzt, um verschiedene Konstrukte zusammenzufassen, darunter etwa die Begriffe *self-beliefs* (z.B. Bong et al., 2012, Valentine et al., 2004), *self-perceptions* (z.B. Humberg et al., 2019a, Skaalvik & Skaalvik, 2008) und *self-evaluations* (z.B. Bonneville-Roussy et al., 2017). Die vielleicht am meisten untersuchten Konstrukte in diesem Forschungsbereich sind das Fähigkeitsselbstkonzept und die Selbstwirksamkeit (Bong et al., 2012; Marsh et al., 2019). Bei beiden Konstrukten geht es um die subjektive Einschätzung eigener Fähigkeiten und Kompetenzen in bestimmten Domänen oder Aufgaben. Der zentrale konzeptionelle Unterschied zwischen dem Fähigkeitsselbstkonzept und der Selbstwirksamkeit besteht darin, dass das Fähigkeitsselbstkonzept die „reine“ Einschätzung einer bestimmten, meist relativ umfassenden Fähigkeit oder Kompetenz darstellt (bspw. der mathematischen Kompetenz), wohingegen es bei der Selbstwirksamkeit darum geht, ob eine Person davon ausgeht, die notwendigen Fähigkeiten und Kompetenzen zu besitzen, um ein gewünschtes Ziel herbeizuführen (Bong et al., 2012; Bong & Clark, 1999; Bong & Skaalvik, 2003). Ein typisches Item zur Erfassung des Fähigkeitsselbstkonzepts, in diesem Beispiel in Mathematik, ist „Ich bin gut in Mathe“ (Spinath & Steinmayr, 2012, S. 1148). Ein typisches Item zur Erfassung der Selbstwirksamkeit ist „Ich bin überzeugt, dass ich auch den kompliziertesten Stoff, den der Lehrer in Mathematik vorstellt, verstehen kann“ (OECD, 2020a). Ursprünglich ging man davon aus, dass das akademische Fähigkeitsselbstkonzept hierarchisch organisiert ist und verschiedene domänenspezifische Komponenten, wie Fähigkeitsselbstkonzepte in Mathematik, Englisch oder Geschichte subsumiert (Shavelson et al., 1976). Inzwischen ist aus zahlreichen Studien zu den Zusammenhängen zwischen verschiedenen domänenspezifischen Fähigkeitsselbstkonzepten allerdings bekannt, dass das mathematische und das verbale Fähigkeitsselbstkonzept nicht oder kaum zusammenhängen (z.B. Arens & Niepel, 2023; Esnaola et al., 2020; Lauermann et al., 2020; Lohbeck, 2020; Wolff et al., 2020; Wolff et al., 2021a; 2021b; für Metaanalysen siehe Möller et al., 2009; Möller et al., 2020; Wan et al., 2021; Wolff & Möller, 2022). Somit sind in neueren Modellen das mathematische und das verbale Fähigkeitsselbstkonzept meist als voneinander unabhängige Faktoren zweiter Ordnung konzipiert. Diesen zugeordnet sind verschiedene spezifischere Fähigkeitsselbstkonzepte als Faktoren erster Ordnung, die sich dahingehend unterscheiden, wie stark sie mit den Faktoren zweiter Ordnung zusammenhängen. So hängt beispielsweise das biologische Fähigkeitsselbstkonzept primär mit dem mathematischen aber in geringerem Maße auch mit dem verbalen Fähigkeitsselbstkonzept zusammen, wohingegen es sich beim Fähigkeitsselbstkonzept in Geschichte andersherum verhält (Marsh et al., 1988; Marsh et al., 2015; Trautwein & Möller, 2016). Ähnlich wie beim Fähigkeitsselbstkonzept kann auch bei der Selbstwirksamkeit zwischen verschiedenen Graden der Spezifität unterschieden werden. Das oben angeführte Beispielitem stammt aus einer Skala zur Erfassung der relativ allgemeinen Selbstwirksamkeit im Schulfach Mathematik. Darüber hinaus wird häufig aber auch die aufgabenspezifische Selbstwirksamkeit erfasst, welche die subjektive Überzeugung einer Person beschreibt, eine bestimmte Aufgabe oder eine bestimmte eng umschriebene Art von Aufgaben lösen zu können. Ein Beispielitem zur Erfassung der aufgabenspezifischen

Selbstwirksamkeit ist „Wie sicher glaubst du, folgende Mathematikaufgaben lösen zu können? Ausrechnen, wie viel billiger ein Fernseher bei 30 % Rabatt wäre“ (OECD, 2020b). Obwohl sich das Fähigkeitsselbstkonzept und die Selbstwirksamkeit somit in ihrer Konzeptionalisierung und Operationalisierung unterscheiden, zeigen sich empirisch hohe bis sehr hohe Korrelationen teils um $r = .90$ oder höher zwischen den Konstrukten (Bong et al., 2012; Marsh et al., 2019). Andere Autor*innen fanden geringere Korrelationen und argumentieren für die empirische Unterscheidbarkeit von Fähigkeitsselbstkonzept und Selbstwirksamkeit (Pfeiffer et al., 2020). Auch die prädiktive Validität von Fähigkeitsselbstkonzepten und Selbstwirksamkeit für akademische Leistung scheint ähnlich ausgeprägt zu sein. In einer Metaanalyse fanden Valentine et al. (2004) keine Unterschiede zwischen der Vorhersagekraft der verschiedenen untersuchten Konstrukte (Fähigkeitsselbstkonzept, Selbstwirksamkeit und Selbstwert) auf akademische Leistung unter Kontrolle der vorausgegangenen akademischen Leistung. Ebenso hatte die Art der Operationalisierung der akademischen Leistung über Noten, standardisierte Tests oder den erreichten Abschluss in dieser Metaanalyse keinen Einfluss auf die prädiktive Validität von Fähigkeitsselbstkonzept oder Selbstwirksamkeit. Insgesamt war die Vorhersagekraft beider Konstrukte mit $\beta = .08$ relativ gering ausgeprägt. In einer neueren Metaanalyse fand Huang (2012), dass das Fähigkeitsselbstkonzept im Vergleich zur Selbstwirksamkeit besser geeignet ist, um Noten vorherzusagen, wohingegen die Selbstwirksamkeit eine höhere Vorhersagekraft für PISA-Testergebnisse aufweist. Allerdings umfasste diese Metaanalyse nicht ausschließlich längsschnittliche Studien. Weitere Metaanalysen über Längsschnittstudien, die nur die Selbstwirksamkeit (Talsma et al., 2018) oder nur das Fähigkeitsselbstkonzept (Wu et al., 2021) betrachteten, fanden eine vergleichbare mittlere prädiktive Validität der beiden Konstrukte für akademische Leistung ($\beta = .07$ für die Selbstwirksamkeit und $\beta = .08$ für das Fähigkeitsselbstkonzept). Zusammenfassend zeigt sich also, dass das Fähigkeitsselbstkonzept und die Selbstwirksamkeit zwei konzeptionell verschiedene Formen der Selbsteinschätzung darstellen, die sich aber sowohl in ihrem empirischen Zusammenhang als auch in ihren Auswirkungen auf die akademische Leistung höchstens geringfügig unterscheiden. Beide Konstrukte weisen dabei einen schwach positiven Effekt auf die akademische Leistung auf. Der Fokus der folgenden Kapitel liegt auf dem Fähigkeitsselbstkonzept, da dieses Konstrukt in den empirischen Beiträgen herangezogen wurde, um die Selbsteinschätzung der Proband*innen zu repräsentieren. Eine zumindest grundlegende Darstellung der Selbstwirksamkeit und ihrer Effekte auf akademische Leistung ist allerdings ebenfalls relevant, da in Studien zu SE Bias Effekten zahlreiche unterschiedliche Formen der Selbsteinschätzung, darunter unter anderem die Selbstwirksamkeit, genutzt wurden (z.B. Hewitt, 2015; Ramdass, 2010; Talsma et al., 2019).

1.1.2 Mediatoren von Selbsteinschätzungseffekten auf akademische Leistung

Während Selbsteinschätzungseffekte auf akademische Leistungen gut empirisch abgesichert sind, ist wesentlich weniger über die Mechanismen bekannt, die diese Effekte vermitteln. In Bezug auf das Fähigkeitsselbstkonzept ist aus zahlreichen Studien zum reciprocal effects model bekannt, dass sich Fähigkeitsselbstkonzept und akademische Leistung wechselseitig beeinflussen (z.B. Arens & Niepel, 2023; Bakadorova & Raufelder, 2020; Fu et al., 2020; Marsh; 2022; Marsh et al., 2022; Marsh, 2023; Preckel et al., 2017; Seaton et al., 2015; Sewasew & Koester, 2019; Sewasew & Schroeders, 2019; Weidinger et al., 2018; Wolff et al., 2021a; Zhang et al., 2023). Dem liegt die Annahme zugrunde, dass der Effekt des Fähigkeitsselbstkonzepts auf die akademische Leistung durch motivationale Faktoren mediiert wird. In den Worten von Marsh und Martin (2012):

> Implicit in the rationale of the REM is the largely untested assumption that the effect of prior self-concept on subsequent achievement is mediated by student characteristics such as increased conscientious effort, persistence in the face of difficulties, enhanced intrinsic motivation, academic choice, and coursework selection … (S. 68).

Die in dem obigen Zitat angesprochenen Mediationen wurden bisher nur in einer relativ kleinen Zahl von Studien empirisch untersucht. Dennoch konnten einzelne Studien zeigen, dass Anstrengung, ein erhöhter Aufgabenfokus, intrinsische Motivation und verminderte Prüfungsangst die Effekte des Fähigkeitsselbstkonzepts auf die akademische Leistung mediieren (Areepattamannil, 2012; Cai et al., 2018; Guay et al., 2010; Helmke, 1990; Trautwein et al., 2009; Wigfield & Eccles, 2000). Ebenso wird im Erwartungs-Wert-Modell (Eccles & Wigfield; 2020; 2023; Wigfield, 1994; Wigfield et al., 2020) angenommen, dass die Effekte des Fähigkeitsselbstkonzepts auf akademische Leistungen über zwei Gruppen von motivationalen Variablen, den Erfolgserwartungen und den subjektiven Werten, mediiert werden. Diese Annahmen basieren ursprünglich auf der klassischen Erwartungs-Wert-Theorie von Atkinson (1964), nach der eine Person um so motivierter ist, eine bestimmte Handlung auszuführen, (1) je größer ihre Überzeugung ist, durch diese Handlung ein gewünschtes Ergebnis herbeiführen zu können (Erfolgserwartung) und (2) je größer der Wert ist, den sie diesem Ergebnis beimisst (subjektive Werte). Konkret sind die Erfolgserwartungen in dem Erwartungs-Wert-Modell von Eccles und Kollegen definiert als die individuellen Überzeugungen einer Person, in zukünftigen Aufgaben kurz- oder langfristig gute Ergebnisse erzielen zu können (Eccles & Wigfield, 2002, S. 119). Der Wert stellt in diesem Modell einen Faktor höherer Ordnung dar, der in drei verschiedene subjektive Wertekomponenten sowie wahrgenommene Kosten untergliedert ist (Eccles & Wigfield, 2023). Die drei subjektiven Wertekomponenten sind die intrinsischen Werte, die Wichtigkeitswerte und die Nützlichkeitswerte. Intrinsische Werte sind das Ausmaß an Freude, das eine Person bei der Ausübung einer Tätigkeit um ihrer selbst willen auch ohne extrinsischen Anreiz empfindet. Sie sind damit eng verwandt mit den Konzepten der intrinsischen Motivation und des Interesses (Wigfield & Eccles, 2020). Die Wichtigkeitswerte spiegeln das Ausmaß wieder, in dem es einer Person wichtig ist, in einem bestimmten Bereich oder einer bestimmten Aufgabe erfolgreich zu sein und ergeben sich nach Wigfield & Eccles (2020) aus der Übereinstimmung einer bestimmten Tätigkeit mit dem grundlegenden Selbstschema und der Identität einer Person. Beispielsweise sollte es jemanden, der sich selbst als Sportler oder sportliche Person identifiziert, wichtig sein, in Aufgaben, welche sportliche Leistung erfordern, gut abzuschneiden. Die Nützlichkeitswerte reflektieren das Ausmaß, in dem eine Person davon ausgeht, dass eine bestimmte Tätigkeit nützlich zum Erreichen ihrer kurzfristigen oder langfristigen Ziele ist und ist demnach verwandt mit extrinsischer Motivation (Wigfield & Eccles, 2020). Die letzte Komponente, die wahrgenommenen Kosten, beinhalten sowohl die Anstrengung, die eine Person für eine gewisse Tätigkeit aufbringen muss, als auch emotionale und psychologische Kosten sowie zeitliche und Opportunitätskosten. Die im Erwartungs-Wert-Modell angenommene Mediation des Effekts des Fähigkeitsselbstkonzepts auf akademische Leistung durch die beschriebenen Komponenten wurde nach meinem Wissen bisher in keiner Längsschnittstudie direkt empirisch überprüft. Dies liegt insbesondere daran, dass in Studien zwar die Effekte von Erfolgserwartung und subjektiven Werten auf akademische Leistung untersucht wurden, seltener aber die Effekte des Fähigkeitsselbstkonzepts auf Erfolgserwartung und subjektive Werte. Die Ergebnisse dieser Studien zeigen positive Effekte sowohl von Erfolgserwartungen (z.B. Benden & Lauermann, 2022; Breitwieser & Brod, 2022; Brown & Putwain, 2021; Fadda et al., 2020; Geng et al., 2022; Kiuru et al., 2020; Liu et al., 2022; Yeung et al., 2022; Wille et al., 2020) als auch von subjektiven Werten (z.B. Benden & Lauermann, 2022; Brown & Putwain, 2021; Geng et al., 2022; Kiuru et al., 2020; Yeung et al., 2022; Weidinger et al., 2020) auf akademische Leistung auf (aber siehe Liu et al., 2022; die lediglich einen Effekt der Erfolgserwartungen aber keinen Effekt der subjektiven Werte nachweisen konnten). Allerdings weisen die Befunde einiger Studien darauf hin, dass Nützlichkeitswerte für die akademische Leistung von geringerer Bedeutung sein könnten, als intrinsische Werte und Wichtigkeitswerte. So fanden etwa Li et al. (2021) lediglich einen positiven Effekt von Wichtigkeitswerten, nicht aber von Nützlichkeitswerten auf akademische Leistung. Der Effekt von intrinsischen Werten wurde in dieser Studie nicht untersucht. Ebenso fanden Wille et al. (2020) zwar einen positiven Effekt eines gemeinsamen Faktors der intrinsischen Werte und Wichtigkeitswerte, aber keinen Effekt der Nützlichkeitswerte. Andere Autorinnen hingegen überprüften explizit, ob sich die Effekte der drei subjektiven Wertekomponenten auf akademische Leistung unterscheiden und konnten keine signifikanten Unterschiede feststellen (Weidinger et al., 2020). Insgesamt spricht die Befundlage somit dafür, dass sowohl Erfolgserwartungen als auch subjektive Werte einen positiven Einfluss auf akademische Leistung ausüben, wobei die Ergebnisse zu Nützlichkeitswerten im Vergleich zu den anderen Wertekomponenten und Erfolgserwartung etwas weniger eindeutig erscheinen. Um eine Mediation, wie sie im Erwartungs-Wert-Modell angenommen wird, empirisch abzusichern, müsste allerdings zusätzlich nachgewiesen werden, dass sich das Fähigkeitsselbstkonzept positiv auf Erfolgserwartungen und subjektive Werte auswirkt. Vermutlich wurden Mediationsstudien in diesem Forschungskontext bisher auch deshalb nicht vorgenommen, da Fähigkeitsselbstkonzept, Erfolgserwartung und subjektive Werte als sämtlich motivationale Variablen sowohl konzeptuell als auch empirisch eng verwandt sind. Insbesondere die Erfolgserwartung weist deutliche inhaltliche Überschneidungen mit dem Fähigkeitsselbstkonzept (wie auch der Selbstwirksamkeit) auf (Wigfield & Eccles, 2020). Diese Tatsache erschwert es, die jeweiligen Konstrukte in empirischen Arbeiten voneinander zu trennen (Eccles et al., 1993; Eccles & Wigfield, 1995; Marsh et al., 2019). Aus diesem Grund haben zahlreiche Autor*innen die Begriffe austauschbar verwendet und Messinstrumente für das Fähigkeitsselbstkonzept und die Erfolgserwartung zusammengefügt oder das eine als Maß für das jeweils andere Konstrukt verwendet (z.B. Gaspard et al., 2018; Guo et al., 2017; Jansen et al., 2021; Nagengast et al., 2011; Trautwein et al., 2012; Wang et al., 2012). Inzwischen betrachten die Autor*innen des Modells es als „vermutlichen Fehler“ (Eccles & Wigfield, 2020, S. 3), die Erfolgserwartungen und Fähigkeitsselbstkonzepte in ihren empirischen Arbeiten nicht voneinander getrennt zu haben. Zwar gibt es nach meinem

Wissen keine empirischen Längsschnittstudien zu den oben diskutierten Mediationseffekten auf akademische Leistung, allerdings wurde vereinzelt untersucht, ob Erfolgserwartungen und subjektive Werte die Effekte des Fähigkeitsselbstkonzepts auf akademische Anstrengung und Entscheidungsverhalten mediieren (Mac Iver et al., 1991; Priess-Groben & Hyde, 2017). Allerdings konnten in diesen Studien keine signifikanten Mediationen in Modellen mit akzeptablen Fitindizes nachgewiesen werden. Alles in allem gibt es somit noch keine eindeutigen empirischen Hinweise darauf, dass Erfolgserwartungen oder subjektive Werte die Effekte des Fähigkeitsselbstkonzepts auf akademische Leistung mediieren. Aufgrund der Popularität des Erwartungs-Wert-Modells in der pädagogischen Psychologie wäre eine empirische Untersuchung der angenommenen Mediationseffekte somit bedeutsam.

1.2 SE Bias Effekte

1.2.1 Definition

Der SE Bias wird in der Literatur definiert als die Diskrepanz zwischen einer Selbsteinschätzung (z.B. Fähigkeitsselbstkonzept oder Selbstwirksamkeit in Mathematik, selbsteingeschätzte Intelligenz) und einem möglichst objektiven Außenkriterium für die jeweilige Selbsteinschätzung (z.B. Ergebnis eines standardisierten Mathematikkompetenztests, Ergebnis eines Intelligenztests; z.B. Bonneville-Roussy et al., 2017; Humberg et al. 2018; 2019a; Lee, 2021). Ein SE Bias kann somit positiv sein (wenn die Ausprägung der Selbsteinschätzung die Ausprägung des Außenkriteriums übersteigt; Selbstüberschätzung), negativ sein (wenn die Ausprägung des Außenkriteriums die Ausprägung der Selbsteinschätzung übersteigt; Selbstunterschätzung), oder einen Wert von null haben (wenn die Ausprägung der Selbsteinschätzung genau der Ausprägung des Außenkriteriums entspricht; akkurate Selbsteinschätzung). Der SE Bias ist alo mathematisch gesprochen kein Absolutwert, da negative Werte nicht durch Multiplikation mit -1 in positive Werte umgewandelt werden. In der Literatur wird ein entsprechender nicht absoluter Wert meist als „bias“ bezeichnet, ein absoluter Wert hingegen als „miscalibration“ (siehe Talsma et al., 2019). Die miscalibration ist demnach also der Betrag des SE Bias (|SE Bias|). Allerdings wird die miscalibration in der vorliegenden Arbeit aus folgenden Gründen nicht näher thematisiert. Einige Autor*innen gehen, wie in den folgenden Abschnitten erläutert wird, davon aus, dass ein positiver SE Bias (also eine eine Selbstüberschätzung) sich anders auf akademische Leistung auswirkt als ein negativer SE Bias (eine Selbstunterschätzung). Dies wäre zum Beispiel dann der Fall, wenn eine Selbstüberschätzung positive, eine Selbstunterschätzung aber negative Effekte auf akademische Leistungen hat. Würde man statt dem SE Bias also die miscalibration heranziehen, könnten entsprechende Unterschiede zwischen Effekten einer Selbstüberschätzung und einer Selbstunterschätzung nicht nachgewiesen werden, weshalb es sinnvoller ist, den SE Bias heranzuziehen. Ein SE Bias Effekt kann definiert werden, als eine Veränderung in einem SE Bias, die ursächlich zu einer Veränderung in einer anderen Variablen, beispielsweise akademischer Leistung, führt.

1.2.2 Theoretischer Hintergrund und empirische Befunde

In der pädagogischen Psychologie wird seit längerem die Frage diskutiert, ob und falls ja welche Effekte ein derartiger Bias selbst eingeschätzter Kompetenzen auf akademische Leistung hat (z.B. Baumeister, 1989; Bonneville-Roussy et al., 2017; Côté et al., 2014; Leduc & Bouffard, 2017; Lee, 2021; 2022; Lopez et al., 1998; Martin & Debus, 1998; Praetorius et al., 2016; Robins & Beer, 2001; Rohr & Ayers, 1973; Talsma et al., 2019; Taylor & Brown, 1988; 1994; Wright, 2000). Eine häufig vertretene Ansicht ist dabei, dass es für Schüler*innen von Vorteil ist, ihre eigenen Kompetenzen zu überschätzen, da dies zu einer Steigerung der Motivation und darüber vermittelt zu einer Steigerung der akademischen Leistung führt, während eine Selbstunterschätzung sich negativ auswirkt (z.B. Bonneville-Roussy et al., 2017; Helmke, 1998; Lee, 2021; Martin & Debus, 1998; Taylor & Brown, 1988; 1994; Wright, 2000). Demnach läge also ein positiver linearer SE Bias Effekt vor, da ein größerer SE Bias (eine geringere Selbstunterschätzung oder eine größere Selbstüberschätzung) zu besserer akademischer Leistung führen würde. In einem einflussreichen Artikel (3650 Zitationen in der Datenbank PsycINFO, Stand 03.07.2023) argumentieren Taylor und Brown (1988) für die Adaptivität „positiver Illusionen" in zahlreichen Lebensbereichen. Unter anderem sollen Selbstüberschätzungen von Vorteil sein, da sie Motivation und Aufgabenpersistenz und darüber vermittelt produktive und kreative Leistungen verbessern. Obwohl der Artikel von Taylor und Brown (1988) mehrfach für die unzureichende theoretische und empirische Fundierung seiner Schlussfolgerungen kritisiert wurde (z.B. Asendorpf & Ostendorf, 1998; Block & Colvin, 1994; Colvin & Block, 1994; Colvin et al., 1995; Humberg et al., 2018; Jopling, 1996; Robins & Beer, 2001), wurde die zugrundeliegende Idee unter anderem im Kontext von SE Bias Effekten auf akademische Leistung immer wieder aufgegriffen (z.B. Bonneville-Roussy et al., 2017; Bouffard et al., 2011; Chung et al., 2016; Côté et al., 2014; Dupeyrat et al., 2011; Lee, 2021; Lopez et al., 1998). Die Kritik an der Argumentation von Taylor & Brown (1988) ist vielfältig, doch der für die vorliegende Diskussion zentrale Kritikpunkt ist die mangelnde Differenzierung zwischen „positiven Illusionen" (beziehungsweise einem positiven SE Bias) und der Selbsteinschätzung per se. So argumentieren die Autor*innen etwa für Effekte positiver Illusionen auf produktive und kreative Leistung auf der Grundlage von Studien, die fanden, dass positive Selbsteinschätzungen mit längerem und härterem Arbeiten an Aufgaben assoziiert sind (Taylor & Brown, 1988, S. 199; siehe Humberg et al., 2018 für eine Diskussion). Beispielsweise fand Felson (1984) in einer Längsschnittstudie, dass sich die Selbsteinschätzung eigener Kompetenzen, vermittelt über Anstrengung, positiv auf spätere Schulnoten von Zehnt- bis Zwölftklässlern auswirkt. Während derartige Befunde relevante Erkenntnisse über Selbsteinschätzungseffekte liefern, stellen sie, entgegen der Darstellung von Taylor und Brown (1988), allerdings keinen Nachweis von Effekten positiver Illusionen beziehungsweise von SE Bias Effekten dar. Dass eine hohe Selbsteinschätzung einen positiven Einfluss auf ein bestimmtes Resultat hat, bedeutet nicht zwangsläufig, dass auch der SE Bias einen positiven Einfluss haben muss. Da es sich bei einem SE Bias Effekt um den

Effekt einer Diskrepanz handelt, liegt ein positiver SE Bias Effekt nur dann vor, wenn eine größere Diskrepanz zwischen Selbsteinschätzung und Außenkriterium und nicht nur eine höher ausgeprägte Selbsteinschätzung per se zu einem positiven Resultat führt. Somit muss in jeder theoretischen Argumentation für positive SE Bias Effekte dargelegt werden, weshalb es speziell die Diskrepanz zwischen der Selbsteinschätzung und dem Außenkriterium statt der Selbsteinschätzung per se sein sollte, die einen positiven Effekt ausübt. Es müssen die Mechanismen aufgezeigt werden, die für einen solchen positiven SE Bias Effekt verantwortlich sein sollen. Beispielsweise ist die oben genannte Argumentation von Taylor und Brown (1988), dass positive Illusionen vorteilhaft für produktive und kreative Leistungen sind, weil sie die Motivation von Personen steigern, auch inhaltlich ohne weitere Erläuterungen wenig überzeugend. Denn es stellt sich die Frage, weshalb es speziell die positive Illusion statt einer positiven Selbsteinschätzung per se sein sollte, die die Motivation einer Person fördert. Dies kann an folgendem Beispiel verdeutlicht werden: Angenommen eine Person (Julia) verfügt sowohl über ein hohes Fähigkeitsselbstkonzept in Mathematik als auch eine hohe Kompetenz in Mathematik, während eine zweite Person (Maria) über ein mittleres Fähigkeitsselbstkonzept und eine niedrige Kompetenz in Mathematik verfügt. Folgt man der Argumentation von Taylor und Brown (1988) über positive Illusionen, sollte Maria unter sonst gleichen Bedingungen in Mathematik mehr Motivation und Anstrengung zeigen als Julia, da sie ihre eigenen Fähigkeiten überschätzt, sie also im Gegensatz zu Julia über eine positive Illusion in Bezug auf ihre mathematischen Fähigkeiten verfügt. Doch dies scheint wenig plausibel zu sein. Sollte nicht jene Person, die sich selbst für kompetenter hält (Julia), mehr Motivation und Anstrengung zeigen, da sie im Sinne der Erwartungs-Wert-Theorie eher als Maria erwarten würde, durch ihre Anstrengung ein wünschenswertes Resultat erzielen zu können? Weshalb sollten Schüler*innen, die über ein mittleres Fähigkeitsselbstkonzept verfügen, sich besonders anstrengen, nur weil ihre Kompetenz niedriger ausgeprägt ist, als ihr Fähigkeitsselbstkonzept? Mit diesem Beispiel soll nicht argumentiert werden, dass es keine positiven SE Bias Effekte geben kann, aber es soll auf die Notwendigkeit stringenter theoretischer Argumentationen für SE Bias Effekte hingewiesen werden, in denen explizit herausgestellt wird, weshalb es der Bias statt der positiven Selbsteinschätzung per se sein sollte, der einen Effekt ausübt. Diese Frage wird von Taylor & Brown (1988) nicht beantwortet. Dieses Problem ist auch in der neueren Literatur nach wie vor relevant. So argumentieren etwa sowohl Bonneville-Roussy et al. (2017, S. 9) als auch Lee (2021, S. 459-460), dass ein positiver SE Bias als eine „selbsterfüllende Prophezeiung" positiv auf akademische Leistung einwirken könnte. Demnach würde ein positiver SE Bias dazu führen, dass eine Person sich entsprechend dieses Bias verhält und daher mehr Engagement und Aufgabenpersistenz, etwa bei Hausaufgaben, zeigt, Fehler external attribuiert, und auftretende Schwierigkeiten eher als Herausforderung interpretiert. Diese positiven Effekte wiederum sollen zu besseren akademischen Leistungen führen. Allerdings liegt dieser Argumentation dasselbe Problem zugrunde wie der Argumentation von Taylor & Brown (1988): Es wird nicht deutlich, weshalb es der SE Bias statt der Selbsteinschätzung an sich sein sollte, der den Effekt bewirkt. Folgt man der Logik, dass Schüler*innen sich im Sinne einer selbsterfüllenden Prophezeiung auf eine Art und Weise verhalten, die ihrem Selbstbild entspricht, erscheint es plausibler, dass es auf die Höhe der Selbsteinschätzung per se statt auf den SE Bias ankommt. Insbesondere in der Literatur zu positiven SE Bias Effekten mangelt es somit bisher an einer fundierten theoretischen Grundlage, um derartige Effekte abzuleiten.

Trotz der oben beschriebenen theoretischen Mängel in den weitaus meisten Studien zu SE Bias Effekten auf akademische Leistung gehen die meisten Autor*innen von positiven Effekten aus und fanden oft auch entsprechende empirische Befunde (z.B. Bonneville-Roussy et al., 2017; Bouffard et al., 2011; Chung et al., 2016; Côté et al., 2014; Dupeyrat et al., 2011; Kurman, 2006; Leduc & Bouffard, 2017; Lee, 2021; Lopez et al., 1998; Martin & Debus, 1998; Praetorius et al., 2016; Willard & Gramzow, 2009; Wright, 2000). Allerdings weisen nahezu alle Studien in diesem Bereich nicht nur theoretische sondern auch methodische

Mängel auf, welche die Interpretierbarkeit der Befunde stark einschränken. Diese methodischen Mängel werden im folgenden Kapitel detailierter erläutert und sollen an dieser Stelle nur kurz zusammengefasst werden. Das zentrale Problem besteht darin, dass Methoden zur Untersuchung von SE Bias Effekten gewählt wurden, die SE Bias Effekte mit Selbsteinschätzungseffekten konfundieren. Somit kann nicht ausgeschlossen werden, dass es sich bei den augenscheinlich gefundenen SE Bias Effekten um Artefakte von Selbsteinschätzungseffekten handelt (siehe Humberg et al., 2018; 2019a).

Manche Autor*innen gehen außerdem davon aus, dass lediglich ein moderat positiver SE Bias förderlich für akademische Leistung ist, während sich ein extremer SE Bias negativ auswirken soll (Baumeister, 1989; Taylor und Brown, 1994). In seiner Theorie einer „optimal margin of illusion" bezieht sich Baumeister (1989) unter anderem auf den oben diskutierten Artikel von Taylor & Brown (1988). Er argumentiert, dass eine Überschätzung eigener Kompetenzen im akademischen Kontext aber auch in anderen für die vorliegende Arbeit weniger relevanten Kontexten sich im Sinne einer selbsterfüllenden Prophezeiung positiv auf Leistung auswirken kann (S. 182). Diese Argumentation entspricht also soweit jener von Bonneville-Roussy et al. (2017, S. 9) und Lee (2021, S. 459-460). Sei die Selbstüberschätzung hingegen zu hoch ausgeprägt, könne sie sich auch negativ auf die Leistung auswirken, da Personen ihre Anstrengungen, gute Leistungen zu erzielen, reduzieren („self-handicapping behavior", S. 177-178). Demnach gäbe es, so Baumeister (1989), ein gewisses Maß an Selbstüberschätzung, dass für das Erbringen von Leistungen optimal ist, während bei noch höher ausgeprägter Selbstüberschätzung, die erbrachten Leistungen wieder abnehmen. Nach dieser Theorie liegt also kein positiver linearer SE Bias Effekt vor, sondern ein kurvilinearer Effekt, welcher bis zu einem gewissen positiven Wert des SE Bias positiv ist und darüber hinaus negativ wird. Allerdings weist auch diese Theorie das Problem einer mangelnden Differenzierung zwischen SE Bias Effekten und Selbsteinschätzungseffekten auf. Zum einen ist die Annahme „selbst erfüllender Prophezeiungen" wie bereits diskutiert, logisch geeigneter um Selbsteinschätzungseffekte statt SE Bias Effekte zu erklären. Auch die Annahme, dass eine zu hoch ausgeprägte Selbstüberschätzung zu einer Abnahme an Anstrengung führe, ist nicht ohne weitere Erklärung unmittelbar ersichtlich. Weshalb sollte eine Person, die ihre Kompetenz in Mathematik als hoch einschätzt, tatsächlich aber über eine niedrige Kompetenz verfügt (hohes Maß an Selbstüberschätzung) eher dazu tendieren, ihre Anstrengungen zu reduzieren, als eine Person die ihre Kompetenz als durchschnittlich einschätzt und tatsächlich über eine durchschnittliche Kompetenz verfügt (keine Selbstüberschätzung)? Zumindest bedarf eine solche Annahme einer näheren theoretischen Erläuterung, welche von Baumeister (1989) nicht angeführt wird. Denkbar wäre etwa, dass zu hohe Selbstüberschätzungen zu Enttäuschungen im Angesicht von Leistungsrückmeldungen führen, welche dann zukünftige Anstrengungen und Motivation verringern. Eine solche Argumentation wäre logisch plausibel, da sie sich speziell auf den SE Bias und nicht auf die Selbsteinschätzung per se bezieht. Eine hohe Selbsteinschätzung per se reicht nicht aus, um zu Enttäuschungen zu führen. Nur wenn die Selbsteinschätzung unrealistisch hoch ist (positiver SE Bias) und somit vermutlich in Kontrast zu zukünftigen Leistungsrückmeldungen stehen wird, wäre dies der Fall. Eine derartige Überlegung ist allerdings rein hypothetisch und wurde nach meinem Wissen bisher weder explizit aufgestellt noch empirisch überprüft. Sie kann allerdings als Beispiel dafür dienen, wie theoretische Argumentationen für SE Bias Effekte aufgebaut sein können, wenn diese unabhängig von Selbsteinschätzungseffekten gelten sollen. Die Theorie der optimal margin of illusion wurde im akademischen Kontext nur relativ selten empirisch überprüft. Die vorhandenen Studien konnten jedoch sämtlich keine entsprechenden Effekte auf akademische Leistung nachweisen (Lee, 2021; Lopez et al., 1998; Praetorius et al., 2016; Wright, 2000). Allerdings sind diese Studien ebenfalls von den oben angesprochenen methodischen Mängeln betroffen, weshalb ihre Ergebnisse nur mit Vorsicht zu interpretieren sind.

Gleichwohl die meisten Autor*innen in diesem Forschungsfeld von positiven SE Bias Effekten auf akademische Leistung ausgehen, gibt es in der Literatur, neben den oben diskutierten Schwächen der genannten Studien, auch inhaltliche Argumente gegen diese Annahme. Studien im Bereich des metakognitiven selbstregulierten Lernens zeigen, dass Lernende Wissen darüber, was sie bereits wie gut gelernt haben, nutzen, um weitere Lernprozesse effektiv zu motivieren und zu steuern (z.B. Dunslosky & Rawson, 2012; Hacker & Bol, 2019; Hadwin & Webster, 2013; Händel et al., 2020; Hewitt, 2015; Hong et al., 2020; Kim et al., 2010; van Loon & Oeri, 2023). Hadwin und Webster (2013) fassen die für die vorliegende Arbeit relevante Interpretation wie folgt zusammen:

> It may be that students who are overconfident take this as a sign that they do not need to adapt their learning or studying, whereas those who are relatively less overconfident (or perhaps even under-confident) respond by increasing their levels of regulation. (S. 45)

Es sollte jedoch erwähnt werden, dass die in der oben zitierten Studie vorsichtig formulierte Möglichkeit potentieller Vorteile von Selbstunterschätzungen nur von wenigen Forschenden explizit vertreten wird. Die weitaus meisten Forschenden zum selbstregulierten Lernen sprechen sich hingegen für die positive Bedeutung akkurater Selbsteinschätzungen aus (Cogliano et al., 2021; Gutierrez de Blume, 2022; siehe auch die oben zitierten Studien). Somit lässt sich aus den Theorien und Befunden dieses Forschungsfelds ableiten, dass eine akkurate Selbsteinschätzung zum Erbringen akademischer Leistungen optimal sein könnte, da sie eine effektivere Regulation von Lernverhalten erlaubt. Allerdings sind bei dieser Schlussfolgerung zwei Einschränkungen zu beachten: 1) sind auch einige Studien zur Akkuratesse von Selbsteinschätzungen beim selbstregulierten Lernen von dem Problem betroffen, dass die Effekte von Selbstüber- und -unterschätzungen empirisch nicht völlig von den Effekten der Selbsteinschätzung an sich getrennt wurden (z.B. Dunlosky & Rawson, 2012). 2) wird Selbstüber- und -unterschätzung in Studien zum selbstregulierten Lernen meist anders operationalisiert als in Studien, die sich spezifisch mit SE Bias Effekten auf akademische Leistung befassen. In den meisten SE Bias Studien bearbeiten die Proband*innen zunächst ein Instrument zur Selbsteinschätzung der Kompetenz in einem bestimmten akademischen Bereich (z.B. Fähigkeitsselbstkonzept oder Selbstwirksamkeit) gefolgt von einem Instrument zur Erfassung der jeweiligen Kompetenz als Außenkriterium (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côté et al, 2014; Leduc & Bouffard, 2017; Lee, 2021; Praetorius et al., 2016). Hingegen liegt der Fokus in Studien zum selbstregulierten Lernen stärker auf dem Lernprozess selbst, statt auf relativ allgemeinen Selbsteinschätzungen. Meist bearbeiten die Proband*innen zunächst gewisse Leistungstestaufgaben und schätzen anschließend ein, wie gut sie in diesen Aufgaben abgeschnitten haben, beziehungsweise wie sicher sie sind, die Aufgaben korrekt gelöst zu haben (z.B. Dunlosky & Rawson, 2012; Gutierrez de Blume et al., 2022; Hadwin & Webster, 2013; Kim et al., 2010; Spencer et al., 2023; van Loon & Oeri, 2023). Aus der Diskrepanz zwischen der tatsächlichen und der eingeschätzten Leistung wird dann, ähnlich wie in SE Bias Studien, auf mögliche Selbstüber- und -unterschätzungen geschlossen. Somit bestehen zwei bedeutsame Unterschiede zwischen den beiden Forschungsfeldern. Zum einen unterscheidet sich die typische zeitliche Abfolge der Messungen (in SE Bias Studien meist zuerst die Selbsteinschätzung gefolgt von dem Leistungstest, in Studien zum selbstregulierten Lernen umgekehrt). Zum anderen unterscheiden sich die eingesetzten Instrumente zur Selbsteinschätzung (in SE Bias Studien meist relativ allgemeine Einschätzung eigener Kompetenzen, in Studien zum selbstregulierten Lernen meist Einschätzung der Leistung in spezifischen Aufgaben). Diese unterschiedliche Operationalisierung von Selbstüber- und -unterschätzungen kann zu unterschiedlichen Ergebnissen führen (Lee, 2022). Somit sind die Ergebnisse der beiden Forschungsfelder zwar nicht unmittelbar aufeinander übertragbar, dennoch weisen die Theorien und Befunde zum selbstregulierten Lernen auf eine Alternative zu der häufig vertretenen Hypothese positiver SE Bias Effekte hin. Nach dieser alternativen

Hypothese wäre eine akkurate Einschätzung eigener Kompetenzen optimal für das Erbringen akademischer Leistung, da akkurate Selbsteinschätzungen die effektivsten Lernprozesse ermöglichen. Nach dieser Hypothese läge also, ähnlich wie nach der optimal margin Theorie, ein kurvilinearer SE Bias Effekt auf akademische Leistung vor. Dieser wäre umgekehrt „u-förmig“, da die akademische Leistung bei einem SE Bias von null (einer akkuraten Selbsteinschätzung) im Mittel am höchsten ausgeprägt wäre und sowohl bei negativen als auch bei positiven SE Bias Werten abnähme.

In einzelnen Arbeiten wird zudem auch von negativen SE Bias Effekten berichtet (Rohr & Ayers, 1973; Talsma et al., 2019). Talsma et al. (2019) berichten, dass Selbstunterschätzungen (erfasst über den Bias zwischen Selbstwirksamkeit und objektiv erfasster Leistung) im Vergleich zu akkuraten Selbsteinschätzungen und Selbstüberschätzungen bessere akademische Leistungen vorhersagen. Rohr und Ayers (1973) argumentieren, dass insbesondere hochleistende Schüler*innen Selbstunterschätzung als eine Technik nutzen, um ihre Motivation zu steigern. Diese Annahme ist vergleichbar mit der oben angesprochenen These von Hadwin und Webster (2013), dass möglicherweise sogar Selbstunterschätzung zu einer erhöhten Selbstregulation führen kann. Auch wenn derartige Annahmen von relativ wenigen Autor*innen vertreten werden, lässt sich aus ihnen somit die zu überprüfende Hypothese ableiten, dass negative lineare SE Bias Effekte bestehen könnten. Selbstunterschätzung wäre somit förderlich, Selbstüberschätzung hingegen hinderlich für akademische Leistung.

Manche Autor*innen fanden keine SE Bias Effekte auf akademische Leistung (Gonida & Leondari, 2011), beziehungsweise heterogene Effekte, die teils positiv, teils negativ und teils nicht signifikant waren (Robins & Beer, 2001). Robins und Beer (2001) argumentieren, dass eine Selbstüberschätzung eigener Kompetenzen und Leistungen zwar kurzfristig adaptiv sein kann, langfristig aber vor allem negative Effekte mit sich bringt.

Fazit. Zusammenfassend lassen sich aus der Literatur somit fünf konkurrierende Hypothesen über SE Bias Effekte auf akademische Leistung ableiten: 1) einen positiven linearen Effekt, 2) einen optimal margin Effekt, 3) einen umgekehrt „u-förmigen" Effekt, 4) einen negativen linearen Effekt und 5) die Hypothese, dass keine SE Bias Effekte auf akademische Leistung bestehen, sondern lediglich positive lineare Effekte der Selbsteinschätzung (und der Kompetenz) per se. Es kann also festgehalten werden, dass es noch keine allgemein akzeptierte und fundierte Theorie zu SE Bias Effekten auf akademische Leistungen gibt. Zwar betonen die meisten Autor*innen die positiven Auswirkungen von Selbstüberschätzungen im akademischen Kontext, doch sind die entsprechenden Studien mit erheblichen theoretischen und methodischen Mängeln behaftet, die bereits seit längerem (z.B. Colvin & Block, 1994) aber auch in jüngeren Diskussionen (z.B. Humberg et al., 2018; 2019a) immer wieder thematisiert werden. Die theoretischen Mängel wurden in diesem Abschnitt dargestellt. Im Folgenden Abschnitt werden die methodischen Herausforderungen bei der Untersuchung von SE Bias Effekten behandelt.

1.2.3 Methodische Herausforderungen

Im vorangegangenen Kapitel wurde erläutert, dass ein SE Bias Effekt nur dann vorliegt, wenn eine Diskrepanz zwischen einer Selbsteinschätzung und einem mit dieser Selbsteinschätzung in Beziehung stehendem Außenkriterium eine Wirkung auf ein gewisses Merkmal (z.B. akademische Leistung) ausübt. Um entsprechende Effekte zu untersuchen, wurden in der Literatur unterschiedliche Verfahren verwendet, welche sich in zwei Klassen unterteilen lassen, den sogenannten „Ein-Schritt-Ansatz" und den „Zwei-Schritte-Ansatz" (siehe Humberg et al., 2018 für eine Übersicht). Die entsprechenden Ansätze und die ihnen zugehörigen spezifischen methodischen Verfahren werden im Folgenden mit Hinblick auf ihre Stärken, Schwächen und Einsatzmöglichkeiten diskutiert. Dabei wird zudem die Bedeutung der jeweiligen Ansätze für die konkrete inhaltliche Fragestellung nach SE Bias Effekten auf akademische Leistung behandelt. Da der Ein-Schritt-Ansatz entwickelt wurde, um die methodischen Schwächen des Zwei-Schritte-Ansatzes zu überwinden, wird der Zwei-Schritte-Ansatz als erstes dargestellt.

1.2.3.1 Der Zwei-Schritte-Ansatz

Verfahren, die dem Zwei-Schritte-Ansatz zuzuordnen sind, laufen entsprechend ihres Namens in zwei methodischen Schritten ab. Im ersten Schritt wird für jede Person ein Wert ihres individuellen SE Bias berechnet. Im zweiten Schritt wird der errechnete SE Bias in weiterführenden Analysen, beispielsweise Regressionsanalysen (z.B. Lee, 2021), Cross-Lagged-Panel-Designs (z.B. Praetorius et al., 2016), oder Gruppenvergleichen (z.B. Bonneville-Roussy et al., 2017), genutzt, um andere Variablen, wie akademische Leistung, vorherzusagen. Wie genau die individuellen SE Bias Werte berechnet werden, unterscheidet sich zwischen den Studien. Die zwei am häufigsten verwendeten Berechnungsmethoden sind algebraische Differenzwerte und Residualwerte. Diese werden im Folgenden näher erläutert und mit Hinblick auf ihre Stärken und Schwächen diskutiert.

Algebraische Differenzwerte: Algebraische Differenzwerte sind die vermutlich intuitiv naheliegendste Möglichkeit, SE Bias Effekte zu untersuchen. Konkret wird bei diesem Verfahren der individuelle SE Bias Wert jeder Person berechnet als die algebraische Differenz aus der Selbsteinschätzung und dem Außenkriterium. Beispielsweise könnte ein SE Bias in mathematischer Kompetenz berechnet werden, indem das Ergebnis eines standardisierten Mathematikkompetenztests von der Selbsteinschätzung eigener mathematischer Kompetenzen subtrahiert wird. Wichtig dabei ist, dass die Selbsteinschätzung und das Außenkriterium auf derselben Skala erfasst werden (Edwards, 1994; Humberg et al., 2018). Beispielsweise könnten Schüler*innen ihren Prozentrang in mathematischer Kompetenz verglichen mit Gleichaltrigen einschätzen. Dieser eingeschätzte Prozentrang könnte dann mit dem erreichten Prozentrang in einem Mathematikkompetenztest verglichen werden. In manchen Studien ist die Berechnung der algebraischen Differenzwerte komplexer, da beispielsweise für bestimmte Antworttendenzen der Proband*innen kontrolliert wird

(Kwan et al., 2004), oder von der Selbsteinschätzung gleichzeitig ein objektives Außenkriterium (z.B. das Ergebnis eines Kompetenztests) als auch eine Fremdeinschätzung subtrahiert wird (z.B. Kwan et al., 2004; Moore & Healy, 2008; siehe Humberg et al., 2018). Algebraische Differenzwerte wurden vor allem in älteren Studien zu SE Bias Effekten eingesetzt (siehe z.B. Colvin et al., 1995; Polzer et al., 1997; Wylie, 1961), finden aber auch in neueren Studien hin und wieder Anwendung (z.B. Dupeyrat et al., 2011; Moore & Healy, 2008; Noble et al., 2011). Kritiken an der Verwendung von algebraischen Differenzwerten im Kontext der Forschung zu SE Bias Effekten (Connell & Illardi, 1987; Griffin et al., 1999; John & Robins, 1994; Paulhus & John, 1998; Robins & John, 1997) haben zahlreiche Autor*innen dazu bewogen, diese nicht mehr zu benutzen und stattdessen zumeist Residualwerte zu verwenden (z.B. Bonneville-Roussy et al., 2017; Côté et al., 2014; Leduc & Bouffard, 2017; Lee, 2021; 2022; Praetorius et al., 2016). Die besagte Kritik bezieht sich darauf, dass es bei der Verwendung von algebraischen Differenzwerten zur Untersuchung von SE Bias Effekten zu einer Konfundierung von SE Bias Effekten mit Selbsteinschätzungseffekten kommt. Dies ist leicht nachvollziehbar: Subtrahiert man von einer Selbsteinschätzung ein Außenkriterium, ist die resultierende Differenz umso größer je größer der Wert der Selbsteinschätzung ist. Wird diese Differenz dann eingesetzt, um andere Variablen vorherzusagen, kann nicht mehr statistisch unterschieden werden, ob tatsächlich die Diskrepanz zwischen der Selbsteinschätzung und dem Außenkriterium oder die Selbsteinschätzung per se ursächlich für einen gefundenen Zusammenhang verantwortlich ist (Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2018; 2019a). Ebenso sind algebraische Differenzwerte zudem mit dem Außenkriterium konfundiert. Denn je größer der Wert des Außenkriteriums ist, desto kleiner ist der Wert der Differenz zwischen Selbsteinschätzung und Außenkriterium. Findet man also empirisch einen Zusammenhang eines algebraischen Differenzwerts und einer zeitlich nachfolgend erfassten weiteren Variablen, so kann dieser Zusammenhang ursächlich durch unterschiedliche Prozesse entstanden sein: 1) durch einen tatsächlichen positiven SE Bias Effekt, 2) durch einen positiven Effekt der Selbsteinschätzung, 3) durch einen negativen Effekt des Außenkriteriums, oder 4) durch eine Mischform der drei zuvor genannten Prozesse (Humberg et al., 2018). Natürlich kann zudem auch der Einfluss von Störfaktoren wie Drittvariablen oder ein unterschiedlicher Ausgangswert der abhängigen Variablen für den empirisch gefundenen Zusammenhang verantwortlich sein, doch sind diese Probleme nicht spezifisch für die SE Bias Forschung, sondern betreffen längsschnittliche Studiendesigns im Allgemeinen und sind daher für die aktuelle Diskussion von geringerer Bedeutung. Es sollte angemerkt werden, dass algebraische Differenzwerte in der SE Bias Forschung nicht grundsätzlich zu kritisieren sind. Ist das Ziel lediglich die Quantifizierung des SE Bias einer Person oder Gruppe, so ist die Berechnung algebraischer Differenzwerte ein angemessenes Vorgehen (Humberg et al., 2018). Zwar sind algebraische Differenzwerte mit der Selbsteinschätzung und dem Außenkriterium konfundiert, doch verringert dies nicht ihre Validität als Maß des SE Bias. Lediglich wenn der Zusammenhang zwischen dem SE Bias und einer oder mehreren weiteren Variablen berechnet werden soll, führt dies zu den oben beschriebenen Problemen.

Residualwerte. Aufgrund der beschriebenen Nachteile algebraischer Differenzwerte, empfehlen einige Autor*innen als Alternative Residualwerte zu verwenden (Connell & Illardi, 1987; John & Robins, 1994; Paulhus & John, 1998; Robins & John, 1997). Dieses Vorgehen wird manchmal als *self-criterion residual strategy* (SCR; John & Robins, 1994; Paulhus & John, 1998; Robins & John, 1997) bezeichnet. Bei diesem Verfahren wird eine lineare Regression der Selbsteinschätzung auf das Außenkriterium gerechnet. Die Residuen dieser Regression werden als Werte des SE Bias abgespeichert. Ähnlich wie bei algebraischen Differenzwerten werden auch bei diesem Verfahren manchmal komplexere Werte für das Außenkriterium verwendet, die zum Beispiel sowohl ein objektives Außenkriterium als auch eine Fremdeinschätzung beinhalten oder bei denen für Antworttendenzen kontrolliert wird (z.B. Dufner et al., 2012; Leising et al., 2016). Aufgrund der Empfehlung in verschiedenen Artikeln, ist die Berechnung des SE Bias über Residualwerte inzwischen zumindest in der Literatur zu SE Bias Effekten auf akademische Leistung der bei weitem verbreitetste Ansatz (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côté et al., 2014; Gramzow et al., 2008; Gramzow & Willard, 2006; Leduc & Bouffard, 2017; Lee, 2021; 2022; Lopez et al., 1998; Praetorius et al., 2016; Robins & Beer, 2001; Willard & Gramzow, 2009). Ein Vorteil dieses Ansatzes gegenüber den algebraischen Differenzwerten ist, dass durch die statistische Kontrolle des Außenkriteriums in der Regression die resultierenden Residualwerte unabhängig von dem Außenkriterium sind (Humberg et al., 2018). Somit sind Residualwerte im Gegensatz zu algebraischen Differenzwerten nicht mit dem Außenkriterium konfundiert. Allerdings weist dieses Verfahren dafür andere bedeutsame Nachteile auf. So sind Residualwerte in der Regel noch stärker als algebraische Differenzwerte mit der Selbsteinschätzung konfundiert (Humberg et al., 2018). Daher kann unter Einsatz von Residualwerten noch schlechter als mit algebraischen Differenzwerten die Frage beantwortet werden, ob ein SE Bias über die Selbsteinschätzung per se hinaus einen Einfluss auf akademische Leistung ausübt. Die Verwendung von Residualwerten in der SE Bias Forschung ist also entgegen der Empfehlung ihrer Befürworter keine Lösung für das zentrale Problem von algebraischen Differenzwerten (siehe Humberg et al., 2018; 2019a; Irving & Meyer, 1999). Zudem führt die Verwendung von Residualwerten zu einer anderen, oft nicht theoriekonformen Interpretation des SE Bias. Ein SE Bias ist konzipiert als die Diskrepanz zwischen einer Selbsteinschätzung und einem damit in Bezug stehenden Außenkriterium, sodass ein positiver Wert einer Selbstüberschätzung und ein negativer Wert einer Selbstunterschätzung entspricht. Jedoch entspricht die Messung des SE Bias über Residualwerte nicht diesem theoretischen Konzept. Die Residuen einer Regression repräsentieren nicht das Ausmaß der Diskrepanz zwischen dem Kriterium und dem Prädiktor, sondern die Diskrepanz zwischen dem gemessenen Kriteriumswert und dem durch den

Prädiktor vorhergesagten Kriteriumswert (Seber & Lee, 2003, S. 38). Mit anderen Worten, die Residueen können als ein Maß dafür angesehen werden wie „untypisch“ der Kriteriumswert gegeben dem Prädiktorwert ist, aber nicht als ein Maß der Diskrepanz zwischen beiden. Anhand des folgenden Beispiels soll die praktische Bedeutung dieser Tatsache erläutert werden. Bonneville-Roussy et al. (2017) untersuchten SE Bias Effekte in schulischen Fähigkeiten auf akademische Leistung. Sie erfassten die selbsteingeschätzten schulischen Fähigkeiten mit Likert-Skalen-Items wie „Diese*r Schüler*in denkt, dass er oder sie gut in der Schule ist“ (S. 5; die berichteten Items sind in der dritten Person formuliert, obwohl die Autor*innen explizit angeben, dass die Schüler*innen sie für sich selbst beantworteten). Die tatsächliche schulische Kompetenz wurde mit einem Kompetenztest erfasst. Die Residuen einer Regression der Selbsteinschätzung auf den Kompetenztest wurden genutzt, um zwischen verschiedenen Gruppen (optimistische, realistische und pessimistische Schüler*innen) zu unterscheiden, welche dann auf Gruppenunterschiede in ihren akademischen Leistungen untersucht wurden. Allerdings ist dieses Vorgehen nicht geeignet, um tatsächlich zwischen optimistischen, realistischen und pessimistischen Personen zu unterscheiden. Falls die Schüler*innen in der Untersuchung sich im Mittel beispielsweise als hoch kompetent einschätzten, ihre tatsächliche Kompetenz aber nur durchschnittlich ausgeprägt war, würde eine Person mit einer hohen Selbsteinschätzung und einer durchschnittlichen tatsächlichen Kompetenz ein Residuum nahe null haben, da eine hohe Selbsteinschätzung in dieser Stichprobe eine durchschnittliche Kompetenz vorhersagt. Sie würde somit als akkurate*r Selbsteinschätzer*in klassifiziert, obwohl sie ihre Kompetenz höher einschätzt, als sie tatsächlich ist. Da die Residuen einer Regression einen Mittelwert von null haben (Urban & Mayerl, 2018, S. 196), schließt die Operationalisierung des SE Bias über Residualwerte von vornherein aus, dass eine Population sich im Mittel überschätzen oder unterschätzen kann. Aus diesem Grund sind Residualwerte im Gegensatz zu algebraischen Differenzwerten auch ungeeignet, um einen SE Bias zu quantifizieren (Humberg et al., 2018).

Andere Operationalisierungen des SE Bias. In einzelnen dem Zwei-Schritte-Ansatz zuzuordnenden Studien wurde der SE Bias statt über algebraische Differenzwerte oder Residualwerte mit anderen Verfahren berechnet. Martin & Debus (1998) verglichen Selbsteinschätzungen von Schüler*innen mit Einschätzungen der Schüler*innen durch ihre Lehrkräfte. Sie klassifizierten sowohl die Selbst- als auch die Lehrkraftseinschätzungen in die Kategorien „niedrig", „mittel" und „hoch". Weiter untersucht wurden lediglich Schüler*innen mit einer mittleren Lehrkraftseinschätzung. Diese wurden als „Unterschätzer*in" klassifiziert, wenn sie eine niedrige Selbsteinschätzung abgegeben hatten, als „akkurate*r Einschätzer*in", wenn sie eine mittlere Selbsteinschätzung abgegeben haben, oder als Überschätzer*in, wenn sie eine hohe Selbsteinschätzung abgegeben haben. Dieses Vorgehen folgt den Empfehlungen von Assor et al. (1990), welche im Gegensatz zu algebraischen Differenzwerten und Residualwerten in Studien zu SE Bias Effekten auf akademische Leistung ansonsten kaum Anwendung fanden. Allerdings ist auch dieses Vorgehen aus mehreren Gründen problematisch. Zum einen wird die Stichprobe systematisch auf lediglich jene Schüler*innen reduziert, die von ihren Lehrkräften als weder besonders hoch noch besonders niedrig kompetent eingeschätzt werden. Da sich aber natürlich auch diese Schüler*innen selbst über- oder unterschätzen können, schränkt dieses Vorgehen die Generalisierbarkeit der Ergebnisse auf die Gesamtpopulation ein. Zum anderen führt auch dieses Vorgehen zu einer potentiell stark ausgeprägten Konfundierung von SE Bias und Selbsteinschätzung. Denn nach diesem Verfahren haben sämtliche als Überschätzer*innen klassifizierte Schüler*innen eine hohe Selbsteinschätzung, sämtliche „akkurate Selbsteinschätzer*innen" haben eine mittlere Selbsteinschätzung und sämtliche „Unterschätzer*innen" haben eine niedrige Selbsteinschätzung. Demgegenüber kommt es zu keiner Konfundierung zwischen SE Bias und Außenkriterium, da das Außenkriterium für alle Schüler*innen in den Analysen konstant ist. Daher können mit diesem Vorgehen, ähnlich wie mit Residualwerten, SE Bias Effekte zwar von Effekten des Außenkriteriums getrennt werden, dafür ist die Konfundierung mit der Selbsteinschätzung aber umso problematischer.

Fazit. Sämtliche dem Zwei-Schritte-Ansatz zuzuordnende Verfahren weisen das Problem der Konfundierung von SE Bias Effekten mit Selbsteinschätzungseffekten und/oder Effekten des Außenkriteriums auf. Dies liegt nicht daran, dass die spezifischen in den einzelnen Verfahren genutzten Berechnungsvorschriften für den SE Bias fehlerhaft sind, sondern an einem grundlegenden Nachteil des Zwei-Schritte-Ansatzes, der das Auftreten dieses Problems quasi unumgänglich macht (Humberg et al., 2018). Berechnet man nämlich eine Variable (den SE Bias) aus zwei anderen Variablen, so wird die neu berechnete Variable in den allermeisten Fällen mit mindestens einer der beiden Variablen, aus denen sie berechnet wurde, konfundiert sein. Dies ist nicht zwangsläufig ein Problem, sofern der SE Bias lediglich quantifiziert werden soll, jedoch macht es den Zwei-Schritte-Ansatz ungeeignet um Effekte des SE Bias auf andere Variablen (wie auch generell Zusammenhänge des SE Bias mit anderen Variablen) zu untersuchen, da die besagte Konfundierung dann dazu führt, dass gefundene Effekte nicht eindeutig als Effekte des SE Bias interpretiert werden können.

1.2.3.2 Der Ein-Schritt-Ansatz

Beim Ein-Schritt-Ansatz entfällt der erste Schritt des Zwei-Schritte-Ansatzes, also die Berechnung des SE Bias. Stattdessen werden die Zusammenhänge der Selbsteinschätzung und des Außenkriteriums mit der interessierenden abhängigen Variablen (z.B. akademische Leistung) untersucht. Es mag kontraintuitiv erscheinen, SE Bias Effekte zu untersuchen, ohne den SE Bias überhaupt zu berechnen, doch ist es, wie in diesem Abschnitt erläutert wird, gerade dieses Vorgehen, dass die Konfundierung von SE Bias Effekten mit Effekten der Selbsteinschätzung und des Außenkriteriums verhindert. Zunächst sei an dem Beispiel eines positiven linearen SE Bias Effekts erläutert, wie man von Zusammenhängen der Selbsteinschätzung und des Außenkriteriums mit der abhängigen Variablen auf SE Bias Effekte schließen kann. Ein positiver linearer SE Bias Effekt liegt dann und nur dann vor, wenn eine größere Differenz zwischen Selbsteinschätzung und Außenkriterium (D = Selbsteinschätzung minus Außenkriterium) zu einer Zunahme in der abhängigen Variablen führt. Allerdings führt die Berechnung der Differenz D zu einer Konfundierung mit der Selbsteinschätzung (und der Kompetenz). Somit bedarf es einer Möglichkeit, die Effekte der Differenz D zu untersuchen, ohne diese Differenz zu berechnen. Nun gilt folgendes: Bei konstantem Außenkriterium führt eine Zunahme der Selbsteinschätzung zu einer Zunahme von D. Und: Bei konstanter Selbsteinschätzung führt eine Zunahme des Außenkriteriums zu einer Abnahme von D. Dies bedeuet widerum, ein positiver linearer SE Bias Effekt liegt dann und nur dann vor, wenn bei konstantem Außenkriterium eine Zunahme der Selbsteinschätzung zu einer Zunahme in der abhängigen Variablen führt und wenn bei konstanter Selbsteinschätzung eine Zunahme des Außenkriteriums zu einer Abnahme in der abhängigen Variablen führt (Edwards & Parry, 1993; Humberg et al., 2018). Denn nur dann führt jede Zunahme der Differenz D (und somit des SE Bias) zu einer Zunahme der abhängigen Variablen. Mit anderen Worten: Es wird betrachtet unter welchen *Bedingungen* eine Differenz größer wird. Dies ist der Fall, wenn der Minuend größer wird oder der Subtrahend kleiner wird. Führen diese Bedingungen nun zu einer Veränderung in der abhängigen Variablen, impliziert dies einen Effekt der Differenz, da *jede* Vergößerung der Differenz (egal ob verursacht durch eine Vergrößerung des Minuenden oder eine Verringerung des Subtrahenden) zu einer Veränderung der abhängigen Variablen führt (Humberg et al., 2018).

Die Begriffe „Ein-Schritt-Ansatz“ und „Zwei-Schritte-Ansatz“ wurden von Humberg et al. (2018) eingeführt, um die zahlreichen verschiedenen Methoden zur Untersuchung von SE Bias Effekten zu klassifizieren. Erste Verfahren, die dem Ein-Schritt-Ansatz zuzuordnen sind, wurden in der psychologischen Forschung hingegen schon weitaus früher entwickelt und verwendet (z.B. Asendorpf & Ostendorf, 1998; Edwards, 1993; 1994; Edwards & Parry, 1993; Irving & Meyer, 1999). Im Folgenden werden zwei grundlegende Verfahren innerhalb des Ein-Schritt-Ansatzes dargestellt: Die *condition-based-regression analysis* (*CRA*; Humberg et al., 2018) und die *response surface analysis* (*RSA*; Edwards 1993; 1994; 2002; Edwards & Parry, 1993; Humberg et al., 2019a; 2022).

Condition-based regression analysis. Bei der CRA handelt es sich, wie der Name nahelegt, um eine Regressionsanalyse. In einer linearen Regression wird die interessierende abhängige Variable (z.B. akademische Leistung; A) durch zwei Prädiktoren vorhergesagt: Die Selbsteinschätzung (S) und das Außenkriterium (K). Somit ergibt sich die in Gleichung 1) dargestellte Regressionsgleichung.

1) $A = a + b_1S + b_2K + \varepsilon$

Aus dem Ergebnis dieser Regression, speziell aus den Werten für die beiden Regressionsgewichte b_1 und b_2, können wie folgt Schlussfolgerungen über etwaige SE Bias Effekte abgeleitet werden. Ein linearer SE Bias Effekt liegt vor, wenn eine größere Differenz (S – K) zu größeren Werten (positiver Effekt) oder kleineren Werten (negativer Effekt) in A führt. Werte für die Differenz (S – K) werden wiederum größer, wenn bei konstantem Wert K die Werte von S zunehmen und auch, wenn bei konstantem Wert S die Werte von K abnehmen. Die Regressionsgewichte in einer multiplen Regression sind zu interpretieren als das Ausmaß, in dem sich der vorhergesagte Wert A verändert, wenn der Wert von einem der Prädiktoren erhöht wird, während die anderen Prädiktoren konstant bleiben (Urban & Mayerl, 2018, S. 79-80). Somit liegt ein positiver linearer SE Bias dann und nur dann vor, wenn b_1 in Regressionsgleichung 1) positiv und b_2 negativ ist. Denn nur in diesem Fall geht eine Vergrößerung des Wertes von S unter Konstanthaltung von K als auch eine Verkleinerung von K unter Konstanthaltung von S (und somit eine größere Differenz S – K) mit einer Vergrößerung des Wertes von A einher. Ebenso liegt umgekehrt ein negativer linearer SE Bias Effekt dann und nur dann vor, wenn b_1 negativ und b_2 positiv ist. Zusammenfassend bedeutet dies, dass ein linearer SE Bias Effekt dann und nur dann vorliegt, wenn b_1 und b_2 unterschiedliche Vorzeichen haben (Humber et al., 2018). Das grundlegende Prinzip der CRA ist somit eine multiple lineare Regression. Das Neue an diesem Ansatz ist, dass die Autor*innen einen Signifikanztest für den Befund entwickelt haben, dass die Vorzeichen von b_1 und b_2 unterschiedlich sind. Das Vorgehen bei der CRA ist somit das Folgende: 1) Es wird eine multiple lineare Regression der abhängigen Variablen auf die Selbsteinschätzung und das Außenkriterium berechnet. 2) Es wird betrachtet, ob sich die Vorzeichen der Regressionsgewichte der Selbsteinschätzung und des Außenkriteriums deskriptiv unterscheiden. 3) Falls sich die Vorzeichen deskriptiv unterscheiden, wird ein Signifikanztest gerechnet, der diesen Befund inferenzstatistisch absichert. Nur falls dieser Test ein statistisch signifikantes Ergebnis erbringt, liegt ein statistisch signifikanter SE Bias Effekt vor. Ob dieser Effekt positiv oder negativ ist, kann dann anhand der Vorzeichen der Selbsteinschätzung und des Außenkriteriums beurteilt werden. Gegenüber allen Verfahren des Zwei-Schritte-Ansatzes weist die CRA einen bedeutsamen Vorteil auf: Der Effekt des SE Bias ist nicht mit den Effekten der Selbsteinschätzung und/oder des Außenkriteriums konfundiert (Humberg et al., 2018). Dies ist darauf zurückzuführen, dass entsprechend des Ein-Schritt-Ansatzes kein SE Bias Wert berechnet wird, sondern Schlussfolgerungen über einen etwaigen SE Bias Effekt ausschließlich auf Basis der Betrachtung der Effekte der Selbsteinschätzung und des Außenkriteriums gezogen werden. Aus diesem Grund stellt die CRA eine wichtige Neuerung bei der Untersuchung von SE Bias Effekten dar. Allerdings ist sie auch mit der Limitation behaftet, dass lediglich lineare SE Bias Effekte untersucht werden können. In vielen Bereichen der SE Bias Forschung, inklusive der Forschung zu SE Bias Effekten auf akademische Leistung, gehen Autor*innen aber auch von der Möglichkeit nonlinearer Effekte aus. Wie in vorigen Abschnitten dargestellt, wird in der Literatur unter anderem die Hypothese vertreten, dass ein gewisses Maß an Selbstüberschätzung förderlich für akademische Leistung ist, während ein noch größeres Maß an Selbstüberschätzung sich negativ auswirkt (z.B. Baumeister et al., 1989; Helmke, 1998). Ebenso weisen Ergebnisse aus Studien zum selbstregulierten Lernen darauf hin, dass akkurate Selbsteinschätzungen am förderlichsten für akademische Leistung sein könnten (z.B. Dunlosky & Rawson, 2012; Hacker & Bol, 2019). Da in derartigen Hypothesen nonlineare Zusammenhängen zwischen einem SE Bias und akademischer Leistung angenommen werden, können sie mit der CRA nicht überprüft werden. Eine Alternative, die auch die Überprüfung nonlinearer Effekte erlaubt, ist die response surface analysis (RSA).

Response surface analysis. Das Grundprinzip der RSA (Box & Draper, 1987; Edwards & Parry, 1993; Khuri & Cornell, 1987) ist dem der CRA sehr ähnlich. Zusammenfassend kann die RSA als der allgemeinere Ansatz bezeichnet werden und die CRA als ein Spezialfall der RSA. Während die CRA auf einer linearen Regression basiert, basiert die RSA allgemein auf einer polynomialen Regression. Welche polynomiale Regression, beispielsweise eine lineare, quadratische oder kubische, verwendet wird, kann von den Anwender*innen anhand inhaltlich-theoretischer Überlegungen über die möglichen zugrundeliegenden Effekte entschieden werden (Humberg et al., 2022). Letztlich handelt es sich bei der CRA also um einen Spezialfall der RSA für lineare Effekte wobei der neuartige Signifikanztest der CRA nur im Falle einer linearen Regression angewendet werden kann. Im Folgenden wird die RSA zunächst am relativ simplen und zudem in der Literatur am häufigsten verwendeten Beispiel einer polynomialen Regression zweiter Ordnung erläutert (z.B. Edwards, 2007; Humberg et al., 2019b; Nestler et al., 2015; 2019; Schönbrodt, 2016; Schönbrodt et al., 2017). Bei Verwendung einer polynomialen Regression zweiter Ordnung in der RSA wird die abhängige Variable durch fünf Prädiktorterme vorhergesagt: 1) einen linearen Term der Selbsteinschätzung (S), 2) einen linearen Term des Außenkriteriums (K), 3) einen quadratischen Term der Selbsteinschätzung (S^2), 4) eine Interaktion zwischen der Selbsteinschätzung und dem Außenkriterium (SK) und 5) einen quadratischen Term des Außenkriteriums (K^2). Somit ergibt sich die in Gleichung 2) dargestellte Regressionsgleichung für die abhängige Variable (A).

2) $A = a + b_1S + b_2K + b_3S^2 + b_4SK + b_5K^2 + \varepsilon$

Das Ergebnis einer solchen Regression kann als Regressionsoberfläche (daher der Name response surface analysis) dargestellt werden, auf der für jede Kombination von Werten für S und K ein bestimmter Wert für A vorhergesagt wird (siehe Abbildung 1).

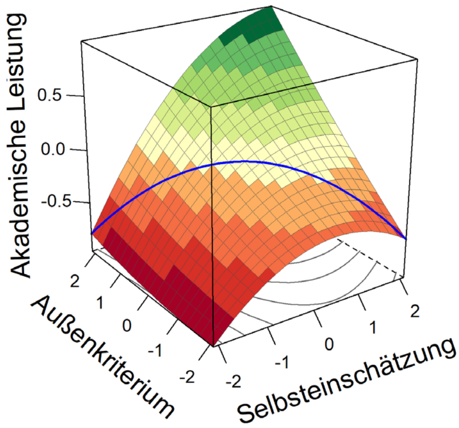


Abbildung 1. Beispielhaftes Ergebnis einer response surface analysis unter Verwendung einer polynomialen Regression zweiter Ordnung.

Ähnlich wie auch in der CRA können in der RSA Schlussfolgerungen über mögliche SE Bias Effekte anhand der Betrachtung der Regressionsgewichte b_1 bis b_5 gezogen werden (siehe unten). Bei der RSA werden darüber hinaus häufig bestimmte geometrische Linien auf der Regressionsoberfläche betrachtet, um SE Bias Effekte zu untersuchen (z.B. Edwards & Parry, 1993; Humberg et al., 2019b; Humberg et al., 2022; Shanock et al., 2010). Die prominenteste und wichtigste dieser Linien ist die sogenannte line of incongrunece (LOIC). Die LOIC ist die Schnittmenge der Fläche, die durch die Gleichung S = -K definiert ist und der Regressionsoberfläche. Mit anderen Worten, sie beinhaltet alle Punkte auf der Regressionsoberfläche für die gilt: S = -K. Sie ist in den Graphiken in Abbildung 1 in blau dargestellt. Die LOIC ist für die explorative graphische Untersuchung von SE Bias Effekten interessant, da sie als ein Maß des SE Bias verstanden werden kann. Der Wert des SE Bias ist groß, wenn S groß und K klein ist, also in Abbildung 1 am rechten Ende der LOIC, und er ist klein, wenn S klein und K groß ist, also am linken Ende der LOIC. Der SE Bias nimmt auf der LOIC somit von links nach rechts zu. Beispielsweise erkennt man bei der LOIC in Abbildung 1 eine Krümmung in etwa entsprechend eines umgekehrt u-förmigen Zusammenhangs. Somit scheint die akademische Leistung in diesem fiktiven Beispiel höher ausgeprägt zu sein, wenn Selbsteinschätzung und Außenkriterium einander in etwa entsprechen (SE Bias nahe null; Mitte der LOIC). Auf diese Weise können unter Betrachtung der Krümmung der LOIC explorativ Annahmen über nonlineare SE Bias Effekte abgeleitet werden (Edwards & Parry, 1993; Humberg et al., 2022; Shanock et al., 2010). Ebenso kann die Steigung wie auch die Krümmung der LOIC berechnet werden, um die entsprechenden Zusammenhänge zu quantifizieren (Edwards & Parry, 1993; Shanock et al., 2010). Dieses Vorgehen bringt allerdings auch Nachteile mit sich, da sehr viele Modellparameter gleichzeitig berücksichtigt und interpretiert werden müssen (unter anderem die Steigung und Krümmung der LOIC, aber auch weitere aus Platzgründen hier nicht näher thematisierte Parameter), um zutreffende Schlussfolgerungen über SE Bias Effekte ziehen zu können (Humberg et al., 2022). Da eine solche parallele Interpretation mehrerer Parameter fehleranfällig ist, wird insbesondere für den Fall mehrerer konkurrierender Hypothesen, wie sie auch in der vorliegenden Arbeit vertreten sind, ein alternatives Vorgehen empfohlen (Humberg et al. 2022). Diese Alternative besteht darin, verschiedene Modelle aufzustellen, die unterschiedlichen inhaltlichen Annahmen über SE Bias Effekte entsprechen und diese in Modellvergleichen gegenüber zu stellen (Humberg et al., 2019a; 2022). Ein weiterer Vorteil dieses Vorgehens ist, dass unterschiedliche inhaltliche Annahmen über SE Bias Effekte direkt verglichen werden können. Die verschiedenen Modelle werden dabei durch Restriktionen der Regressionsgewichte $b1$ bis $b5$ aufgestellt. Dies soll im Folgenden beispielhaft erläutert werden. Eine häufige Annahme der SE Bias Forschung ist, dass ein SE Bias eigener Kompetenzen in einer bestimmten Domäne sich positiv auf die akademische Leistung in selbiger Domäne auswirkt (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côté et al., 2014; Lee, 2021; 2022; Leduc & Bouffard, 2017; Martin & Debus, 1998; Praetorius et al., 2016; Wright, 2000). Wie im Abschnitt zur CRA erläutert, liegt ein solcher positiver linearer SE Bias Effekt in einer linearen Regression genau dann vor, wenn das Regressionsgewicht der Selbsteinschätzung $b1$ positiv und das Regressionsgewicht des Außenkriteriums $b2$ negativ ist. In einer polynomialen Regression zweiter Ordnung müssen zudem die Regressionsgewichte der nonlinearen Terme $b3$ bis $b5$ auf null gesetzt werden, damit lediglich lineare Effekte modelliert werden. Somit ergibt sich ein Modell für einen positiven linearen SE Bias Effekt durch folgende Regressionsgleichung:

3) $A = a + b_1S + b_2K + b_3S^2 + b_4SK + b_5K^2 + \varepsilon \mid b_1 > 0; b_2 < 0; b_3 = b_4 = b_5 = 0$

Diese Restriktionen der Werte der Regressionsgewichte spiegeln die Annahme wieder, dass es lediglich einen positiven linearen SE Bias Effekt auf akademische Leistung gibt ($b_1 > 0$; $b_2 < 0$), aber keine nonlinearen Effekte ($b_3 = b_4 = b_5 = 0$). Ebenso können andere Annahmen über SE Bias Effekte, etwa die Annahme, dass ein SE Bias von null optimal für akademische Leistung sein sollte, durch andere Restriktionen der Regressionsgewichte in Gleichung 2) modelliert werden (Humberg et al., 2019a). Die entsprechenden Modelle können dann auf ihre Güte geprüft und verglichen werden, um das oder die beste(n) Modell(e) zu identifizieren.

Die bisherigen Erläuterungen zur RSA basieren auf der Annahme einer polynomialen Regression zweiter Ordnung wie in Gleichung 2) dargestellt. Für viele Fragestellungen in der SE Bias Forschung ist eine solche Regressionsgleichung ausreichend, um die entsprechenden Effekte zu modellieren. Falls aber inhaltlich-theoretische Annahmen vorliegen, dass es SE Bias Effekte geben könnte, die über polynomiale Effekte zweiter Ordnung (also lineare Effekte, quadratische Effekte und Interaktionseffekte) hinausgehen (beispielsweise kubische Effekte), können diese ebenfalls unter Verwendung einer polynomialen Regression höherer Ordnung modelliert werden. Ein Beispiel dafür wäre etwa die Annahme, dass sich sowohl Selbstüberschätzung als auch Selbstunterschätzung negativ auswirken, aber der negative

Effekt einer Selbstunterschätzung größer ist als der einer Selbstüberschätzung. Derartige nonlineare asymmetrische SE Bias Effekte können nicht durch eine polynomiale Regression zweiter Ordnung modelliert und geprüft werden. Stattdessen sind dafür polynomiale Regressionen dritter Ordnung notwendig, welche zusätzlich zu den Prädiktoren in Gleichung 2) die Prädiktoren S^3, S^2K, SK^2 und K^3 beinhalten (Humberg et al., 2022). Prinzipiell erlaubt die RSA somit jegliche SE Bias Effekte zu testen, die sich durch eine polynomiale Funktion ausdrücken lassen. Aufgabe der Forschenden ist es also, inhaltlich-theoretische Annahmen in ein bestimmtes polynomiales Regressionsmodell zu übersetzen, indem die entsprechenden Annahmen durch Parameterrestriktionen dargestellt werden können. Da in der Literatur zu SE Bias Effekten auf akademische Leistung nach meinem Wissen nicht von derartigen asymmetrischen Effekten berichtet wird, liegt der Fokus in dieser Arbeit auf der RSA mit einer polynomialen Regression zweiter Ordnung. Sowohl die Annahme linearer SE Bias Effekte (z.B. Bonneville-Roussy et al., 2017; Lee, 2021) als auch die Annahme einer optimal margin of illusion (z.B. Baumeister, 1989, Helmke 1998) und die Annahme, dass ein SE Bias von null optimal ist (z.B. Dunlsoky & Rawson 2012; Hacker & Bol, 2019), lassen sich durch eine solche Regression modellieren und prüfen (Humberg et al., 2019a; 2022).

Die unterschiedlichen Modelle können dann empirisch mit Hinblick auf ihre Modellgüte verglichen werden. Dazu können gängige Verfahren zum empirischen Modellvergleich genutzt werden, wie Chi-Quadrat-Differenz-Tests (Pavlov et al., 2020) und/oder informationstheoretische Kriterien wie das Akaike-Informationskriterium (AIC; Akaike, 1973; Burnham & Anderson, 2002; Humberg et al., 2019a; 2022; Kuha, 2004; Wagenmakers & Farrell, 2004) und das Bayessche Informationskriterium (BIC; Humberg et al., 2022; Kuha, 2004; Schwarz, 1978). Chi-Quadrat-Differenztests weisen jedoch den Nachteil auf, dass sie nur zum Vergleich genesteter Modelle geeignet sind (Pavlov et al, 2020). Allerdings sind viele der gängigen Hypothesen über SE Bias Effekte nicht ineinander genestet (siehe Humberg et al., 2019a für eine Übersicht). Unter anderem aus diesem Grund sind informationstheoretische Kriterien oft nützlicher, um unterschiedliche Hypothesen über SE Bias Effekte vergleichend zu evaluieren. In diesen informationstheoretischen Ansätzen, wird die Güte jedes Modells durch ein oder mehrere Kriterien (z.B. AIC und/oder BIC) quantifiziert. Die Berechnungsvorschrift der Kriterien besteht dabei aus zwei Komponenten: Einer likelihood-Funktion, die einen umso kleineren (negativen) Wert annimmt, je besser die Passung des Models zu den Daten ist, sowie einem Strafterm, der umso größer ausfällt, je mehr Parameter in dem Modell geschätzt werden und der auf die likelihood-Funktion addiert wird (Akaike, 1973; Schwarz, 1978). Kleinere Werte der Informationskriterien sprechen dann für bessere Modelle. Modelle sind nach dieser Logik also umso besser, je besser sie zu den Daten passen und je sparsamer sie sind, da weniger Parameter geschätzt werden. Ein zweiter Vorteil des informationstheoretischen Ansatzes ist, dass er erlaubt, die sehr einfach und intuitiv zu interpretierenden Akaike-Gewichte zu berechnen. Akaike-Gewichte basieren auf einer korrigierten Version des AIC (AICc), die entwickelt wurde, um das Problem der Überanpassung (overfitting) des klassischen AIC bei kleinen Stichproben zu lösen (Burnham & Anderson, 2002; Cavanaugh, 1997, Hurvich and Tsai, 1989). Akaike-Gewichte können interpretiert werden als die a priori Wahrscheinlichkeit (genauer likelihood) eines Modells, nach Logik des AICc das beste Modell gegeben der Daten und konkurrierenden Modelle zu sein (Burnham & Anderson, 2001; Wagenmakers & Farrell, 2004). Die Summe der Akaike-Gewichte aller Modelle in einem Modellset beträgt somit stets 1, beziehungsweise 100%. Beispielsweise bedeutet ein Akaike-Gewicht von $w = .87$, dass das entsprechende Modell mit einer Wahrscheinlichkeit von 87% das beste Modell gegeben der Daten und konkurrierenden Modelle darstellt. Akaike-Gewichte sind somit nicht nur einfach zu interpretieren, sondern erlauben neben der reinen Modellselektion zudem eine Quantifizierung der Unsicherheit bei der Modellselektion. So bleibt etwa im oben genannten Beispiel eine Restwahrscheinlichkeit von 13%, dass das entsprechende Modell nicht das beste Modell gegeben der Daten und konkurrierenden Modelle ist. Dies wiederum erlaubt analog zum Nullhypothesentesten, gewisse Kriterien festzulegen, um ein einzelnes Modell als das beste Modell anzunehmen. Beispielsweise kann analog zum Alpha-Niveau von .05 beim Nullhypothesentesten festgelegt werden, dass ein Modell nur dann als das alleinige beste Modell angenommen wird, wenn es ein Akaike-Gewicht größer .95 aufweist (Humberg et al., 2019a).

Spline Regression. Der Vollständigkeit halber sei kurz eine dritte Möglichkeit umrissen, innerhalb des Ein-Schritt-Ansatzes SE Bias Effekte zu testen, die spline regressions (Edwards & Parry, 2018). Durch spline regressions können bestimmte SE Bias Effekte überprüft werden, die sich nicht durch eine polynomiale Regression in der RSA abbilden lassen. Speziell können in diesem Verfahren nicht monotone Zusammenhänge zwischen einem SE Bias und anderen Variablen geprüft werden. Mit anderen Worten, es können Zusammenhänge zwischen einem SE Bias und anderen Merkmalen geprüft werden, bei denen sich die Steigung der Regressionsoberfläche für bestimmte Werte plötzlich verändert. Beispielsweise könnte so die Annahme geprüft werden, dass bei negativen SE Bias Werten ein positiver linearer Zusammenhang mit akademischer Leistung besteht, bei positiven SE Bias Werten hingegen ein negativer linearer Zusammenhang. Ebenso ist es mit diesem Ansatz möglich, Modelle aufzustellen, in denen es mehrere Wertebereiche gibt, für die sich die Steigung der Regressionsoberfläche plötzlich verändert. Beispielsweise kann so die Annahme modelliert werden, dass die abhängige Variable ein Plateau erreicht, wenn der SE Bias innerhalb eines gewissen Wertebereichs, beispielsweise [-1; 1], liegt. Die entsprechenden Modelle können dann wie auch bei der RSA durch Modellvergleiche beurteilt werden. Spline regressions sind nützlich zur Untersuchung verschiedener in der psychologischen Literatur diskutierter SE Bias Effekte (Edwards & Parry, 2018). Im Kontext von Effekten auf akademische Leistung ist ihre Bedeutung allerdings geringer, da die angenommenen Effekte hier auch durch polynomiale Regressionen abgebildet werden können (Humberg et al., 2019; siehe auch die empirischen Beiträge der vorliegenden Arbeit).

Fazit: Der zentrale Vorteil des Ein-Schritt-Ansatzes gegenüber dem Zwei-Schritte-Ansatz ist die Vermeidung einer Konfundierung von SE Bias Effekte mit Effekten der Selbsteinschätzung und des Außenkriteriums. Aus diesem Grund sollten Verfahren, die dem Zwei-Schritte-Ansatz folgen, nicht mehr zur Untersuchung von SE Bias Effekten verwendet werden (Humberg et al., 2018). Der Problematik des Zwei-Schritte-Ansatzes wird in unterschiedlichen psychologischen Disziplinen unterschiedlich stark Rechnung getragen. So wurden methodische Arbeiten zur RSA (z.B. Box & Draper, 1987; Khuri & Cornell, 1987) etwa innerhalb der Arbeits- und Organisationspsychologie, speziell in der Forschung zur Bedeutung der Passung zwischen Organisationsmerkmalen und Mitarbeitendenmerkmalen durch umfassende methodisch-anwendungsorientierte Arbeiten etabliert (Edwards, 1993; 1994; 2002; Edwards & Parry 1993) und finden daher auch aktuell vielfach Anwendung (z.B. Bäcklander, & Richter, 2022; Bernerth et al., 2022; Breetzke & Wild, 2022; Chênevert et al., 2022; de Vries et al., 2022; Leichner et al., 2022; Li & Xie, 2022; McLarty et al., 2022). In der pädagogischen Psychologie und Persönlichkeitspsychologie fehlten hingegen bis vor kurzem vergleichbare Arbeiten, in denen die Bedeutung und Anwendbarkeit des Ein-Schritt-Ansatzes theoretisch erläutert und anhand empirischer Untersuchungen belegt wurde. Erst kürzlich wurde diese Forschungslücke durch Arbeiten der Forschungsruppe um Sarah Humberg und Mitja Back geschlossen (Humberg et al., 2018, 2019a; 2019b; 2022; Humberg & Grund, 2022). Seitdem verwenden unter Bezug auf die oben genannten Arbeiten auch Forschende in der pädagogischen und Persönlichkeitspsychologie zunehmend Verfahren innerhalb des Ein-Schritt-Ansatzes, um SE Bias Effekte wie auch allgemein Effekte der (In-)kongruenz zwischen zwei Merkmalen zu untersuchen (z.B. Elsaadawy & Carlson, 2022; Förster et al., 2022; Ilmarinen et al., 2019; Joel et al., 2022; Mota et al., 2020; Rivers et al., 2022; Tamm et al., 2021; Wang et al., 2022; Wright & Jackson, 2023; Zhou et al., 2023). Zuvor wurden hingegen weitaus häufiger Verfahren verwendet, die dem Zwei-Schritte-Ansatz zuzuordnen sind. Speziell gibt es nach meinem Wissen, außer den empirischen Beiträgen in der vorliegenden Arbeit, noch keine publizierten Studien zu SE Bias Effekten auf akademische Leistung, die dem Ein-Schritt-Ansatz zuzuordnen sind. Aus diesem Grund können aus empirischen Arbeiten bisher nur sehr bedingt Schlussfolgerungen über SE Bias Effekte auf akademische Leistung gezogen werden.

1.3 Fragestellungen und Hypothesen

Aus zahlreichen Studien ist bekannt, dass sich hohe Selbsteinschätzungen eigener Kompetenzen positiv auf akademische Leistung auswirken (für Metaanalysen siehe Talsma et al., 2018; Valentine et al., 2004; Wu et al., 2021). Unter anderem zeigt sich, dass ein höheres Fähigkeitsselbstkonzept mit einer positiveren Entwicklung zukünftiger Noten innerhalb derselben Domäne einhergeht (Valentine et al., 2004; Wu et al., 2021). Dieser Befund lässt sich theoretisch auf Basis des Erwartungs-Wert-Modells (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020) erklären, da ein höheres Fähigkeitsselbstkonzept, vermittelt über höhere subjektive Werte und Erfolgserwartung, zu besseren akademischen Leistungen führen sollte. Daher stelle ich folgende inhaltliche Hypothesen auf:

Hypothese 1: Ein höheres Fähigkeitsselbstkonzept innerhalb einer Domäne (z.B. Mathematik oder Deutsch) führt zu besseren zukünftigen Noten, selbst wenn die objektiv erfasste Kompetenz und der Ausgangswert der Noten kontrolliert wird.

Hypothese 2a: Die intrinsischen Werte in einer Domäne mediieren den Effekt des Fähigkeitsselbstkonzepts auf Noten in selbiger Domäne.

Hypothese 2b: Die Nützlichkeitswerte in einer Domäne mediieren den Effekt des Fähigkeitsselbstkonzepts auf Noten in selbiger Domäne.

Hypothese 2c: Die Wichtigkeitswerte in einer Domäne mediieren den Effekt des Fähigkeitsselbstkonzepts auf Noten in selbiger Domäne.

Hypothese 2d: Die Erfolgserwartung in einer Domäne mediiert den Effekt des Fähigkeitsselbstkonzepts auf Noten in selbiger Domäne.

Im Gegensatz zu Selbsteinschätzungseffekten, welche gut empirisch wie theoretisch abgesichert sind, gibt es nach wie vor keine methodisch und theoretisch fundierten Arbeiten zu potentiellen SE Bias Effekten auf akademische Leistung. Aus diesem Grund stelle ich keine spezifischen inhaltlichen Hypothesen zu SE Bias Effekten in der vorliegenden Arbeit auf. Stattdessen soll die folgende Fragestellung untersucht werden.

Fragestellung 1: Gibt es Effekte eines Bias im Fähigkeitsselbstkonzept auf Noten und falls ja, um was für einen Effekt handelt es sich?

Speziell sollen zur Beantwortung von Fragestellung 1 die folgenden in der Literatur aufgestellten beziehungsweise aus der Literatur abgeleiteten Hypothesen vergleichend evaluiert werden: 1) es gibt einen positiven linearen SE Bias Effekt auf Schulnoten, 2) es gibt einen negativen linearen SE Bias Effekt af Schulnoten, 3) es gibt einen nonlinearen, umgekehrt u-förmigen SE Bias Effekt auf Schulnoten, sodass ein SE Bias von null zu den besten Noten führt, während sowohl ein positiver SE Bias (Selbstüberschätzung) als auch ein negativer SE Bias (Selbstunterschätzung) zu schlechteren Noten führt, 4) es gibt einen nichtlinearen *optimal margin* SE Bias Effekt auf Schulnoten, in dem Sinne, dass ein moderat positiver SE Bias zu den besten Noten führt, während sowohl ein geringerer als auch ein höherer SE Bias zu schlechteren Noten führen, 5) es gibt keine SE Bias Effekte auf Schulnoten.

Da zahlreiche Autor*innen davon ausgehen, dass SE Bias Effekte auf akademische Leistung über motivationale Variablen vermittelt sein könnten (z.B. Bonneville-Roussy et al., 2017; Lee, 2021), soll zudem überprüft werden ob etwaige SE Bias Effekte analog zu Effekten des Fähigkeitsselbstkonzepts von subjektiven Werten und Erfolgserwartung mediiert werden. Speziell sollen somit folgende sekundäre Fragestellungen untersucht werden: Falls SE Bias Effekte auf Schulnoten gefunden werden, werden diese mediiert durch

Fragestellung 2a: intrinsische Werte,

Fragestellung 2b: Wichtigkeitswerte,

Fragestellung 2c: Nützlichkeitswerte,

Fragestellung 2d: Erfolgserwartung?

Literaturverzeichnis I

Arens, A. K., & Niepel, C. (2023). Formation of academic self-concept and intrinsic value within and across three domains: Extending the reciprocal internal/external frame of reference model. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.1257

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966. https:// doi.org/10.1037//0022-3514.74.4.955

Assor, A., Tzelgov, J., Thien, R., Ilardi, B.C., & Connell, J.P. (1990). Assessing the correlates of over- and underrating of academic competence: a conceptual clarification and a methodological proposal. *Child Development, 61*, 2085-2097. https://doi.org/10.2307/1130862

Bäcklander, G., & Richter, A. (2022). Relationships of task–environment fit with office workers' concentration and team functioning in activity-based working environments. *Environment and Behavior*, *54*(6), 971–1004. https://doi.org/10.1177/00139165221115181

Bakadorova, O., & Raufelder, D. (2020). The relationship of school self-concept, goal orientations and achievement during adolescence. *Self and Identity, 19*(2), 235–249. https://doi.org/10.1080/15298868.2019.1581082

Baumeister, R. F. (1989). The optimal margin of illusion. *Journal of Social and Clinical Psychology, 8*(2), 176-189. https://doi.org/10.1521/jscp.1989.8.2.176

Benden, D. K., & Lauermann, F. (2022). Students' motivational trajectories and academic success in math-intensive study programs: Why short-term motivational assessments matter. *Journal of Educational Psychology*, *114*(5), 1062–1085. https://doi.org/10.1037/edu0000708.supp (Supplemental)

Bernerth, J. B., Carter, M. Z., & Cole, M. S. (2022). The (in)congruence effect of leaders' narcissism identity and reputation on performance: A socioanalytic multistakeholder perspective. *Journal of Applied Psychology*, *107*(10), 1725–1742. https://doi.org/10.1037/apl0000974

Block, J., & Colvin, C. R. (1994). Positive illusions and well-being revisited: Separating fiction from fact. *Psychological Bulletin*, *116*(1), 28. https://doi.org/10.1037/0033-2909.116.1.28

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research, 105*(5), 336–352. https://doi.org/10.1080/00220671.2011.627401

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*(3), 139-153. https://doi.org/10.1207/s15326985ep3403_1

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*, 1-40. https://doi.org/10.1023/A:1021302408382

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. https://doi.org/10.1016/j.jsp.2017.02.002

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*, 221-229. https://doi.org/10.1016/j.ijer.2011.08.003

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces.* Wiley.

Breetzke, J., & Wild, E.-M. (2022). Social connections at work and mental health during the first wave of the COVID-19 pandemic: Evidence from employees in Germany. *PLoS ONE*, *17*(6). https://doi.org/10.1371/journal.pone.0264602

Breitwieser, J., & Brod, G. (2022). The interplay of motivation and volitional control in predicting the achievement of learning goals: An intraindividual perspective. *Journal of Educational Psychology*, *114*(5), 1048–1061. https://doi.org/10.1037/edu0000738.supp (Supplemental)

Burnham, K. P., & Anderson, D. R. (2001). Kullback–Leibler information as a basis for strong inference in ecological studies. *Wildlife Research, 28*, 111-119. https://doi.org/10.1071/WR99107

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer. https://doi.org/10.1007/b97636

Cai, D., Viljaranta, J., & Georgiou, G. K. (2018). Direct and indirect effects of self-concept of ability on math skills. *Learning and Individual Differences*, *61*, 51–58. https://doi.org/10.1016/j.lindif.2017.11.009

Cavanaugh, J. E. (1997). Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters, 31*(2), 201-208. https://doi.org/10.1016/s0167-7152(96)00128-9

Chênevert, D., Hill, K., & Kilroy, S. (2022). Employees perceptions of non-monetary recognition practice and turnover: Does recognition source alignment and contrast matter? *Human Resource Management Journal*, *32*(1), 40–57. https://doi.org/10.1111/1748-8583.12354

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: A longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*(10), 1384-1401. https://doi.org/10.1177/0146167216662866

Cogliano, M., Bernacki, M. L., & Kardash, C. M. (2021). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. *Journal of Educational Psychology*, *113*(7), 1421–1440. https://doi.org/10.1037/edu0000624.supp (Supplemental)

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116*, 3–20. https://doi.org/10.1037/0033-2909.116.1.3

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68,* 1152-1162. https://doi.org/ 10.1037//0022-3514.68.6.1152

Connell, J. P., & Illardi, B. C. (1987). Self-system concomitants of discrepancies between children's and teachers' evaluations of academic competence. *Child Development, 58*, 1297–1307. https://doi.org/10.2307/1130622

Côté, S., Bouffard, T., & Vezeau, C. (2014). The mediating effect of self-evaluation bias of competence on the relationship between parental emotional support and children's academic functioning. *British Journal of Educational Psychology, 84*(3), 415–434. https://doi.org10.1111/bjep.12045

de Vries, A., Broks, V. M. A., Bloemers, W., Kuntze, J., & de Vries, R. E. (2022). Self-, other-, and meta-perceptions of personality: Relations with burnout symptoms and eudaimonic workplace well-being. *PLoS ONE*, *17*(7). https://doi.org/ 10.1371/journal.pone.0272095

Dong, L., & Kang, Y. (2022). Cultural differences in mindset beliefs regarding mathematics learning. *Current Opinion in Behavioral Sciences*, *46*. https://doi.org/10.1016/j.cobeha.2022.101159

Dufner, M., Denissen, J. J. A., van Zalk, M., Matthes, B., Meeus, W. H. J., van Aken, M. A. G., & Sedikides, C. (2012). Positive intelligence illusions: On the relation between intellectual self-enhancement and psychological adjustment. *Journal of Personality, 80,* 537–572. http://dx.doi.org/10.1111/j.1467-6494.2011.00742.x

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https//doi.org10.1016/j.learninstruc.2011.08.003

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250. https://doi.org/10.1016/j.ijer.2011.08.005

Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61.* https://doi.org/10.1016/j.cedpsych.2020.101859.

Eccles, J. S., & Wigfield, A. (2023). Expectancy-value theory to situated expectancy-value theory: Reflections on the legacy of 40+ years of working together. *Motivation Science*, *9*(1), 1–12. https://doi.org/10.1037/mot0000275

Edwards, J. R. (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. *Personnel Psychology*, *46*(3), 641–665. https://doi.org/10.1111/j.1744-6570.1993.tb00889.x

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 51–100. https://doi.org/10.1006/obhd.1994.1029

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Hrsg.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (S. 350–400). Jossey-Bass.

Edwards, J. R. (2007). Polynomial regression and response surface methodology. In C. Ostroff & T. A. Judge (Hrsg.), *Perspectives on organizational fit* (S. 361–372). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36,* 1577–1613. http://dx.doi.org/10.2307/256822

Edwards, J. R., & Parry, M. E. (2018). On the Use of Spline Regression in the Study of Congruence in Organizational Research. *Organizational Research Methods, 21*, 68-110. https://do.org/10.1177/1094428117715067

Elsaadawy, N., & Carlson, E. N. (2022). Do you make a better or worse impression than you think? *Journal of Personality and Social Psychology*, *123*(6), 1407–1420. https://doi.org/10.1037/pspp0000434

Esnaola, I., Sesé, A., Antonio, A. I., & Azpiazu, L. (2020). The development of multiple self-concept dimensions during adolescence. *Journal of Research on Adolescence*, *30*(Suppl 1), 100–114. https://doi.org/10.1111/jora.12451

Fadda, D., Scalas, L. F., Morin, A. J. S., Marsh, H. W., & Gaspard, H. (2020). Value beliefs about math: A bifactor-ESEM representation. *European Journal of Psychological Assessment*, *36*(2), 259–268. https://doi.org/10.1027/1015-5759/a000513 (Supplemental)

Felson, R. B. (1984). The effect of self-appraisals of ability on academic performance. *Journal of Personality and Social Psychology, 47,* 944-952. https://doi.org/10.1037/0022-3514.47.5.944

Förster, N., Humberg, S., Hebbecker, K., Back, M. D., & Souvignier, E. (2022). Should teachers be accurate or (overly) positive? A competitive test of teacher judgment effects on students' reading progress. *Learning and Instruction*, *77*. https://doi.org/10.1016/j.learninstruc.2021.101519

Fu, R., Lee, J., Chen, X., & Wang, L. (2020). Academic self-perceptions and academic achievement in Chinese children: A multiwave longitudinal study. *Child Development*, *91*(5), 1718–1732. https://doi.org/10.1111/cdev.13360

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology, 52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.00

Geng, S., Lu, Y., & Shu, H. (2022). Cross-cultural generalizability of expectancy-value theory in reading: A multilevel analysis across 80 societies. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. https://doi.org/10.1007/s12144-022-03014-0

Gramzow, R. H., & Willard, G. (2006). Exaggerating Current and Past Performance: Motivated Self-Enhancement Versus Reconstructive Memory. *Personality and Social Psychology Bulletin*, *32*(8), 1114–1125. https://doi.org/10.1177/0146167206288600

Gramzow, G. H., Willard, G., & Berry Mendes, W. (2008). Big tales and cool heads: Academic exaggeration is related to cardiac vagal reactivity. *Emotion, 8*, 138-144. https://do.org/10.1037/1528-3542.8.1.138

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6,* 505–518. https://doi.org/10.1111/j.1475-6811.1999.tb00206.x

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science:

Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction, 49*, 81–91. http://dx.doi.org/10.1016/j.learninstruc.2016.12.007

Guo, X., Qin, H., Jiang, K., & Luo, L. (2022). Parent-child discrepancy in educational aspirations and depressive symptoms in early adolescence: A longitudinal study. *Journal of Youth and Adolescence*, *51*(10), 1983–1996. https://doi.org/10.1007/s10964-022-01644-y

Gutierrez de Blume, A. P. (2022). Calibrating calibration: A meta-analysis of learning strategy instruction interventions to improve metacognitive monitoring accuracy. *Journal of Educational Psychology*, *114*(4), 681–700. https://doi.org/10.1037/edu0000674.supp (Supplemental)

Gutierrez de Blume, A. P., Montoya Londoño, D. M., & Hederich- Martínez, C. (2022). An exploratory study of the relation between cognitive style and metacognitive monitoring in a sample of Colombian university students. *Psicología Desde El Caribe*, *39*(2). https://doi.org/10.14482/psdc.39.2.153

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Hrsg.), *The Cambridge handbook of cognition and education*. (S. 647–677). Cambridge University Press. https://doi.org/10.1017/9781108235631.026

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction*, *24*, 37–47. https://doi.org/10.1016/j.learninstruc.2012.10.001

Händel, M., Harder, B., & Dresel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction*, *65*. https://doi.org/10.1016/j.learninstruc.2019.101245

Hecht, C. A., Yeager, D. S., Dweck, C. S., & Murphy, M. C. (2021). Beliefs, affordances, and adolescent development: Lessons from a decade of growth mindset interventions. In J. J. Lockman (Hrsg.), *Advances in child development and behavior, Volume 61.* (S. 169–197). Elsevier Academic Press. https://doi.org/10.1016/bs.acdb.2021.04.004

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (S. 115-132). Beltz.

Hewitt, M. P. (2015). Self-efficacy, self-evaluation, and music performance of secondary-level band students. *Journal of Research in Music Education, 63*(3), 298–313. https://doi.org/10.1177/0022429415595611

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology*, *112*(7), 1409–1430. https://doi.org/10.1037/edu0000445

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019a). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859. https://doi.org/10.1037/pspp0000204

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology, 114*(2), 303–322. https://doi.org/10.1037/pspp0000134

Humberg, S., & Grund, S. (2022). Response surface analysis with missing data. *Multivariate Behavioral Research*, *57*(4), 581–602. https://doi.org/10.1080/00273171.2021.1884522

Humberg, S., Nestler, S., & Back, M. D. (2019b). Response surface analysis in personality and social psychology: Checklist and clarifications for the case of congruence hypotheses. *Social Psychological and Personality Science*, *10*(3), 409–419. https://doi.org/10.1177/1948550618757600

Humberg, S., Schönbrodt, F. D., Back, M. D., & Nestler, S. (2022). Cubic response surface analysis: Investigating asymmetric and level-dependent congruence effects with third-order polynomial models. *Psychological Methods*, *27*(4), 622–649. https://doi.org/10.1037/met0000352

Hurvich, C., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika, 76*, 297-307. https://doi.org/ 10.1093/BIOMET/76.2.297

lmarinen, V., Vainikainen, M., Verkasalo, M., & Lönnqvist, J. (2019). Peer sociometric status and personality development from middle childhood to preadolescence. *European Journal of Personality*, *33*(5), 606–626. https://doi.org/10.1002/per.2219

Irving, P. G., & Meyer, J. P. (1999). On using residual difference scores in the measurement of congruence: The case of met expectations research. *Personnel Psychology*, *52*(1), 85–95. https://doi.org/10.1111/j.1744-6570.1999.tb01814.x

Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology, 113*(2), 330–350. https://doi.org/10.1037/edu0000524.supp (Supplemental)

Joel, S., Maxwell, J. A., Khera, D., Peetz, J., Baucom, B. R. W., & MacDonald, G. (2022). Expect and you shall perceive: People who expect better in turn perceive better behaviors from their romantic partners. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000411.supp (Supplemental

John, O. P., & Robins, R. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219. https://doi.org/10.1037//0022-3514.66.1.206

Jopling, D. A. (1996). "Take away the life-lie …": Positive illusions and creative self-deception. *Philosophical Psychology*, *9*(4), 525–544. https://doi.org/10.1080/09515089608573198

Khuri, A. I., & Cornell. J. A. (1987). *Response surfaces: Designs and analyses*. Marcel Dekker.

Kim, Y.-H., Chiu, C., & Zou, Z. (2010). Know thyself: Misperceptions of actual performance undermine achievement motivation, future performance, and subjective well-being. *Journal of Personality and Social Psychology, 99*(3), 395–409. https://doi.org/10.1037/a0020555

Kiuru, N., Spinath, B., Clem, A.-L., Eklund, K., Ahonen, T., & Hirvonen, R. (2020). The dynamics of motivation, emotion, and task performance in simulated achievement situations. *Learning and Individual Differences*, *80*. https://doi.org/10.1016/j.lindif.2020.101873

Kuha, J. (2004). Comparisons of Assumptions and Performance. *Sociological methods & research, 33*(2), 188-229.https://doi.org/10.1177/0049124103262065

Kurman, J. (2006). Self-enhancement, self-regulation and self-improvement following failures. *British Journal of Social Psychology, 45*, 339-356. https://doi.org/10.1348/014466605X42912

Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology*, *112*(2), 364–383. https://doi.org/10.1037/edu0000377.supp

Leduc, C., & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences, 55*, 193–201. https://doi.org/10.1016/j.lindif.2017.04.006

Lee, E. J. (2021). Biased self-estimation of maths competence and subsequent motivation and achievement: Differential effects for high- and low-achieving students. *Educational Psychology, 41*(4), 446–466. https://doi.org/10.1080/01443410.2020.1821869

Lee, E. J. (2022). Are overconfidence and the accurate calibration of performance mutually incompatible? *Japanese Psychological Research*. https://doi.org/10.1111/jpr.12409

Leichner, N., Ottenstein, C., Eckhard, J., Matheis, S., Weis, S., Schmitt, M., & Lischetzke, T. (2022). Examining the congruence hypothesis in vocational interest research: The case of teacher students. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. https://doi.org/10.1007/s12144-022-03509-w

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the association of self-enhancement bias with self-ratings of personality and life satisfaction. *Assessment, 23,* 588–602. http://dx.doi.org/10.1177/1073191115590852

Li, X., Huebner, E. S., & Tian, L. (2021). Relations between achievement task values and academic achievement and depressive symptoms in Chinese elementary school students: Variable-centered and person-centered perspectives. *School Psychology*, *36*(3), 167–180. https://doi.org/10.1037/spq0000384.supp (Supplemental)

Li, G., & Xie, L. (2022). The effects of job involvement and supervisor developmental feedback on employee creativity: A polynomial regression with response surface analysis. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues.* https://doi.org/10.1007/s12144-022-02901-w

Liu, A. S., Rutherford, T., & Karamarkovich, S. M. (2022). Numeracy, cognitive, and motivational predictors of elementary mathematics achievement. *Journal of Educational Psychology*, *114*(7), 1589–1607. https://doi.org/10.1037/edu000077

Lohbeck, A. (2020). Does integration play a role? Academic self-concepts, self-esteem, and self-perceptions of social integration of elementary school children in inclusive and mainstream classes. *Social Psychology of Education: An International Journal*, *23*(5), 1367–1384. https://doi.org/10.1007/s11218-020-09586-8

Lopez, D. F., Little, T. D., Oettingen, G., & Baltes, P. B. (1998). Self-regulation and school performance: Is there optimal level of action-control? *Journal of Experimental Child Psychology, 70*, 54-74. https://doi.org/10.1006/jecp.1998.2446

Mac Iver, D. J., Stipek, D. J., & Daniels, D. H. (1991). Explaining within-semester changes in student effort in junior high school and senior high school courses. *Journal of Educational Psychology, 83*(2), 201–211. https://doi.org/10.1037/0022-0663.83.2.201

Marsh, H. W. (2022). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000750.supp (Supplemental)

Marsh, H. W. (2023). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology*, *115*(2), 193–211. https://doi.org/10.1037/edu0000750.supp

Marsh, H. W., Byrne, B. M., & 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. https://doi.org/10.1037/0022-0663.80.3.366

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. https://doi.org/ 10.1016/j.learninstruc.2014.08.005

Marsh, H. W., Pekrun, R., & Lüdtke, O. (2022). Directional ordering of self-concept, school grades, and standardized tests over five years: New tripartite models juxtaposing within- and between-person perspectives. *Educational Psychology Review.* https://doi.org/10.1007/s10648-022-09662-9

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, *111*(2), 331–353. https://doi.org/10.1037/edu0000281.supp

Martin, A. J., & Debus, R. L. (1998). Self-reports of mathematics self-concept and educational outcomes: The roles of ego-dimensions and self-consciousness. *British Journal of Educational Psychology, 68*(4), 517–535. https://doi.org/10.1111/j.2044-8279.1998.tb01309.x

McLarty, B. D., Whitman, D. S., Kluemper, D. H., & Tao, S. (2022). An identity and reputation approach to understanding the Dark Triad in the workplace. *Journal of Organizational Behavior*, *43*(3), 524–545. https://doi.org/10.1002/job.2569

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*(3), 1129-1167. https://doi.org/10.3102/0034654309337522

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research*, *90*(3), 376–419. https://doi.org/10.3102/0034654320919354

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115,* 502–517. http://dx.doi.org/10.1037/0033-295X.115.2.502

Mota, S., Humberg, S., Krause, S., Fatfouta, R., Geukes, K., Schröder-Abé, M., & Back, M. D. (2020). Unmasking Narcissus: A competitive test of existing hypotheses on (agentic, antagonistic, neurotic, and communal) narcissism and (explicit and implicit) self-esteem across 18 samples. *Self and Identity*, *19*(4), 435–455. https://doi.org/10.1080/15298868.2019.1620012

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the “x” out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science, 22*, 1058–1066. http://dx.doi.org/10.1177/0956797611415540

Nestler, S., Grimm, K. J., & Schönbrodt, F. D. (2015). The Social consequences and mechanisms of personality: How to analyse longitudinal data from individual, dyadic, round-robin and network designs. *European Journal of Personality, 29,* 272–295. http://dx.doi.org/10.1002/per.1997

Nestler, S., Humberg, S., & Schönbrodt, F. D. (2019). Response surface analysis with multilevel data: Illustration for the case of congruence hypotheses. *Psychological Methods, 24,* 291–308. http://dx.doi.org/10.1037/met0000199

Noble, R. N., Heath, N. L., & Toste, J. R. (2011). Positive illusions in adolescents: The relationship between academic self-enhancement and depressive symptomatology. *Child Psychiatry and Human Development, 42,* 650–665. http://dx.doi.org/10.1007/s10578-011-0242-5

OECD (2020a, June 30). *Scale: Selbstwirksamkeit Mathematik.* https://www.fdz-bildung.de/skala.php?la=en&skala_id=2776&erhebung_id=287

OECD (2020b, June 30). *Scale: Selbstwirksamkeit in Mathematik*. https://www.fdz-bildung.de/skala.php?la=en&skala_id=2749&erhebung_id=287

Paulhus, D. L. & John, O. P. (1998). Egoistic and moralistic bias in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025–1060. https://doi.org/10.1111/1467-6494.00041

Pavlov, G., Shi, D., & Maydeu-Olivares, A. (2020). Chi-square Difference Tests for Comparing Nested Models: An Evaluation with Non-normal Data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(6), 908-917. https://doi.org/10.1080/10705511.2020.1717957

Peiffer, H., Ellwart, T., & Preckel, F. (2020). Ability self-concept and self-efficacy in higher education: An empirical differentiation based on their factorial structure. *PLoS ONE*, *15*(7). https://doi.org/10.1371/journal.pone.0234604

Polzer, J. T., Kramer, R. M., & Neale, M. A. (1997). Positive illusions about oneself and one's group: Antecedents and consequences. *Small Group Research, 28,* 243–266. http://dx.doi.org/10.1177/1046496497282004

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. https://doi.org/10.1177/0016986216687824

Priess-Groben, H. A., & Hyde, J. S. (2017). Implicit theories, expectancies, and values predict mathematics motivation and behavior across high school and college. *Journal of Youth and Adolescence, 46*(6), 1318–1332. https://doi.org/10.1007/s10964-016-0579-y

Ramdass, D. H. (2010). Improving fifth grade students' mathematics self-efficacy calibration and performance through self-regulation training [ProQuest Information & Learning].

In Dissertation Abstracts International Section A: Humanities and Social Sciences (Vol. 70, Issue 7–A, p. 2388).

Rivers, A. S., Bosmans, G., Piovanetti Rivera, I., Ruan-Iu, L., & Diamond, G. (2022). Maternal and paternal attachment in high-risk adolescents: Unique and interactive associations with anxiety and depressive symptoms. *Journal of Family Psychology*, *36*(6), 954–963. https://doi.org/10.1037/fam0000989

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*(2), 340-352. https://doi.org/10.1037/0022-3514.80.2.340

Robins, R. W. & John, O. P. (1997). The quest for self-insight: Theory and research on the accuracy of self-perception. In H. Hogan, J. Johnson & S. Briggs (Hrsg.), *Handbook of personality psychology* (S. 649–679). Academic Press.

Rohr, M. E., & Ayers, J. B. (1973). Relationship of student grade expectations, selected characteristics, and academic performance. *Journal of Experimental Education, 41*(3), 58-62. https://doi.org/10.1080/00220973.1973.11011410

Schönbrodt, F. D. (2016). *Testing fit patterns with polynomial regression models*. Retrieved from osf.io/3889z

Schönbrodt, F. D., Humberg, S., & Nestler, S. (2017). Testing similarity effects with dyadic response surface analysis. *European Journal of Personality, 32,* 627–641. http://dx.doi.org/10.1002/per.2169

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly, 59*(3), 143–156. https://doi.org/10.1177/0016986215583870

Seber, G. A. F., & Lee, A. J. (2003). *Linear regression analysis* (2. Aufl.). Wiley Interscience.

Sewasew, D., & Koester, L. S. (2019). The developmental dynamics of students' reading self-concept and reading competence: Examining reciprocal relations and ethnic-background patterns. *Learning and Individual Differences, 73*, 102–111. https://doi.org/10.1016/j.lindif.2019.05.010

Sewasew, D., & Schroeders, U. (2019). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology, 58*, 204–212. https://doi.org/10.1016/j.cedpsych.2019.03.009

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology*, *25*(4), 543–554. https://doi.org/10.1007/s10869-010-9183-4

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46* , 407–441. https://doi.org/10.2307/1170010

Skaalvik, E. M., & Skaalvik, S. (2008). Self-concept and self-efficacy in mathematics: Relation with mathematics motivation and achievement. In F. M. Olsson (Hrsg.), *New developments in the psychology of motivation*. (S. 105–128). Nova Science Publishers.

Spencer, D., Nietfeld, J. L., Cao, L., & Difrancesca, D. (2023). Exploring the interplay between attributions and metacognitive monitoring ability in a post-secondary classroom. *Journal of Experimental Education*, *91*(1), 46–61. https://doi.org/10.1080/00220973.2021.1897773

Sun, X., Nancekivell, S., Gelman, S.A., & Shah, P. (2021). Growth mindset and academic outcomes: a comparison of US and Chinese students. *npj Science of Learning, 6*. https://doi.org/10.1038/s41539-021-00100-z

Talsma, K., Schüz, B., & Norris, K. (2019). Miscalibration of self-efficacy and academic performance: Self-efficacy ≠ self-fulfilling prophecy. *Learning and Individual Differences, 69,* 182–195. https://doi.org/10.1016/j.lindif.2018.11.002

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Tamm, A., Tulviste, T., & Tõnissaar, M. (2021). Values of adolescents and values prevailing in the classroom are related to adolescents' psychological adjustment. *European Journal of Developmental Psychology*, *18*(1), 56–74. https://doi.org/10.1080/17405629.2020.1755250

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210. https://doi.org/10.1037/0033-2909.103.2.193

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21–27. https://doi.org/10.1037/0033-2909.116.1.21

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97,* 1115–1128. http://dx.doi.org/10.1037/a0017048

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. http://dx.doi.org/10.1037/a0027470

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Hrsg.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (S. 187-214). Springer International Publishing.

Urban, D., & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Praxis (5. Aufl.)*. Springer VS.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

van Loon, M. H., & Oeri, N. S. (2023). Examining on-task regulation in school children: Interrelations between monitoring, regulation, and task performance. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000781

Wagenmakers, E.-J., Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*, 192-196. https://doi.org/10.3758/BF03206482

Wan, S., Lauermann, F., Bailey, D. H., & Eccles, J. S. (2021). When do students begin to think that one has to be either a “math person” or a “language person”? A meta-analytic review. *Psychological Bulletin*, *147*(9), 867–889. https://doi.org/10.1037/bul0000340.supp

Wang, M.-T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*(6), 1643–1657. https://doi.org/10.1037/a0027247

Wang, K., Li, F., Xu, J., Chen, S., & Zhou, M. (2022). Insecure attachment may not hamper relationships: A dyadic fit perspective. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. https://doi.org/10.1007/s12144-022-04005-x

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2020). The value of valuing math: Longitudinal links between students' intrinsic, attainment, and utility values and grades in math. *Motivation Science*, *6*(4), 413–422. https://doi.org/10.1037/mot0000179.supp (Supplemental)

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development, 89*(2), 138–156. https://doi.org/10.1111/cdev.12806

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*(1), 49–78. https://doi.org/10.1007/BF02209024

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Hrsg.), *Advances in motivation science, Volume 7* (S. 161–198). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.05.002

Willard, G., & Gramzow, R. H. (2009). Beyond Oversights, Lies, and Pies in the Sky: Exaggeration as Goal Projection. *Personality and Social Psychology Bulletin, 35*(4), 477-492. https://doi.org/10.1177/0146167208329631

Wille, E., Stoll, G., Gfrörer, T., Cambria, J., Nagengast, B., & Trautwein, U. (2020). It takes two: Expectancy-value constructs and vocational interests jointly predict STEM major choices. *Contemporary Educational Psychology*, *61*. https://doi.org/10.1016/j.cedpsych.2020.101858

Wolff, F., Helm, F., Junge, F., & Möller, J. (2020). Are dimensional comparisons performed unconsciously? An investigation of the internal/external frame of reference model using implicit self-concepts. *Journal of Educational Psychology*, *112*(2), 397–415. https://doi.org/10.1037/edu0000375.supp

Wolff, F., & Möller, J. (2022). An individual participant data meta-analysis of the joint effects of social, dimensional, and temporal comparisons on students' academic self-concepts. *Educational Psychology Review*, *34*(4), 2569–2608. https://doi.org/10.1007/s10648-022-09686-1

Wolff, F., Sticca, F., Niepel, C., Götz, T., Van Damme, J., & Möller, J. (2021a). The reciprocal 2I/E model: An investigation of mutual relations between achievement and self-concept levels and changes in the math and verbal domain across three countries. *Journal of Educational Psychology*, *113*(8), 1529–1549. https://doi.org/10.1037/edu0000632

Wolff, F., Zitzmann, S., & Möller, J. (2021b). Moderators of dimensional comparison effects: A comprehensive replication study putting prior findings on five moderators to the test and going beyond. *Journal of Educational Psychology*, *113*(3), 621–640. https://doi.org/10.1037/edu0000505.supp

Wright, S. (2000). Looking at the self in a rose-colored mirror: Unrealistically positive self-views and academic performance. *Journal of Social and Clinical Psychology, 19*(4), 451-462. https://doi.org/10.1521/jscp.2000.19.4.451

Wright, A. J., & Jackson, J. J. (2023). Is parent personality associated with adolescent outcomes for their child? A response surface analysis approach. *Infant and Child Development*. https://doi.org/10.1002/icd.2395

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09600-1

Wylie, R. (1961). *The self-concept: A critical survey of pertinent research literature.* University of Nebraska Press.

Ximénez, M. C., & San Martín, R. (2000). Application of response surface methodology to the study of person–organization fit. *Psicothema*, *12*(1), 151–158.

Yeung, S. S. S., King, R. B., Nalipay, M. J. N., & Cai, Y. (2022). Exploring the interplay between socioeconomic status and reading achievement: An expectancy-value perspective. *British Journal of Educational Psychology*, *92*(3), 1196–1214. https://doi.org/10.1111/bjep.12495

Zhang, J., Chiu, M. M., & Lei, H. (2023). Achievement, self-concept and anxiety in mathematics and English: A three-wave cross-lagged panel study. *British Journal of Educational Psychology*, *93*(1), 56–72. https://doi.org/10.1111/bjep.12539

Zhou, Y., Qu, D., Xu, C., Zhang, Q., & Yu, N. X. (2023). How does (in)congruence in perceived adolescent–parent closeness link to adolescent socioemotional well-being? The mediating role of resilience. *Journal of Happiness Studies: An Interdisciplinary Forum on Subjective Well-Being*. https://doi.org/10.1007/s10902-023-00624-8

2 Beiträge der kumulativen Dissertation

2.1 Beitrag I

Separating the effects of self-Evaluation bias and self-View on grades

Paschke, P., Weidinger, A. F., & Steinmayr, R. (2020). Separating the effects of self-evaluation bias and self-view on grades [Trennung der Effekte des Selbsteinschätzungsbias und der Selbsteinschätzung auf Schulnoten]. *Learning and Individual Differences, 83-84*. https://doi.org/10.1016/j.lindif.2020.101940

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Note: This article is not the copy of record and may not exactly replicate the final, authoritative version of the article.

Abstract

The present study aimed at investigating whether the absolute level of a person's self-view of competence or the discrepancy between the self-view and actual competence (self-evaluation bias) affect academic achievement operationalized as grades. An important theoretical and methodological problem in studies on this topic has been pointed out: The effects of the absolute level of the self-view and the self-evaluation bias have been confounded. We used a novel approach (condition-based regression analysis; Humberg et al., 2018) to disentangle these effects. We investigated 284 German students in Grade 10. Ability self-concepts and competences in math and German were assessed at *t*1. Additionally, we received the students' last report card grades and the grades from the following year (*t*2). The absolute level of ability self-concept significantly predicted subsequent grades within both domains when controlling for former grades, gender and socioeconomic variables. No effect of self- evaluation bias on grades was found.

Keywords: ability self-concept, self-evaluation bias, condition-based regression analysis, academic achievement

Zusammenfassung

Ziel der vorliegenden Studie war die Untersuchung der Fragestellung, ob die Ausprägung der Kompetenzselbsteinschätzung von Personen oder die Diskrepanz zwischen der Selbsteinschätzung und der tatsächlichen Kompetenz (Selbsteinschätzungsbias) sich auf deren akademische Leistung, operationalisiert über Noten, auswirkt. Ein bedeutendes theoretisches und methodisches Problem voriger Studien zu diesem Thema wurde bereits herausgestellt: Die Effekte der Selbsteinschätzung an sich und des Selbsteinschätzungsbias wurden konfundiert. Wir verwendeten einen neuen Ansatz (condition-based regression analysis; Humberg et al., 2018), um diese Effekte voneinander zu trennen. Wir untersuchten 284 deutsche Schüler*innen der zehnten Klassenstufe. Zum ersten Messzeitpunkt (*t*1) erfassten wir das Fähigkeitsselbstkonzept und die Kompetenzen der Schüler*innen in Mathematik und Deutsch. Zusätzlich erhielten wir die letzten Zeugnisnoten der Schüler*innnen und die Zeugnisnoten des Folgejahres (*t*2). Das Fähigkeitsselbstkonzept sagte in beiden Domänen die Zeugnisnoten des Folgejahres auch unter Kontrolle der vorigen Noten, des Geschlechts und sozioökonomischer Variablen vorher. Es wurden keine Effekte eines Selbsteinschätzungsbias auf Noten gefunden.

Stichwörter: Fähigkeitsselbstkonzept, Selbsteinschätzungsbias, condition-based regression analysis, Akademische Leistung

1. Introduction

Having high opinions on one's own competences and abilities positively affects students' academic achievement, for instance their grades, standardized test scores, or attainment (e.g., Huang, 2011; Trautwein & Möller, 2016; Valentine, DuBois, & Cooper, 2004). We call these effects self-view effects. Self-view effects on academic achievement have been demonstrated for different self-view related constructs such as ability self-concept (ASC), self-efficacy, or self-esteem (e.g., Usher, Li, Butz, & Rojas, 2019; Weidinger, Steinmayr, & Spinath, 2018; see Valentine et al., 2004 for a meta-analysis). But does the bias in the self-view, which is the discrepancy between the self-view and a related objective measure, also have an effect on academic achievement? We call effects of a bias in a self-view self-estimation bias effects (SE bias effects) and investigated whether these SE Bias effects on achievement occur beyond the self-view effects. Therefore, in the present study, we contrast self-view effects against SE bias effects. Many authors argue for positive SE bias effects (e.g., Bonneville-Roussy, Bouffard, &Vezeau, 2017; Chung, Schriber, & Robins, 2016; Dupeyrat, Escribe, Huet, & Régner, 2011). Others, however, warn against negative consequences of overly optimistic self-views (e.g., Buckelew, Byrd, Key, Thornton, & Merwin, 2013; Chiu & Klassen, 2009, 2010; Dunning, Johnson, Ehrlinger, & Kruger, 2003). The results of empirical studies on associations between SE bias and academic achievement are heterogeneous with some showing positive and some showing negative relations (e.g., Buckelew et al., 2013; Chiu & Klassen, 2009, 2010; Chung et al., 2016; Dupeyrat et al., 2011). Additionally, the interpretability of studies on SE bias effects is often limited by confounding SE bias effects with self-view effects, both theoretically and statistically (Humberg et al., 2018; Humberg et al., 2019). In the present study we adopt a novel approach called condition-based regression analysis (CRA, Humberg et al., 2018) that was developed to disentangle self-view effects and SE bias effects. Thus, unlike traditional approaches, the CRA allowed us to test whether there are SE bias effects that are independent of self-view effects. We used this approach to analyze self-view effects and SE bias effects on academic achievement measured as grades in two school subjects, namely math and German.

1.1. Conceptual Distinction between Self-view and SE Bias Effects

It is important to distinguish the effects of the SE bias from the effects of the self-view alone because they imply different ways in which self-view and competence relate to the outcome, for example school grades. In order to illustrate this, consider the following example. Two students, Peter and Tom, have rated their own mathematical competence and completed a standardized math competence test. The results of the self-rating and the competence test are located on the same scale ranging from 0 to 10. Peter scored 7 on the self-rating and 4 on the competence test, while Tom scored 8 on the self-rating and 6 on the competence test. If only a (positive) self-view effect on subsequent grades is present, we would expect Tom's math grade to be better since his self-rating exceeds Peter's. If, however, only a (positive) SE bias effect is present, we would expect Peter's grade to be better because his SE bias (7-4=3) is greater than Tom's (8-6=2). Since standardized competence test scores predict subsequent grades (Möller, Zimmermann, & Köller, 2014) and teachers base their grading processes, among other factors, on competence (Brookhart et al., 2016), subsequent grades will likely also rely on the results in the competence tests. Thus, there are three possible effects on students' grades that need to be separated: an effect of the competence alone (competence effect), an effect of the self-view alone (self-view effect), and an effect of the discrepancy between the two (SE bias effect).

1.2. Ability self-concept and academic achievement

Self-view is not a distinct psychological construct. Rather there are different self-view related constructs such as ability self-concept (ASC), self-efficacy, or self-esteem. Researchers studying SE bias effects on academic achievement have used different constructs, most commonly ASC or self-efficacy, to measure SE bias. In the present study we assessed ASC which is one of the most prominent self-view constructs in psychological and educational research (Bong, Cho, Ahn, & Kim, 2012; Bong & Skaalvik, 2003). According to Bong and Skaalvik (2003), "self-concept represents one's general perceptions of the self in given domains of functioning" (p. 5). By now, there is a large body of evidence for a positive reciprocal relation between ASCs and academic achievement (reciprocal effects model; e.g., Arens et al., 2017; Marsh & Martin, 2011; but see Burns, Crisp, & Burns, 2019). The reciprocal effects model appears to be highly cross-cultural (e.g., Kalogiannis, Papaioannou, Sagovich, & Abatzoglu, 2011; Lee & Kung, 2018; Marsh, Hau, & Kong, 2002) and hold equally for students in mixed-ability and academically selective groups (Seaton, Marsh, Parker, Craven, & Yeung, 2015; Preckel et al., 2017).

The effect of ASC on academic achievement is of particular interest in the present study because it is an example of a self-view effect. We chose to assess ASC as the self-view construct because of the large body of literature outlined above which allowed us to derive theoretically well-founded hypotheses about self-view effects in our study. Note however, that a meta-analysis (Valentine et al., 2004) found small effects of self-view constructs on subsequent academic achievement, independent of which specific self-view construct was assessed (ASC, self-efficacy, or self-esteem). Thus, we would not expect large differences in self-view effects if we assessed another self-view construct than ASC.

Regarding mechanisms by which ASC affects academic achievement, adaptive academic behavior such as effort (Helmke, 1990; Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009; Wigfield & Eccles, 2000; see Trautwein & Möller, 2016) as well as intrinsic and autonomous academic motivation (Areepattamannil, 2012; Guay, Ratelle, Roy, & Litalien, 2010) and lower test anxiety (Helmke, 1990) have been identified as relevant mediators.

1.3. SE bias and its effects on academic achievement

Several different terms have been used to describe differences between self-views and reality criteria (e.g. self-enhancement and self-evaluation bias; Humberg et al., 2018;

Bonneville-Roussy et al., 2017). In the present study, we opted for the term self-evaluation bias to avoid terminological confusion with the self-enhancement-model (Calsyn & Kenny, 1977) that describes a positive effect of ASC on academic achievement.

Some researchers argue for academic benefits of overrating one's competences (i.e., positive self-evaluation bias). However, the theoretical rationales behind these SE bias effects are often rather weak, especially due to confusing arguments for SE bias effects with arguments for self-view effects. For example, in their influential paper on "positive illusions", Taylor and Brown (1988) argue for positive SE bias effects on the ground that "Positive conceptions of the self are associated with working harder and longer on tasks" (p. 199). However, this not an argument for an SE bias effect, but an argument for a self-view effect (see also Humberg et al., 2018). Only if the difference between the self-conception and some kind of reality criterion was associated with working harder and longer on tasks, then that would be an argument for an SE bias effect. Concerning SE bias effects on academic achievement, it has been argued that a positive SE bias can make children perceive academic difficulties as challenges and attribute failures externally, which increases engagement and thus achievement (Bonneville-Roussy et al., 2017). Positive SE biases have also been argued to result in a greater intent to pursue careers in the respective domain (Bench, Lench, Liew, Miner, & Flores, 2015), psychological adjustment in the academic context (Chung et al., 2016; Dufner, Reitz, & Zander, 2015), better self-regulation and less school alienation (Leduc & Bouffard, 2017), and task persistence (Helmke, 1998). However, these rationales are ambiguous, since it does not become clear why they are arguments for SE bias effects instead of self-view effects (see also pp. 838 - 839 and footnote 3 in Humberg et al., 2019). For example, why should a greater SE bias instead of a greater self-view per se make children perceive academic difficulties as challenges? What is it about the discrepancy between the self-view and actual competence that affects the outcome variable beyond the main effects of the two variables? To our knowledge, no theoretical rationale for positive SE bias effects on academic achievement that is thoroughly distinguished from self-view effects has yet been publicized. Negative SE bias effects on academic achievement have been discussed much less than positive effects. A negative SE bias effect would imply that underestimation is beneficial while overestimation is detrimental for academic achievement. The only rationale of that kind that we know of is that high achieving students use underestimation as a motivational technique to increase effort while poor students develop unrealistic study patterns due to overestimating themselves (Rohr & Ayers, 1973).

Additionally, other authors argue against benefits of SE biases on the ground that recognizing one's own limitations is important to facilitate behavior change and learning (Dunlosky & Rawson, 2012; Dunning, Heath, & Suls, 2004; Dunning et al., 2003; Epley & Dunning, 2006; Hacker & Bol, 2019; Nietfeld & Schraw, 2002; Thiede, 1999; Thiede, Anderson, & Therriault, 2003). Only if people know where they have strengths or weaknesses, they can make efficient decisions on where to apply effort (Dunning et al., 2003).

Empirical studies have found both positive (e.g., Bonneville-Roussy et al., 2017; Côte, Bouffard, & Vezeau, 2014; Dupeyrat et al., 2011; Leduc & Bouffard, 2017) and negative associations between SE bias and academic achievement (e.g., Bol, Hacker, O'Shea, & Allen, 2005; Buckelew, et al., 2013; Chiu & Klassen, 2009; 2010). However, there is a general tendency for studies assessing SE bias prior to academic achievement to find positive associations (but see Robins & Beer, 2001). However, to date, there is still no consensus on whether SE biases have positive, negative, or no consequences for academic achievement (Bonneville-Roussy et al., 2017; Trautwein & Möller, 2016). Moreover and most importantly, it was shown that virtually all studies on associations between SE bias and other variables suffer from a major methodological problem. That is, methods were used that statistically confounded the SE bias with the self-view alone. Because of that, it can't be ruled out that any observed relation between SE bias and a second variable is an artifact of the relation between the self-view alone and the second variable (Humberg et al., 2018).

1.4. Statistical confounding of self-view and SE bias effects

Various different measures have been used to quantify SE bias. These measures can be classified as algebraic difference scores or residual scores (Humberg et al., 2018). Using algebraic difference scores in this context has been criticized for confounding the level of the SE bias with the levels of both the self-view and the reality criterion (e.g., Asendorpf & Ostendorf, 1998; Edwards & Parry, 1993; Griffin, Murray, & Gonzalez, 1999; Humberg et al., 2018). Additionally, Humberg et al. (2018) have shown that both approaches, the use of algebraic difference scores and the use of residual scores, statistically confound SE bias and self-view effects. Since it is well-known that students' self-view of their own abilities alone does indeed affect their academic achievement, including grades (e.g., Trautwein & Möller, 2016; Valentine et al., 2004), actual self-view effects on academic achievement might have been falsely interpreted as SE bias effects in the literature. Therefore, based on prior results, it is not possible to draw a clear conclusion about the existence and direction of SE bias effects on subsequent academic achievement.

1.5. The Present Research

The aim of the present study is to help answer the question whether there are positive, negative, or any SE bias effects on grades by separating SE bias effects from self-view effects. In order to achieve this, we used the newly developed CRA (Humberg et al., 2018) to analyze SE bias and self-view effects in two academic domains, namely math and German, in 10th grade students. We chose students relatively late in their school career as participants because there is some evidence that self-view effects increase with age (Chen, Yeh, Hwang, & Lin, 2013; Skaalvik & Hagtvet, 1990; Weidinger et al., 2018) and we expected this might also be the case for SE bias effects. Thus, since this is the first study to separate self-view from SE bias effects, we considered it useful to examine a sample where these effects might be most pronounced.

The influence of ASCs on grades is empirically well established (Trautwein & Möller, 2016; Valentine et al., 2004). This effect is likely mediated by academic effort, motivation, and test anxiety (e.g., Guay et al., 2010; Trautwein et al., 2009). Since in the present study we chose a relatively long interval of 8 to 9 months between our *t*1 measurements and the date on which students received the *t*2 grades from the schools, we expected that there would be ample opportunities for the relevant mediator effects to take place. For example, if students with a high ASC show increased effort and motivation during the 8 to 9 month interval, this should positively affect their grades.

We also included a set of additional predictors, namely *t*1 grades in the same domain, gender, as well as two indicators of socioeconomic background – parental educational level and number of books present at the students' home. ASCs, standardized competence test scores, and grades are usually all positively correlated within the same domain (Möller, Pohlmann, Köller, & Marsh, 2009; Möller et al., 2014). Thus, without including *t*1 grades in our models, the common link between *t*1 grades and both *t*2 grades as well as ASC and competence could lead to a spurious effect of ASC and competence on *t*2 grades. In other words, without including *t*1 grades, any observed path from ASC or competence on *t*2 grades might not reflect a causal link but simply a correlation of grades with ASC and competence which is relatively stable over time. By including *t*1 grades, we were able to analyze how ASC and competence predict *t*2 grades when the influence of *t*1 grades is controlled. We included gender and the two socioeconomic background variables for similar reasons. Gender is known to be correlated with verbal and math ASCs as well as verbal grades (Skaalvik & Skaalvik, 2004; Wilgenbusch & Merrell, 1999). Additionally, in large scale assessments it has been shown that in Germany, socioeconomic variables such as parental educational level and number of books are related to reading comprehension (Hußmann et al., 2017) and math competence (Stubbe et al., 2016; see also OECD, 2016). Moreover, in Germany, students' socioeconomic background is related to the grades assigned by teachers beyond the effect of the students' competence (Maaz, Baeriswyl, & Trautwein, 2013). Thus, since socioeconomic variables and competence are correlated and there is a direct relation of socioeconomic variables with grades, omitting socioeconomic variables as predictors could lead to part of the observed path coefficient from competence on grades being caused by an effect of socioeconomic variables.

We hypothesized that ASC would positively predict grades in both domains. Moreprecisely, we expect that *t*1 math ASC predicts *t*2 math grades when controlling for *t*1 math grades, *t*1 math competence, gender, parental educational level, and number of books at home (Hypothesis 1a). Moreover, we expect that *t*1 German ASC predicts *t*2 German grades when controlling for *t*1 German grades, *t*1 German competence, gender, parental educational level, and number of books at home (Hypothesis 1b). Because of the confounding of SE bias effects with self-view effects in virtually every study on SE bias effects as well as the contradictory and ambiguous theoretical assumptions behind SE bias effects, current knowledge does not allow deriving a hypothesis about the existence and direction of SE bias effects. We thus investigated the research question (RQ) whether there are SE bias effects on students' math grades (RQ 1a) and German grades (RQ 1b) beyond the effects of *t*1 grades, gender, parental educational level and number of books at home, when using a statistical approach that separates SE bias effects from self-view effects.

Additionally, we tested SE bias effects using the two most prominent approaches so far applied in the literature on this subject, algebraic difference scores and residual scores. This allowed us to compare the results produced by the application of the novel CRA to those produced by more traditional methods. Since these methods have yielded contradicting results for SE bias effects (Bonneville-Roussy et al., 2017; Trautwein & Möller, 2016), we did not derive hypotheses about whether or not applying these methods would result in positive, negative, or any SE bias effects. Instead, our aim was an explorative comparison between traditional and more advanced methods.

2. Method

2.1. Participants

We examined 284 10th grade students (160 girls and 121 boys; 3 students did not indicate their gender) from two academic high schools ("Gymnasium") in a town in western Germany. In Germany, after four years of elementary education, students are split into one of three types of schools that vary in their academic demands or an integrated school that combines the three types. The Gymnasium, which the participants in the present investigation attended, is the most academically advanced of the three types and the only one besides the integrated school that leads to the Abitur, a degree that qualifies students to enter a university. Students followed a G8-track which leads to the Abitur after 8 years attending a Gymnasium (in 12th grade). Thus, they were already in the first grade of the secondary phase of scholastic education after elementary school (Oberstufe) which begins in 10th grade. In this phase, students are not taught in classes anymore, but in courses, which means that the composition of students varies between different subjects.

At *t*1, participants were on average 15.26 years old (SD = 0.54). Of all participants, 92.3% were born in Germany, 2.8% were born in another country, and 4.9% did not indicate their country of birth. Regarding the participants' first language, 82.0% were native German speakers, 7.4% did not learn German as a first language, and 10.6% did not provide information on this variable. Regarding the parental educational level, 180 participants (63.4%) had at least one parent with an Abitur or Fachabitur (a degree that does not qualify students to enter a university but a Fachhochschule which is comparable to a university of applied sciences) which is similar to the overall percentage in Germany in 2015 (61%; Statistisches Bundesamt, 2016). Additionally, 82 (28.9%) students did not have a parent with an Abitur or Fachabitur and 22 (7.7%) did not provide information on this variable. There was no student who reported that both of their parents did not graduate from secondary school. We also assessed the number of books at each student's home and 154 (54.2%) had more than 100 books, while 112 (39.4%) had 100 or fewer books, and 18 (6.3%) did not provide information on this variable.

2.2. Procedure

The present project is in accordance with established ethical guidelines for psychological research. The participants are not part of a vulnerable group and were not misled about the purpose of the study. We did not use any exclusion or inclusion criteria, possibly stigmatizing questions, or interventions that could cause mental or physical harm. Approval by an ethics committee was not required as per the institution's guidelines and applicable regulations in the federal state in which the study was conducted. We received informed consent forms from the parents of all participating students. Moreover, the responsive school administrations approved the study design and the data collection procedure beforehand. Participation was voluntary, and all present students agreed to participate. Students not attending at one of the measurement occasions were either ill, participated in a different extra-curricular activity, or spend the semester in which *t*1 took place abroad.

The *t*1 measurements took place in October/November 2014 when students were at the beginning of Grade 10. About 85% of the basic population participated at this measurement time point. Trained research assistants administered the measures and instructed the participants in groups of about 20 students. Students were informed that their responses were anonymous and that their teachers would have no access to the data. Self-reports, including the ASC measures, were administered before the competence tests. Only the tests relevant for the present study are reported below. Additionally, at *t*1 we received students' grades on the last report card they received from the schools. At *t*1, these were the report cards from the end of Grade 9 and thus from July 2014.

At *t*2, we again received the students' grades from their last report card. These were the grades from the end of Grade 10 and thus from July 2015. Thus, the interval between the *t*1 measurement of ASCs and competences and the time point at which the students received the *t*2 grades from the schools was 8 to 9 months.

Data were entered by trained research assistants who were advised to look for response patterns that could indicate random, careless, or other non-valid answering.

2.3. Measures

2.3.1. Math and German ability self-concepts

Students' ASCs in math and German were each assessed with four subject-specific items from the absolute school self-concept scale of the SESSKO (German Scales for the Assessment of School-Related Competence Beliefs; Schöne, Dickhäuser, Spinath, & Stiensmeier-Pelster, 2002, p. 26). English versions of these four items are reported by Spinath and Steinmayr (2012, p. 1148). The same items have been successfully used to assess ASCs in other studies before (e.g., Steinmayr, Weidinger, & Wigfield, 2018). All items are made up of self-referencing statements about the students' competences in math and German which they rate on a scale from 1 (totally disagree) to 5 (totally agree). An example item is "I am good in math/German". We computed separate scales for math and German ASCs. The internal consistencies of both scales were high (math: $\alpha = .95$; German: $\alpha = .90$).

2.3.2. Math competence

Students' math competence was assessed with 42 items from the Trends in International Mathematics and Science Study (TIMSS) of the IEA (International Association for the Evaluation of Educational Achievement; Baumert et al., 1998). Items were selected based on item-test correlation and difficulty to ensure that the whole ability range was covered. A version covering 36 of the selected 42 items of the TIMSS-test was used successfully before (Steinmayr & Meißner, 2013). The items include multiple choice as well as open-answer questions that cover six content domains: algebra, data display and analysis, numbers and numerical reasoning, geometry, measures and measuring units, and proportionality. Each content domain was assessed with 7 items. All items were combined to obtain a global scale of math competence. The internal consistency of the total scale was good (α = .82). Since students could get 1 point per correct answer, the possible range for the total score was 0 to 42.

2.3.3. German competence

Students' German competence was computed as a composite score by building the average of z-standardized reading comprehension scores and z-standardized orthography scores.

2.3.3.1. Reading comprehension. Students' reading comprehension was assessed with the LGVT 6-12 (Reading speed and comprehension test for classes 6-12; Schneider, Schlagmüller, & Ennemoser, 2007). In the LGVT 6-12 participants are asked to read a specific text, containing about 2000 words, as quickly as possible within 4 minutes. In some sentences, one word is replaced by three different words. Only one of these three words fits the context of the sentence and text. The participants have to underline the word they believe to be the correct choice. The reading comprehension score is calculated as the number of correct choices multiplied by two minus the number of wrong choices. There are 23 instances where the participants have to underline a word, so the possible score ranges from -23 to 46. Since usually less than 1% of the participants finish the text within the given time (Schneider et al., 2007), the LGVT 6-12 has a pronounced speed aspect. Thus, internal consistency and split-half reliability are not useful indicators of the instrument's reliability, and we did not compute these measures. Schneider et al. (2007) report a retest-reliability over a period of 6 weeks for the reading comprehension score of $r_{tt} = .87$.

2.3.3.2. Orthography. Students' orthography was assessed with the RT orthography test (Kersting & Althoff, 2004). In the RT, participants are asked to read a text in which some words are replaced by different orthographic versions of the word, only one of which is correct. The participants chose the word they believe to be orthographically correct. The orthography score is computed as the number of correct choices. Since there are 44 instances where the participants have to choose a word, the possible score ranges from 0 to 44. The internal consistency of the RT was acceptable (α = .70).

2.3.4. Math and German grades

We chose grades as the measure of academic achievement because of their high practical relevance for students (Brookhart et al., 2016). It should be noted that grades are not a measure of competence but "a mixture of multiple factors that teachers value" (Brookhart et al., 2016, p. 834) such as effort and motivation (Cizek, Fitzgerald, & Rachor, 1996; Kelly, 2008; Guskey, 2002; McMillan, 2001; see Brookhart et al., 2016). Thus, what we investigated was not how SE bias affects students' competence but a more general measure of academic achievement, which is the most common indicator of achievement in schools. We assessed students' grades on the last report card they received, which was the report card at the end of Grade 9 at $t1$ and the report card at the end of Grade 10 at $t2$. We received students' report card grades from the schools. In Germany, at the time the present study was conducted, the grading system changed between Grade 9 and Grade 10. Up to Grade 9 grades ranged from 1 (very good) to 6 (insufficient). From Grade 10 onwards grades ranged from 0 to 15 with higher scores indicating better grades. We recoded the students' grades on the report card from Grade 9 so that they are located on the same scale as the grades from Grade 10. In the German school system, each of the grades from 1 to 6 corresponds to one of the grades from 0 to 15 in the following way: 1 = 14; 2 = 11; 3 = 8; 4 = 5; 5 = 2; 6 = 0. We used this system to recode the $t1$ grades.

2.3.5 Socioeconomic variables

We also assessed two indicators of students' socioeconomic background, the number of books at the students' home and parental educational level. The number of books at home was assessed with the question "How many books are present at your home? *Approximately 40 books fit on one bookshelf. Magazines, newspapers, and your school books are not included.*" There were six response options ranging from "fewer than 10" to "more than 500".

We dichotomized the answers into the categories "up to 100" (coded as 1) and "more than 100" (coded as 2) which is the common procedure in large scale assessments like TIMSS and IGLU (e.g., Stubbe, Schwippert, & Wendt, 2016; Hußman, Stubbe, & Kasper, 2017). Parental educational level was assessed with the question "Which qualification do your parents have? Please select only the highest qualification." followed by a list of secondary school qualifications students can obtain in the German school system. Students selected one qualification for each parent and we treated the higher one as the parental educational level. We dichotomized the responses into the categories "Fachabitur or Abitur" (coded as 1) and "lower qualifications" (coded as 2).

2.4. Statistical Analysis

An adaption of the condition-based regression analysis, confirmatory factor analyses, as well as all regression analyses were computed in R 3.4.3 with the packages lavaan and semTools and an adaption of the script provided by Humberg et al. (2018). Further analyses were computed in IBM SPSS Statistics 26. To account for the missing data, full information maximum likelihood estimation (FIML) with auxiliary variables (Enders, 2010) was used in all structural equation models (Little, 2013). Detailed information on missing data can be found in the online supplementary material (OSM S1).

2.4.1. Condition-based regression analysis

We used an adaption of the condition-based regression analysis developed by Humberg et al. (2018) to analyze SE bias and self-view effects. In this approach, no SE bias scores are computed. Instead, a multiple linear regression with a self-view (e.g., math ASC) and a reality criterion (e.g., math competence) as the predictors and the outcome of interest (e.g., math grades) as the criterion is computed. The self-view effect is tested by a significance test of the regression coefficient of the self-view. On the other hand, an SE bias effect is present if, and only if, the regression coefficients of the self-view and the reality criterion have opposite algebraic signs. However, even if this condition holds, it has only been shown that an SE bias effect is present, not that it is significant. Thus, if this condition holds, the CRA computes a significance test for that same condition. If the condition does not hold, no SE bias effect is present and thus no interpretable significance test can be computed. For more detailed information on the computation of the significance test, see the example R code for the CRA provided by Humberg et al. (2018).

We adapted the CRA slightly so that we could analyze our data on a latent level. We modelled the two predictor variables, *t*1 ASC and *t*1 competence test score, as latent instead of observed variables. Thus, we estimated two structural equation models, one for math and one for German, which are depicted in Figure 1 and Figure 2, respectively. In order to test the SE bias effect, we analyzed the signs of the path coefficients in these structural equation models instead of regression coefficients. In each model, an SE bias effect would be present if the path coefficients of the *t*1 ASC and the *t*1 competence test score on the *t*2 grades have opposite algebraic signs. Only if that is the case, that effect is then tested for significance. Additionally to ASC and competence, we included *t*1 grades in the same domain, gender, parental educational level, and number of books at home as exogenous variables.

In summary, *t*2 grades were predicted by six variables: *t*1 grades, *t*1 ASC, and *t*1 competence test scores (called *t*1 variables in the following), as well as gender, parental educational level, and number of books (called covariates in the following).

In both models, we freely estimated the correlations between the *t1* variables because these variables are known to be correlated within the same domain (Möller et al., 2009; Möller et al., 2014). In order to control the effects of the covariates on the effects of the *t*1 variables on *t*2 grades, we also modelled effects of the covariates on the *t*1 variables. The correlations between the covariates were freely estimated. For a graphical representation of the models and the results see Figures 1 and 2.

Due to the large number of individual items in the competence tests, we used item parcels as observed variables in the models. We computed six parcels for math competence, each consisting of the seven items from one of the six content domains of the TIMSS items. For German, we computed two parcels per test, two for the LGVT and two for the RT. We used two parcels per test instead of just one because otherwise German competence would not be an identified construct in our model (Little, 2013; p. 85). We allocated the items to the parcels based on the recommendations by Little (2013; p. 24). Since orthography and reading comprehension are different aspects of German competence, there was a theoretical reason to assume that the two LGVT parcels and the two RT parcels each would have a certain amount of shared variance not explained by German competence. Therefore, we freely estimated the error covariances between parcels of the same test but not between parcels of different tests. We computed the following fit indices to evaluate the goodness of fit of the estimated models: the chi-square test statistic (χ^2), the root mean square error of approximation (RMSEA) along with associated 90% confidence interval (CI), the standardized root mean square residual (SRMR), and the comparative fit index (CFI). The following cutoff criteria for the fit indices were used to evaluate model fit: RMSEA < .06, SRMR < .08, CFI > .95 (see West, Taylor, & Wu, 2012).

2.4.2 Preliminary analyses and traditional approaches to SE bias effects

We tested the structural validity of the measurement instruments with confirmatory factor analyses (CFAs) and by inspecting the item's cross-loadings. For details on the computation and the results of these analyses, see OSM S2 and S3. In order to test if and how the inclusion of the covariates in the models changed the content of the residualized ASC and competence scores, we computed the ASC and competence scores' double-entry intraclass correlations as a similarity index (r_{ICC}; Vize, Collison, Miller, & Lynam, 2018; for more details, see OSM S4 and OSM S5). Additionally to the CRA, we also tested SE bias effects on grades by using algebraic difference and residual scores. For details on the computation of these effects and the results (see OSM S6 and S7).

3. Results

3.1 Preliminary analysis

The results of the CFAs of the different measures are reported in detail in OSM S3. Overall, all instruments showed good structural validity according to all established cutoff-criteria except for the RMSEA of the ASC scales (RMSEA = .076). The computation of r_{ICC} indicated that the content of the ASC and competence scores before and after partialling out the covariates was highly similar (all r_{ICC} between the same constructs and the same domain ≥ .94). For detailed results see OSM S5.

3.2. Descriptive Results and correlations

Sample sizes, means, standard deviations, skewness, kurtosis, and internal consistencies of the nondichotomous model variables are reported in Table 1 and correlations between all analysis variables are reported in Table 2. Within each domain, all correlations between students' competence, ASCs, and grades were significantly positive. In both domains, ASC correlated moderately to strongly with the $t1$ and $t2$ grades ($.46 \leq r \leq .57$; all $p < .001$). The correlations of competence with $t1$ and $t2$ grades were slightly smaller ($.38 \leq r \leq .44$; all $p < .001$). The correlations between ASC and competence were moderate (math: $r = .37$, $p < .001$; German: $r = .33$; $p < .001$) and the correlations between $t1$ and $t2$ grades were high (math: $r = .62$, $p < .001$; German: $r = .56$, $p < .001$). Regarding cross-domain correlations, students' math and German ASCs correlated weakly and negatively ($r = -.25$, $p < .001$), while students' math and German competences ($r = .39$, $p < .001$) and their math and German grades ($t1$: $r = .43$, $p < .001$; $t2$: $r = .34$, $p < .001$) correlated moderately and positively. Gender and the two socioeconomic variables showed several significant correlations with ASCs, competence scores and grades, however, all of them were small ($-.29 \leq r \leq .23$).

Table 1

Sample sizes (n), means (M), standard deviations (SD), skewness, kurtosis, and Cronbach's α for all nondichotomous analyses variables

	n	*M*	*SD*	Skewness	Kurtosis	α
Math ASC *t*1[a]	270	3.45	1.14	-0.39	-0.75	.95
German ASC *t*1[a]	271	3.44	0.78	-0.03	-0.13	.90
Math Competence *t*1[b]	269	28.71	5.83	-0.44	0.08	.82
German Competence *t*1[c]	271	0.00	0.80	0.15	-0.28	-
Reading comprehension *t*1[d]	270	16.83	6.51	0.24	0.32	-
Orthography *t*1[e]	271	27.42	4.92	-0.35	-0.27	.70
Math Grade *t*1[f]	267	9.15	3.29	-0.01	-1.03	-
Math Grade *t*2[f]	235	9.08	2.95	0.17	-0.83	-
German Grade *t*1[f]	267	8.43	2.66	0.34	-0.50	-
German Grade *t*2[g]	235	8.80	2.59	0.33	-0.46	-

Note. ASC = ability self-concept; [a]Range 1 to 5; [b]Range 0 to 42; [c]Range -1.94 to 2.14; [d]Range -1 to 37; [e]Range 12 to 38; [f]Range 2 to 14; [g]Range 4 to 14.

3.3 Structural Equation Models

Path coefficients, fit indices, and explained variance for the math and the German models are reported in Figure 1 and Figure 2, respectively. Fit indices indicated a very good fit for the math model (χ^2 (74) = 81.608, p = .255; RMSEA = .019 [.000, .040], SRMR = .029, CFI = .995) and a good fit for the German model (χ^2 (47) = 66.031, p = .035; RMSEA = .038 [.011, .058], SRMR = .027, CFI = .980). Both models explained a substantial amount of variance in the respective *t*2 grade scores (math: R^2 = .52; German: R^2 = .41). In both models, the path coefficient of *t*1 ASC on *t*2 grades was positive and significant (math: β = .32, p < .001; German: β = .19, p = .028). The path coefficients of the *t*1 competence scores on *t*2 grades were both positive as well but reached significance only in the math model (β = .18, p = .009). While the respective path coefficient for German (β = .22, p = .139) was larger than the one for math, it failed to reach significance because of the larger standard error (math: *SE*

= .068; German: SE = .149). Since all path coefficients of ASCs and competence scores on $t2$ grades were positive, no SE bias effect after controlling for t1 grades, gender, parental educational level, and number of books was found, and therefore, no significance test of an SE bias effect was computed. Additionally, the path coefficients of $t1$ grades on $t2$ grades were significant in both models (both models: $\beta = .35$, $p < .001$), while the path coefficients of gender on $t2$ grades were significant in neither model (math: $\beta = -.04$, $p = .371$; German: $\beta = .05$, $p = .371$). Likewise, neither the path coefficients of parental educational level (math: $\beta = -.04$, $p = .348$; German: $\beta = -.02$, $p = .697$) nor number of books (math: $\beta = .09$, $p = .071$; German: $\beta = .03$, $p = .578$) on $t2$ grades were significant.

3.4 Regression on algebraic difference scores and residual scores

For detailed results on algebraic difference and residual scores, see OSM S7. When computing SE bias as an algebraic difference score, we did not obtain significant regression coefficients of SE bias on $t2$ grades in either domain (math: $\beta = .07$, $p = .200$; German: $\beta = .02$, $p = .677$). On the other hand, when computing SE bias as residual scores, we obtained significant regression coefficients for the SE bias in both domains (math: $\beta = .22$, $p < .001$; German: $\beta = .16$, $p = .005$). However, it should be noted that these results are not considered valid indicators of true SE bias effects, but are used for a comparison between traditional and more advanced methods, as will be discussed in detail in the discussion section.

Table 2

Correlation matrix of the analysis variables

	2	3	4	5	6	7	8	9	10	11	12	13
1 Math ASC *t*1	.37**	.56**	.57**	-.25**	-.01	.00	-.03	.07	.11	.23**	-.05	-.00
2 Math Competence *t*1	-	.44**	.41**	-.02	.39**	.32**	.30**	.36**	.23**	.10	.03	.18**
3 Math Grade *t*1		-	.62**	-.06	.21**	.13*	.22*	.43**	.33**	.01	-.05	.01
4 Math Grade *t*2			-	-.06	.15*	.08	.19**	.29**	.34**	.06	-.04	.11
5 German ASC *t*1				-	.33**	.25**	.28**	.48**	.46**	-.29**	.14*	.20**
6 German Competence *t*1					-	.80**	.80**	.41**	.38**	-.17*	.07	.21**
7 Reading comprehension *t*1						-	.27**	.34**	.29**	-.05	.08	.19**
8 Orthography *t*1							-	.32**	.33**	-.21**	.03	.14*
9 German Grade *t*1								-	.56**	-.27**	.14*	.22**
10 German Grade *t*2									-	-.17**	.09	.19**
11 Gender										-	-.04	-.10
12 Parental Educational Level											-	.33**
13 Number of books at home												-

Note. N = 228 – 271; ASC = ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur; 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100; 2 = more than 100; *p < .05, **p < .01.

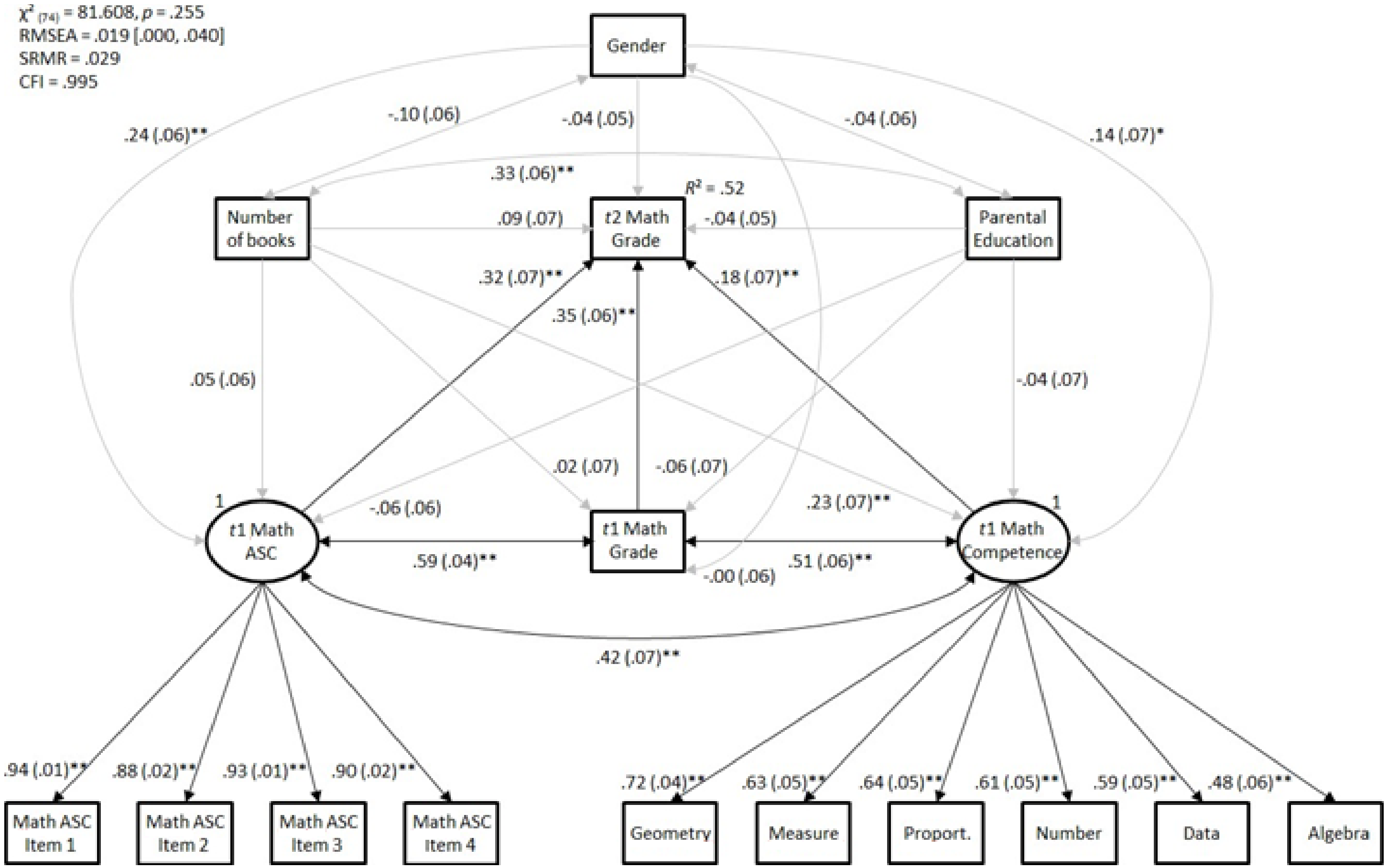


Figure 1. Structural equation model of the effects of math ability self-concept (ASC), competence, grades, gender, parental education, and number of books at home at measurement occasion 1 (*t*1) on grades at measurement occasion 2 (*t*2). Proport. = Proportionality; Gender: 1 = female, 2 = male; Parental Education: 1 = No parent with Abitur or Fachabitur; 2 = At least one parent with Abitur or Fachabitur; Number of books: 1 = up to 100; 2 = more than 100; Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 284$. Standardized solution (standard errors in parentheses). Gray lines = paths involving covariates. *($p < .05$), **($p < .01$).

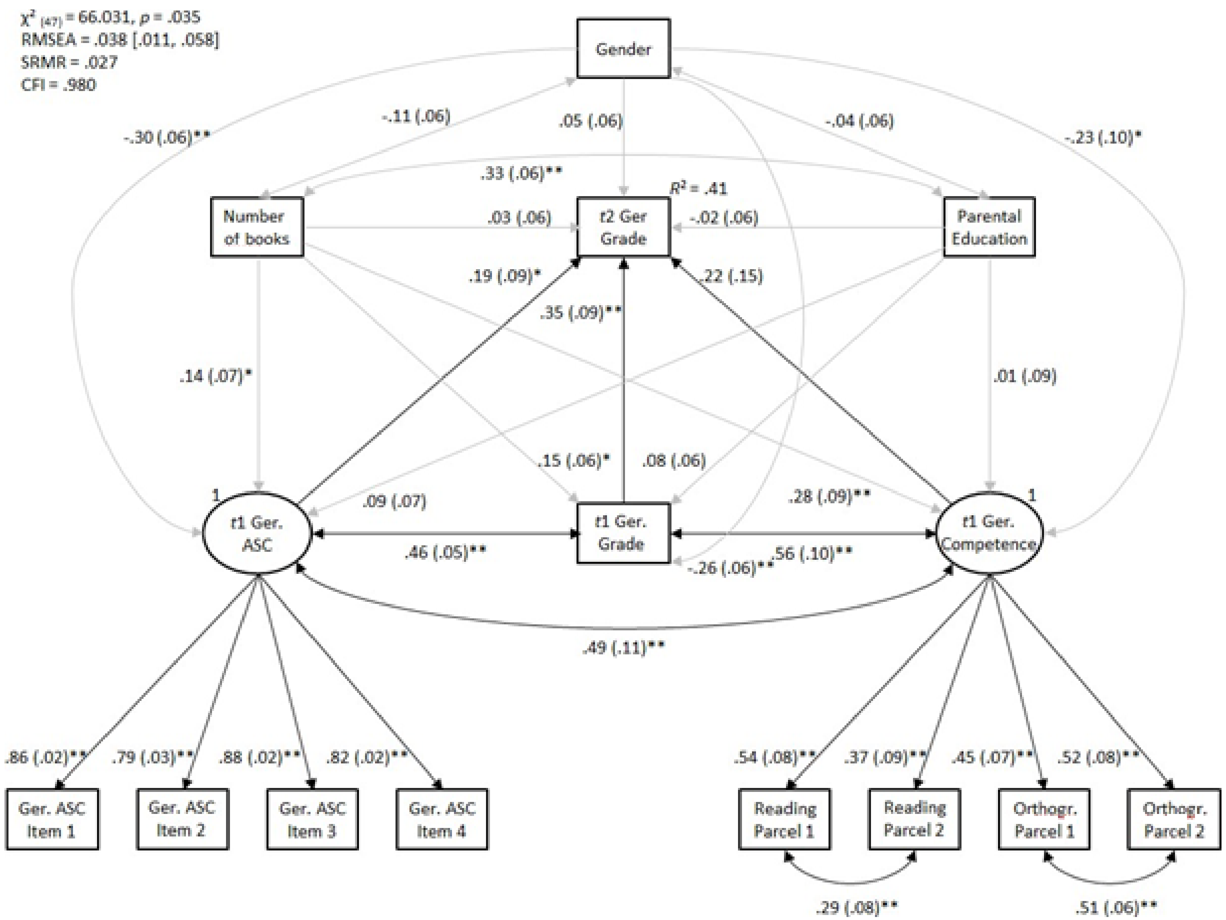


Figure 2. Structural equation model of the effects of German ability self-concept (ASC), competence, grades, gender, parental education, and number of books at home at measurement occasion 1 (*t*1) on grades at measurement occasion 2 (*t*2). Orthogr. = Orthography; Gender: 1 = female, 2 = male; Parental Education: 1 = No parent with Abitur or Fachabitur; 2 = At least one parent with Abitur or Fachabitur; Number of books: 1 = up to 100; 2 = more than 100; Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 284$. Standardized solution (standard errors in parentheses). Gray lines = paths involving covariates. *($p < .05$), **($p < .01$).

4. Discussion

The primary goal of the present study was to separate and analyze the effects of self-view and SE bias on change in school grades. We have done so by computing structural equation models in which *t*1 ASC and competence as well as *t*1 grades, gender, and socioeconomic variables predicted *t*2 grades. Thus, we were able to analyze, how ASC and competence affected grades when the influence of former grades, gender, and socioeconomic variables are controlled.

The results showed positive self-view effects after controlling for *t*1 grades, gender and socioeconomic variables on *t*2 grades in the domains of math and German, therefore supporting Hypotheses 1a and 1b. Thus, we found further evidence for the notion of the reciprocal effects model that ASCs positively affect subsequent grades in the same domain (see Marsh & Martin, 2011). This result is in line with current knowledge and meta-analytic evidence on effects of ASCs on various academic achievement measures including grades (Trautwein & Möller, 2016; Valentine et al., 2004).

By contrast, we did not find any support for the existence of SE bias effects on grades after controlling for *t*1 grades, gender, and socioeconomic variables (RQ 1a and 1b). Taken together, our results suggest that it is indeed the self-view per se rather than the SE bias that affected change in students' grades.

4.1 Self-Estimation Bias, Self-View, and Grades

As has been mentioned above, other studies, in which SE bias was assessed prior to a measure of academic achievement, did mostly find positive SE bias effects (e.g., Bonneville-Roussy et al., 2017; Côte et al., 2014; Dupeyrat et al., 2011; Leduc & Bouffard, 2017). This discrepancy between our and prior results can be interpreted in light of the different methodology. It is a particular strength of our study that we were able to disentangle the effect of SE bias from the effect of the self-view alone. In prior studies, methods were used that confounded these constructs. Since self-view effects on grades are well-documented, it is likely that positive SE bias effects found in the literature are, at least to some degree, artifacts of self-view effects. In the present study, one of the two traditional methods of calculating SE bias effects, the use of residual scores, resulted in significant SE bias effects in both domains after controlling for *t*1 grades, gender, and socioeconomic variables, while the other method, algebraic difference scores, did not. This discrepancy between results of the two traditional methods might be a consequence of the fact that residual scores compared to algebraic difference scores are usually more strongly confounded with self-view scores. For example, if the correlation between a self-view and a reality criterion is $r = .30$ and both variables are standardized, the correlation between the residual score and the self-view is $r = .95$, whereas the correlation between the algebraic difference score and the self-view is "only" $r = .59$ (Humberg et al., 2018, p. 309). Because we observed significant self-view effects after controlling for *t*1 grades, gender and socioeconomic variables in the present analyses, it is plausible to assume that this stronger confounding led to residual scores being better predictors of grade changes than algebraic difference scores. Since the CRA is advantageous over traditional methods in disentangling SE bias effects from self-view effects, our findings indicate that it is not the discrepancy between self-view and competence, but rather the self-view per se that contributes to change in grades after controlling for socioeconomic variables and gender.

The effect sizes of self-view effects on grade changes in the present study are comparatively large (Valentine et al., 2004), especially in math. We theorize that positive self-view effects found in the present study might have been mediated by motivation and adaptive academic behavior, such as effort, task choice, and persistence (Areepattamannil, 2012; Helmke, 1990; 1998; Trautwein et al., 2009; Trautwein & Möller, 2016; Wigfield & Eccles, 2000), within the interval between the measurement time points. However, further studies are needed to test this hypothesis.

4.2 Cross- and Within-domain correlations

The results of our correlational analyses for within-domain correlations between ASC, competence test scores, and grades closely mirror those found in a meta-analysis (Möller et al., 2009). For example, our correlations between ASC and competence test scores (math: r = .37; German: r = .33) are virtually identical to the meta-analytical averages (math: r = .37; German: r = .34; Möller et al., 2009, p. 23). This strengthens the confidence in the generalizability of our findings since our sample does not seem to be unusual in terms of the relations between ASC, competence, and grades within the same domain.

However, our cross-domain correlations between grades and test scores, respectively, are substantially lower than the one's found by Möller et al. (2009). We also observed a significant weak and negative correlation between ASCs across domains. This result was not surprising given that correlations between math and verbal ASCs are usually close to zero (Möller et al., 2009) and have even been found to be negative in some German studies (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2006; Steinmayr & Spinath, 2009), which is in line with our findings.

4.3 Limitations and implications for future research

By using an approach that separated self-view effects from SE bias effects, we were able to show that it is the self-view alone, not the SE bias, that predicted grades after controlling for former grades gender, and socioeconomic variables. We recommend that future studies on SE bias effects should likewise make use of the CRA or similar approaches that disentangle these effects in order to avoid interpreting actual self-view effects as SE bias effects. However, the present study also has important limitations.

First, we analyzed a highly homogenous sample of 10th grade students from two academic schools. The restriction to relatively old students is appropriate because self-view effects might be more pronounced in older students (Chen et al., 2013; Skaalvik & Hagtvet, 1990; Weidinger et al., 2018). Since the present study is the first to disentangle self-view and SE bias effects on academic achievement, we analyzed a sample in which the chance of detecting these effects would supposedly be highest. On the other hand, the academic level of the school does not seem to influence self-view effects in students (Preckel et al., 2017; Seaton et al., 2015), which increases the confidence in the generalizability of the results to other school forms. Still, future studies should analyze whether the results of the present study are generalizable to younger students and students in less academic schools.

Second, we only tested for linear SE bias effects. However, there is no inherent reason to assume that decreasing underestimation must have the same effect as increasing overestimation. In other words, the effects of SE bias might not be linear but vary between different levels of SE bias. Indeed, if proponents of benefits of realistic self-estimations are correct (e.g., Dunlosky & Rawson, 2012; Dunning et al., 2003; Hacker & Bol, 2019) this would imply a nonlinear relation between SE bias and academic achievement with maximum academic achievement at an SE bias score of 0, a positive slope for negative SE bias values, and a negative slope for positive SE bias values (self-knowledge hypothesis; Humberg et al., 2019). In future studies, in order to test whether there are indeed nonlinear relations between SE bias and academic achievement, bivariate spline regressions (Edwards & Parry, 2018) or the regression models derived by Humberg et al. (2019) for various hypotheses about the relations between self-views, reality criteria, and certain outcome variables can be used.

Third, we did not analyze the possible mechanisms by which self-view and SE bias can affect academic achievement suggested in the literature. Possible relevant mediator variables are for example effort, task choice, task persistence, and motivation (Areepattamannil, 2012; Helmke, 1990; 1998; Trautwein et al., 2009; Trautwein & Möller, 2016; Wigfield, & Eccles, 2000).

Fourth, with ASCs we measured only one of several self-view constructs. Future studies should seek to answer whether using different constructs (e.g., task specific self- efficacy) leads to different results regarding SE bias effects. However, evidence suggests that varying between different self-view constructs should not have a large impact on effect sizes (Valentine et al., 2004).

Fifth, our aim was to independently test the effects of self-view and SE bias on academic achievement. However, since our study had a predictive, not an experimental design, we were not able to directly test causal relations. This is a limitation shared by virtually any study on SE bias effects to date, which we tried to mitigate by including a set of covariates which could also be responsible for the observed path coefficients of ASC and competence on subsequent grades. Thus, controlling for these covariates strengthens the confidence in the interpretation of path coefficients from ASC and competence on subsequent grades as effects.

It should be noted that most of these suggestions have already been applied to studies on either self-view or SE bias effects. For example, the mediating role of effort between ASCs and academic achievement has been studied (Helmke, 1990; Trautwein et al., 2009). However, if these approaches were combined with a method that disentangles self-view effects from SE bias effects, it would offer further insight into the nature of both self-view and possible SE bias effects. In the present study, we did not perform the recommended additional analyses because the CRA is a highly novel approach that, to our knowledge, has never been used to analyze SE bias effects on grades. Thus, we first aimed to test SE bias effects in general before turning to subgroup and mediator analyses.

4.4 Conclusions and practical implications

The present study indicates that students' ability self-views are more important for improved future grades than the discrepancy between the self-view and actual competence (i.e., SE bias). Notably, both math and German ASCs were significant predictors of grades within the same domain even when prior grades, competence, gender, and socioeconomic variables were controlled. These results support the idea of encouraging positive ASCs in students, which has been the aim of a large number of intervention studies (for a meta- analysis see O'Mara, Marsh, Craven, & Debus, 2006). However, further studies are needed to replicate results and clarify whether SE bias effects and self-view effects vary between different groups. Most importantly, it should be investigated whether these effects vary between students with different levels of SE bias (e.g., between overestimators and underestimators).

References

Areepattamannil, S. (2012). Mediational role of academic motivation in the association between school self-concept and school achievement among Indian adolescents in Canada and India. *Social Psychology of Education: An International Journal, 15*(3), 367–386. doi:10.1007/s11218-012-9187-1

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*(5), 621–634. doi:10.1037/edu0000163

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966. doi:10.1037//0022-3514.74.4.955

Baumert, J., Lehmann, R., Lehrke, M., Clausen, M., Hosenfeld, I., Neubrand, J. et al. (1998). *Testaufgaben Mathematik TIMSS 7./8. Klasse (Population 2) [Test items mathematics TIMSS 7./8. grade (population 2)].* Berlin: Max-Planck-Institut für Bildungsforschung. Retrieved on from https://pure.mpg.de/rest/items/item_2103201/component/file_2103200/content

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles*, *72*(11-12), 536-546. doi:10.1007/s11199- 015-0486-9

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education*, *73*(4), 269-290. doi:10.3200/JEXE.73.4.269-290

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research, 105*(5), 336–352. doi:10.1080/00220671.2011.627401

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. doi:10.1023/A:1021302408382

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. doi:10.1016/j.jsp.2017.02.002

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H, Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. doi:10.3102/0034654316672069

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*(2), 134-138. doi:10.1177/0098628312475034

Burns, R. A., Crisp, D. A., & Burns, R. B. (2019). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the cross-lagged panel and random-intercept cross-lagged panel frameworks. *British Journal of Educational Psychology*. doi:10.1111/bjep.12265

Calysn, R. J., & Kenny, D. A. (1977). Self-concept of ability and perceived evaluation of others: Cause or effect of academic achievements? *Journal of Educational Psychology, 69*, 136–145. doi: 10.1037/0022-0663.69.2.136

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. doi:10.1016/j.lindif.2012.07.02

Chiu, M. M., & Klassen, R. M. (2009). Calibration of reading self-concept and reading achievement among 15-year-olds: Cultural differences in 34 countries. *Learning and Individual Differences*, *19*(3), 372-386. doi:10.1016/j.lindif.2008.10.004

Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction*, *20*(1), 2-17. doi:10.1016/j.learninstruc.2008.11.002

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: A longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*(10), 1384-1401. doi:10.1177/0146167216662866

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159–179. doi:10.1207/s15326977ea0302_3

Côté, S., Bouffard, T., & Vezeau, C. (2014). The mediating effect of self- evaluation bias of competence on the relationship between parental emotional support and children's academic functioning. *British Journal of Educational Psychology, 84*(3), 415–434. doi:10.1111/bjep.12045

Dufner, M., Reiz, A. K., & Zander, L. (2015). Antecedents, consequences, and mechanisms: On the longitudinal interplay between academic self- enhancement and psychological adjustment. *Journal of Personality*, *83*(5), 511-522. doi:10.1111/jopy.12128

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. doi:10.1016/j.learninstruc.2011.08.003

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. doi:10.1111/j.1529-1006.2004.00018.x

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current directions in psychological science, 12*(3), 83-87. doi:10.1111/1467-8721.01235

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250. doi:10.1016/j.ijer.2011.08.005

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613. doi:10.2307/256822

Edwards, J. R., & Parry, M. E. (2018). On the use of spline regression in the study of congruence in organizational research. *Organizational Research Methods, 21*(1), 68-110. doi:10.1177/1094428117715067

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press. Epley, N., & Dunning, D. (2006). The Mixed Blessings of Self-Knowledge in Behavioral Prediction: Enhanced Discrimination but Exacerbated Bias. *Personality and Social Psychology Bulletin, 32*(5), 641–655. doi:10.1177/0146167205284007

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6*, 505-518. doi:10.1111/j.1475-6811.1999.tb00206.x

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*(6), 644–653. doi:10.1016/j.lindif.2010.08.001

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice, 8*(3), 381–391. doi:10.1080/135406002100000512

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education.* (pp. 647–677). New York, NY: Cambridge University Press. doi:10.1017/9781108235631.026

Helmke, A. (1990). Mediating processes between children's self-concept of ability and mathematical achievement: A longitudinal study. In H. Mandl, E. de Corte, S. N. Bennett, & H. F. Friedrich (Eds.), *Learning and instruction. European research in an international context. Volume 2.2: Analysis of complex skills and complex knowledge domains* (pp. 537–549). Oxford: Pergamon Press.

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe [From an optimist to a realist? On the development of the ability self-concept from kindergarten to 6th grade]. In F. E. Weinert (Ed.), *Entwicklung im Kindesalter* (pp. 115-132). Weinheim, DE: Beltz.

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*, 505–528. doi: 10.1016/j.jsp.2011.07.001

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835-859. doi:10.1037/pspp0000204

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology, 114*(2), 303–322. doi:10.1037/pspp0000134

Hußmann, A., Stubbe, T. C., & Kasper, D. (2017). Kapitel VI. Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, & R. Valtin (Eds.), *IGLU 2016. Lesekompetenz von Grundschulkindern in Deutschland im* *internationalen Vergleich [IGLU 2016. Reading competence of grade schoolers in Germany by international comparison].* (S. 195-218). Münster, DE: Waxmann.

Kalogiannis, P., Papaioannou, A., Sagovich, A., & Abatzoglou, G. (2011). Reciprocal effects between self-concept and school performance, preparation for school, and life satisfaction: A longitudinal study. *Hellenic Journal of Psychology, 8*(1), 96–122.

Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education, 81*(1), 32–52. doi:10.1177/003804070808100102

Kersting, M., & Althoff, K. (2004). *RT Rechtschreibungstest [RT orthography test].* Göttingen, DE: Hogrefe.

Leduc, C., & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences, 55*, 193–201. doi:10.1016/j.lindif.2017.04.006

Lee, C.-Y., & Kung, H.-Y. (2018). Math self-concept and mathematics achievement: Examining gender variation and reciprocal relations among junior high school students in Taiwan. *Eurasia Journal of Mathematics, Science & Technology Education, 14*(4), 1239– 1252.

Little, T. D. (2013). *Longitudinal structural equation modeling.* New York, NY: Guilford Press.

Maaz, K., Baeriswyl, F., & Trautwein, U. (2013). II. Studie: „Herkunft zensiert?“ Leistungsdiagnostik und soziale Ungleichheiten in der Schule. In Deißner, D. (Ed.), *Chancen bilden [Building chances].* (pp. 185–334). Wiesbaden, DE: Springer.

Marsh, H. W., Hau, K.-T., & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39*(3), 727–763. doi:10.3102/00028312039003727

Marsh, H. W. & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77. doi: 10.1348/000709910X503501

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of Multidimensional Self-Concept and Core Personality Constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, *74*(2), 403–456. doi:10.1111/j.1467-6494.2005.00380.x

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20–32. doi:10.1111/j.1745-3992.2001.tb00055.x

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*(3), 1129-1167. doi:10.3102/0034654309337522

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*(4), 591–611. doi:10.1111/bjep.12047

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research, 95*(3), 131–142. doi:10.1080/00220670209596583

OECD (2016). *PISA 2015 Ergebnisse (Band 1): Exzellenz und Chancengerechtigkeit in der Bildung [PISA 2015 Results (Volume 1): Excellence and Equity in Education].* Bielefeld, DE: Bertelsmann.

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*(3), 181–206. doi:10.1207/s15326985ep4103_4

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. doi:10.1177/0016986216687824

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*(2), 340-352. doi:10.1037/0022-3514.80.2.340

Rohr, M. E., & Ayers, J. B. (1973). Relationship of student grade expectations, selected characteristics, and academic performance. *Journal of Experimental Education*, *41*(3), 58-62. doi:10.1080/00220973.1973.11011410

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits-und -verständnistest für die Klassen 6–12 [LGVT 6-12: reading speed and comprehension test for classes 6-12].* Göttingen, DE: Hogrefe.

Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzeptes SESSKO [Scales for the assessment of the school self-concept SESSKO].* Göttingen, DE: Hogrefe.

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly, 59*(3), 143–156. doi:10.1177/0016986215583870

Skaalvik, E. M., & Hagtvet, K. A. (1990). Academic achievement and self-concept: An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology, 58*(2), 292–307. doi:10.1037/0022-3514.58.2.292

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles: A Journal of Research, 50*(3–4), 241–252. doi:10.1023/B:SERS.0000015555.40976.e6

Spinath, B., & Steinmayr, R. (2012). The role of competence beliefs and goal orientations for change in intrinsic motivation. *Journal of Educational Psychology, 104*(4), 1135-1148. doi:10.1037/a0028115

Statistisches Bundesamt. (2016). *Bildung der Eltern beeinflusst die Schulwahl für Kinder [Parental education influences the choice of school for children]*. Retrieved from https://www.destatis.de/DE/Presse/Pressemitteilungen/2016/09/PD16_312_122pdf.pdf?__blob=publicationFile

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*(4), 273–282. doi:10.1024/1010-0652/a000113

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80–90. doi:10.1016/j.lindif.2008.05.004

Steinmayr, R., Weidinger, A. F., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology, 53*, 106-122. doi:10.1016/j.cedpsych.2018.02.004

Stubbe, T. C., Schwippert, K., & Wendt, H. (2016). Kapitel X. Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im* *internationalen Vergleich [TIMSS 2015. Mathematical and science competences of grade schoolers in Germany by international comparison].* (S. 299-316). Münster, DE: Waxmann.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193-210. doi:10.1037/0033-2909.103.2.193

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6*(4), 662–667. doi:10.3758/BF03212976

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. doi:10.1037/0022-0663.95.1.66

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97*(6), 1115–1128. doi:10.1037/a0017048

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187-214). Cham, CH: Springer International Publishing.

Usher, E. L., Li, C. R., Butz, A. R., & Rojas, J. P. (2019). Perseverant grit and self-efficacy: Are both essential for children's academic success? *Journal of Educational Psychology, 111(5)*, 877–902. doi:10.1037/edu0000324

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111-133. doi:10.1207/s15326985ep3902_3

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. do: 10.1002/per.2137

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development*, *89*(2), 138–156. doi:10.1111/cdev.12806

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. (pp. 209–231). New York, NY: The Guilford Press.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. doi:10.1006/ceps.1999.1015

Wilgenbuscg, T., & Merrel, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly*, *14*(2), 101–120. doi:10.1037/h0089000

Supplemental Material

Overview

Supplement 1. Missing data

Supplement 2. Computation of confirmatory factor analyses and expected cross-loadings

Supplement 3. Results of confirmatory factor analyses and for expected cross-loadings

Supplement 4. Computation of double-entry intraclass correlations

Supplement 5. Results and discussion of double-entry intraclass correlations

Supplement 6. Computation of SE bias effects based on algebraic difference scores and residual scores

Supplement 7. Results of SE bias effects based on algebraic difference scores and residual scores

Supplement 8. R and SPSS syntaxes

Supplement 1. Missing data

Missing data in designs with repeated measurements can be categorized as resulting from nonresponse or from attrition (Little, 2013). Missing data due to nonresponses amounted to 1.0% for the individual math ASC items and to 0.9% for the individual German ASC items. Because the competence measures are performance tests with a speed component, missing data were analyzed on the scale level rather than on the item level. For both the TIMSS scores and the LGVT reading comprehension, 0.7% of the data were missing. There were no missing data of the RT orthography score. In regard to missing data due to attrition, 74.6% of the participants took part in both measurement occasions. At *t*1, 6.0% of the grade scores were missing, e.g., because students changed schools between Grades 9 and 10. At *t*2, 17.3% of the grade scores were missing. To account for the missing data, full information maximum likelihood estimation (FIML) was used in all structural equation models (Little, 2013). Additionally, we followed the recommendations by Enders (2010) by incorporating a set of auxiliary variables into the FIML estimation. Consequently, we used the auxiliary function from R's semTools package v0.4-14 which is based on the extra-dependent-variables and saturated-correlates approaches by Enders (2008). The auxiliary variables used were general test anxiety (all models), math test anxiety (math models) and German test anxiety (German models).

Supplement 2. Computation of confirmatory factor analyses and expected cross-loadings

In order to test the structural validity of the measures we computed confirmatory factor analyses and inspected the item's expected cross-loadings using the package lavaan in R version 3.4.3 For the two ASC scales we computed a single two-factor CFA with correlated factors and the individual ASC items as indicators. We also inspected the expected cross-loadings of the items on the respective other scale. Because we used parcels of the TIMSS items in our structural equation models, we computed a CFA with overall math competence as the latent variable and the six TIMSS parcels as the indicators. In our structural equation models, we used two parcels each for the LGVT and the RT. This was sufficient because the parcels were embedded in a larger model. However, in order to test the structural validity of the individual instruments, a larger number of indicators are needed to ensure that the degrees of freedom are greater than 0. Thus, we computed four parcels per instrument and used these parcels as indicators in the CFAs.

Supplement 3. Results of confirmatory factor analyses and for expected cross-loadings

The CFA of the two ASC scales showed a good model fit according to cutoff criteria for the CFI and SRMR but not the RMSEA (χ^2 (19) = 47.793, p < .001; RMSEA = .076 [.050, .104], SRMR = .031, CFI = .984). However, an inspection of the standardized factor loadings (.79 < λ < .94) and standardized expected cross-loadings (-.08 < λ < .06) clearly indicated that each item loaded only on its intended scale (see Figure S2.1). The CFA of the TIMSS showed a good model fit (χ^2 (9) = 14.254, p = .114; RMSEA = .046 [.000, .090], SRMR = .033, CFI = .985) and all factor loadings were in an acceptable range (.47 < λ < .72; see Figure S2.2). The same was true for the CFA of the RT (χ^2 (2) = 3.673, p = .159; RMSEA = .056 [.000, .144], SRMR = .023, CFI = .991, .60 < λ < .65; see Figure S.2.3) and the LGVT (χ^2 (2) = 2.803, p = .246; RMSEA = .039 [.000, .133], SRMR = .024, CFI = .992, .37 < λ < .61) with the exception of one slightly weaker factor loading of the LGVT parcel 2 (λ = .37; see Figure S.2.4). Overall, all instruments showed good structural validity.

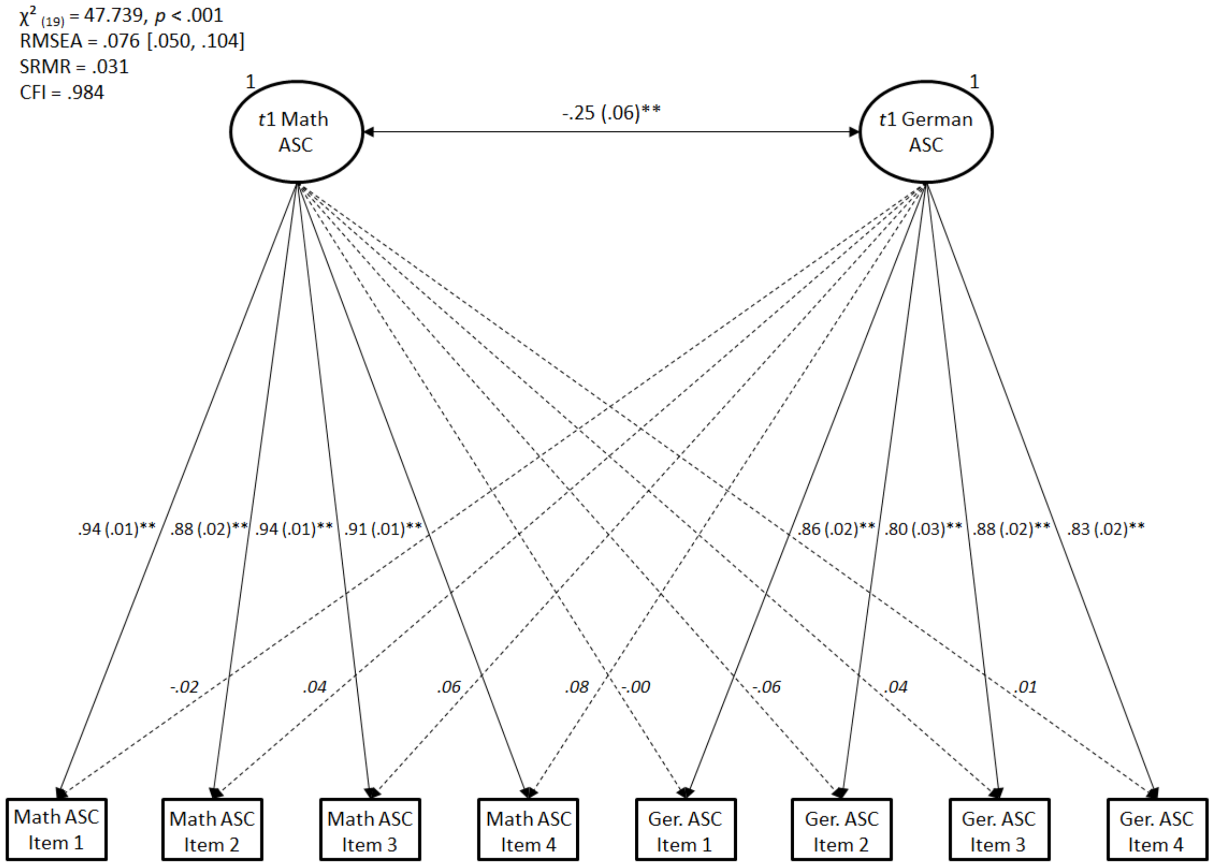


Figure S3.1. Confirmatory factor analysis of the ability self-concept scales in math and German. ASC = ability self-concept. Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 259$. **($p < .01$).

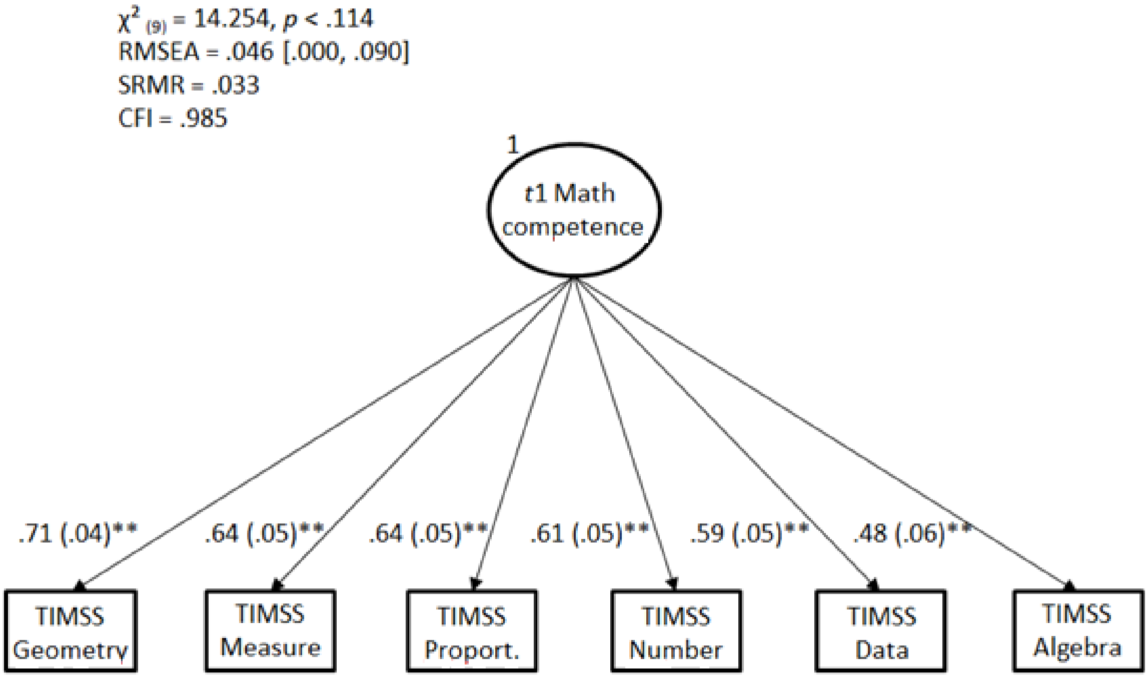


Figure S3.2. Confirmatory factor analysis of the TIMSS items. Proport. = Proportionality. Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 270$. **($p < .01$).

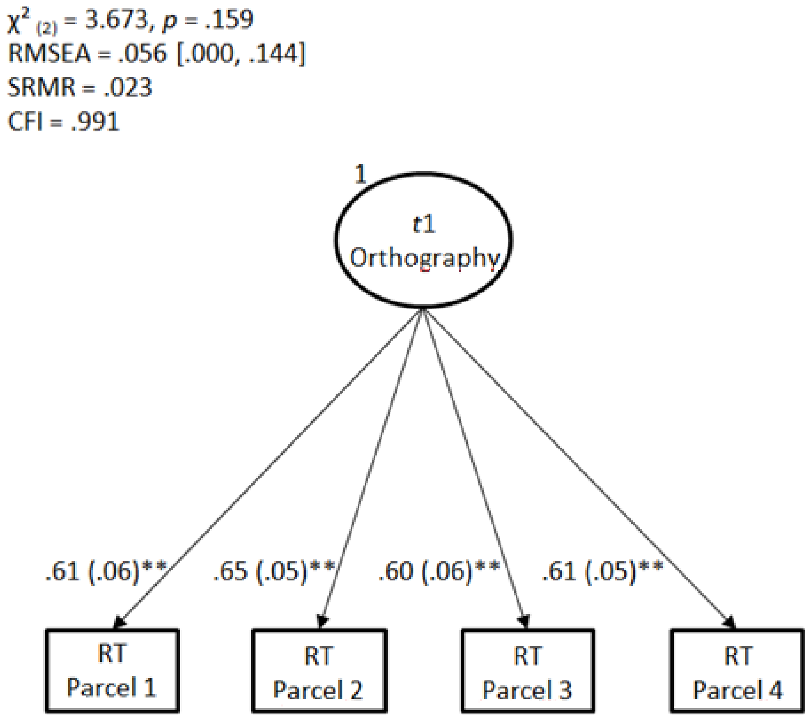


Figure S3.3. Confirmatory factor analysis of the RT. Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 271$. **($p < .01$).

$\chi^2_{(2)} = 2.803$, $p = .246$
RMSEA = .039 [.000, .133]
SRMR = .024
CFI = .992

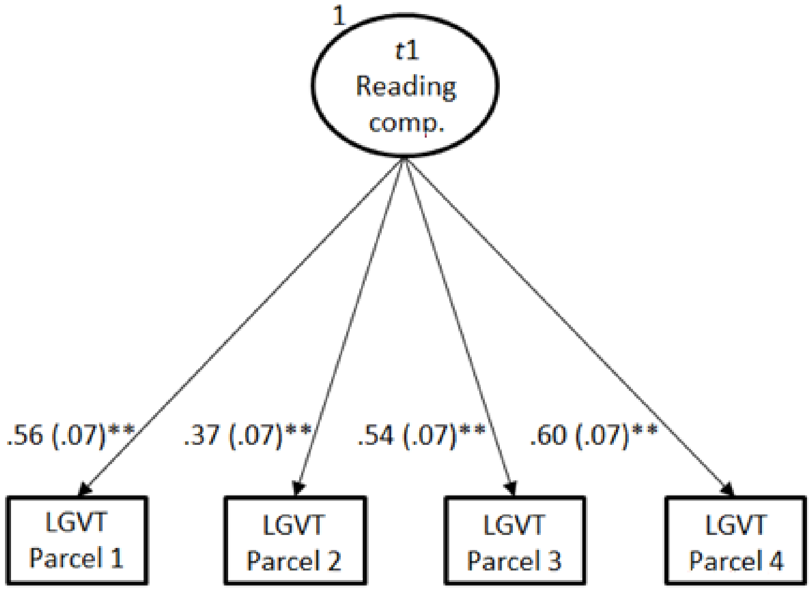


Figure S3.4. Confirmatory factor analysis of the LGVT 6-12. Reading comp. = Reading comprehension. Error terms have been estimated but are omitted in the graphical representation for visual clarity. $N = 270$. **($p < .01$).

Supplement 4. Computation of double-entry intraclass correlations

In order to test whether the inclusion of the covariates (gender, parental educational level, and number of books at home) in our models substantially altered the relation between the ASC and competence scores with related variables, we computed double-entry intraclass correlations (r_{ICC}; Vize, Collison, Miller, & Lynam, 2018). First, we computed two different residual scores for each combination of domain (math or German) and construct (ASC or competence) resulting in a total of 8 residual scores. In each of these domain-construct combinations one of the residual scores (“Residual 1”) was computed by partialling out the respective other construct (ASC from competence or competence from ASC) as well as the *t*1 grades in the same domain. The other residual score (“Residual 2”) was computed by partialling out the same variables, but also the covariates. Second, the eight obtained residual scores were then correlated with other constructs that have been assessed in the research project the data stem from. For a list of these constructs and the instruments used to assess them, see Table S4. Thus, by comparing the correlations of the two different residual scores of the same domain-construct combination, it was possible to estimate the similarity or difference between the relations of the two residual scores with related variables. Thus, we could estimate whether the nomological networks in which the two residual scores are embedded are similar. Since the only difference between the two residual scores of the same domain-construct combination is whether the covariates have been partialled out, this is an estimation on whether partialling out the covariates substantially changed the nomological networks of the constructs beyond partialling out the respective other construct and the *t*1 grade scores. If this were the case, and the correlations of the different residual scores within the same domain-construct combination showed large differences, this would be a warning sign for the interpretability of the ASC and competence effects in our models. In that case, depending on the severity of the differences, we might have to exclude the covariates from our analyses. We chose to compare the correlations of two different residuals instead of the correlations of a residual score and the unresidualized ASC and competence scores because 1) due to the logic of the CRA it was necessary to include both ASC and competence in the same analyses in our models and 2) due to the high correlation between ASCs and grades, we also considered it necessary to include $t1$ grades as predictors in our models. Thus, what we wanted to test was whether the inclusion of the covariates beyond these necessary predictors was feasible. In order to obtain a single score that quantifies the extent to which the correlations of the two residuals within the same domain-construct combination are similar or not, we computed the double-entry intraclass correlations for the correlations between the residual scores and the related variables.

Table S4

Constructs correlated with ASC and competence residuals and their measurement instruments

Construct	Measurement instrument
Neuroticism Extraversion Openness Agreeableness Conscientiousness	NEO five factor inventory (NEO-FFI; Borkenau & Ostendorf, 2008)
Math/German intrinsic value Math/German utility value Math/German attainment value Math/German expectancy of success	Scale assessing subjective educational task values (SESSW; Steinmayr & Spinath, 2010)
Math/German agitation Math/German worry	Short version (Schwarzer & Jerusalem, 1999) of the German test anxiety inventory (TAI-G; Hodapp,1991,1996)
Practical/technical orientation Intellectual/researching orientation Artistic/linguistic orientation Social orientation Business orientation Conventional orientation	General interest structure test with environment structure test (AIST-R/UIST-R; Bergmann & Eder, 2005)

Supplement 5. Results and discussion of double-entry intraclass correlations

An inspection of the differences between the correlations of different residual scores (i.e., Residual 1 and Residual 2) within the same domain-construct combination revealed that these differences were small (all $\Delta r \leq .15$) and all double-entry intraclass correlations between those correlations were very high and significant (all $r_{ICC} \geq .94$, all $p < .001$). Detailed results are reported in Table S5. Therefore, the inclusion of the covariates seemed unproblematic and likely did not substantially affect the interpretability of the ASC and competence effects in our models. It is noteworthy, that in both domains the competence residuals showed less significant and smaller correlation than the ASC residuals. This result is not unexpected given that all variables correlated with the residuals were self-report measures like ASC while competence was assessed with performance tests.

Table S5

Correlations between residualized ASC and competence scores and related variables and their double-entry intraclass correlations (r_{ICC})

	Math ability self-concept		Math competence		German ability self-concept		German competence	
	Residual 1	Residual 2	Residual 1	Residual 2	Residual 1	Residual 2	Residual 1	Residual 2
Neuroticism	-.18**	-.12	-.02	-.03	-.04	-.07	.01	-.02
Extraversion	-.09	-.08	-.18**	-.14*	.10	.09	-.24***	-.22**
Openness	-.07	-.05	.20**	.16**	.16*	.15*	.11	.09
Agreeableness	-.17**	-.09	-.08	-.03	.12	.07	-.16**	-.16**
Conscientiousness	.11	.16*	-.09	-.05	.12	.12	-.08	-.04
Math/German intrinsic value	.57***	.56***	-.01	.00	.54***	.49***	-.07	-.08
Math/German utility value	.28***	.24***	.04	.07	.30***	.29***	-.05	-.05
Math/German attainment value	.40***	.39***	-.01	.03	.38***	.39***	-.09	-.06
Math/German expectancy of success	.62***	.59***	-.04	-.03	.48***	.50***	-.01	.01
Math/German agitation	-.36***	-.34***	-.10	-.11	-.33***	-.31***	.02	.03
Math/German worry	-.36***	-.35***	-.10	-.11	-.26***	-.23***	.01	.02
Practical/technical orientation	.28***	.13*	.12	.05	-.14*	-.05	-.01	-.00
Intellectual/researching orientation	.21**	.11	.18**	.16*	-.06	.02	.04	.06
Artistic/linguistic orientation	-.21**	-.11	.02	.04	.37***	.30***	.00	-.02
Social orientation	-.22***	-.13*	-.10	-.06	.22***	.18**	-.04	-.06

Table S5 - continued

	Math ability self-concept		Math competence		German ability self-concept		German competence	
	Residual 1	Residual 2	Residual 1	Residual 2		Residual 1	Residual 2	Residual 1
Business orientation	-.09	-.07	-.11	-.07	.27***	.28***	-.16**	-.14*
Conventional orientation	.11	.05	-.01	-.00	.12*	.19**	-.09	-.07
Similarity indices (r_{ICC})								
Ability self-concept Resiudal 1	-				-			
Ability self-concept Residual 2	.97***	-			.98***	-		
Competence Resiudal 1	.19	.14	-		-.34	-.36*	-	
Competence Residual 2	.19	.17	.94***	-	-.34*	-.35*	.97***	-

Note. N = 284; Residual 1 = $t1$ grades in the same domain and ability self-concept/competence in the same domain partialled out; Residual 2 = $t1$ grades in the same domain, ability self-concept/competence in the same domain, gender, parental educational level, and number of books partialled out. Intrinsic value, utility value, attainment value, expectancy of success, agitation, and worry: correlations are based on the values in the same domain as the respective residual score.

Supplement 6. Computation of SE bias effects based on algebraic difference scores and residual scores

In order to compute SE bias as algebraic difference scores, we first had to ensure that the ASC scores and the competence scores were located on the same scale. Thus, we standardized both the ASC scores and the competence scores. Then, we subtracted the standardized competence scores from the standardized ASC scores. In order to compute the residual scores, we computed a linear regression of the ASC scores on the competence scores. The residuals of this regression were then treated as the SE bias measure. We then computed two multiple linear regression analyses in each domain, math and German, one with the residual scores and one with the competence scores as predictors of $t2$ grades. In order to ensure comparability with the CRAs, we also included $t1$ grades, gender, parental educational level, and number of books as predictors in all regression analyses.

Supplement 7. Results of SE bias effects based on algebraic difference scores and residual scores

The results of all regression analyses are presented in Table S7. When computing SE bias as an algebraic difference score, we did not obtain significant regression coefficients from SE bias on *t*2 grades in either domain (math: $\beta = .07$, $p = .200$; German: $\beta = .02$, $p = .677$). On the other hand, when computing SE bias as residual scores, we obtained significant regression coefficients for the SE bias in both domains (math: $\beta = .22$, $p < .001$; German: $\beta = .16$, $p = .005$). Additionally, gender ($-.00 \leq \beta \leq .06$, $.225 \leq p \leq .993$) and parental educational level ($-.07 \leq \beta \leq -.02$, $.182 \leq p \leq .687$) did not significantly predict *t*2 grades in any analysis. Number of books did significantly predict *t*2 grades in math (both analyses: $\beta = .16$, $p = .002$) but not in German (algebraic difference scores: $\beta = .09$, $p = .111$; residual scores: $\beta = .08$, $p = .145$).

Table S7

Results of the regression analyses predicting t2 math and German grades by SE bias computed either as an algebraic difference score or a residual score with control variables

Regression analysis	β	*SE*	β 95% *CI*	*p*	R^2
*t*2 math grade on algebraic difference scores					.433
Intercept	-.023	.049	[-.119, .073]	.637	
SE Bias difference score	.073	.057	[-.038, .184]	.200	
Gender	.059	.049	[-.036, .155]	.225	
Parental Educational Level	-.067	.050	[-.166, .031]	.182	
Number of books at home	.164	.053	[.061, .268]	.002	
*t*1 math grade	.625	.041	[.545, .705]	< .001	
*t*2 math grade on residual scores					.460
Intercept	-.023	.048	[-.117, .070]	.626	
SE Bias residual score	.217	.060	[.100, .334]	< .001	
Gender	.023	.049	[-.074, .119]	.644	
Parental Educational Level	-.058	.050	[-.156, .040]	.243	
Number of books at home	.157	.050	[.060, .255]	.002	
*t*1 math grade	.538	.050	[.440, .636]	< .001	
*t*2 German grade on algebraic difference scores					.339
Intercept	-.046	.053	[-.151, .059]	.389	
SE Bias difference score	.024	.059	[-.091, .140]	.677	
Gender	-.001	.056	[-.111, .110]	.993	
Parental Educational Level	-.022	.055	[-.130, .086]	.687	
Number of books at home	.094	.059	[-.021, .209]	.111	
*t*1 German grade	.557	.048	[.463, .651]	< .001	

Table S7 - continued

Regression analysis	β	*SE*	β 95% *CI*	*p*	R^2
*t*2 German grade on residual scores					.360
Intercept	-.045	.053	[-.149, .058]	.388	
SE Bias residual score	.165	.059	[.049, .280]	.005	
Gender	.024	.056	[-.087, .134]	.676	
Parental Educational Level	-.034	.054	[-.140, .071]	.521	
Number of books at home	.084	.058	[-.029, .197]	.145	
*t*1 German grade	.508	.055	[.401, .614]	< .001	

Note. β = standardized regression weight. *SE* = standard error of β; *p* = *p*-value; R^2 = proportion of explained variance in the criterion. Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur; 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100; 2 = more than 100. *N* = 284.

Supplement 8. R and SPSS syntaxes

The R and SPSS syntaxes for the statistical analyses of the present study are freely available at: https://doi.org/10.23668/PSYCHARCHIVES.4285

References

Bergmann, C., & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (AIST-R/UST-R) [General Interest Structure Test with Environment Structure Test (AIST-R/UST-R)]*. Göttingen, DE: Beltz Test.

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI) - Manual [NEO Five Factor Inventory (NEO-FFI) by Costa and McCrae - Manual], 2nd ed. (rev)*. Göttingen, DE: Hogrefe.

Enders, C. K. (2008). A note of the use of missing auxiliary variables in full information maximum likelihood-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 434-448. doi: 10.1080/10705510802154307

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The test anxiety inventory TAI-G: An extended and modified version with four components]. *Zeitschrift für Pädagogische Psychologie, 5*, 121–130.

Hodapp, V. (1996). The TAI-G: A multidimensional approach to the assessment of test anxiety. In C. Schwarzer, & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 95-130). Tübingen, DE: Francke.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for the assessment of teacher and student attributes. Documentation of psychometric instruments as part of the scientific support of the pilot project self-efficient schools]*. Berlin, DE: Freie Universität Berlin.

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and first validation of a scale assessing subjective educational task values (SESSW)]. *Diagnostica, 56*(4), 195-211. doi:10.1026/0012-1924/a000023

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. doi:10.1002/per.2137

2.2 Beitrag II

Linear and nonlinear relationships between self-evaluation and self-evaluation bias with grades

Paschke, P, Weidinger, A. F., & Steinmayr, R. (2023). Linear and nonlinear relationships between self-evaluation and self-evaluation bias with grades [Lineare und nicht lineare Zusammenhänge von Selbsteinschätzung und Selbsteinschätzungsbias mit Schulnoten]. *Learning and Individual Differences, 102*. https://doi.org/10.1016/j.lindif.2023.102266

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Note: This article is not the copy of record and may not exactly replicate the final, authoritative version of the article.

Abstract

There is an ongoing debate whether a bias in self-evaluated competence (SE bias) affects academic achievement, typically assessed as grades. Different rationales about possible relations (e.g., linear, nonlinear, or no relations) between SE bias and academic achievement have been proposed. We compared these rationales in a study with 504 secondary school students in highest academic track secondary schools in math, using response surface analysis (RSA). At *t*1, students' self-evaluated math competence, their competence in math (i.e., objective competence test scores), and their academic achievement in math (i.e., math grades) were assessed. At *t*2, students' math grades were assessed again. The model that fit the data best was the beneficial self-evaluation and competence model, positing only linear relationships of self-concept and competence with change in grades beyond control variables (i.e., gender, parental education, number of books at home) and no SE bias effects.

Keywords: Self-evaluation, self-evaluation bias, response surface analysis, math grades, academic achievement

Zusammenfassung

Es gibt eine fortlaufende Diskussion, ob sich Verzerrungen in den Einschätzungen eigener Kompetenzen (Selbsteinschätzungsbias; SE Bias) auf akademische Leistung, typischerweise erfasst über Noten, auswirken. Es wurden unterschiedliche Theorien über mögliche Zusammenhänge zwischen SE Bias und akademischer Leistung aufgestellt (z.B. lineare, nonlineare oder keine Zusammenhänge). Wir vergleichen diese Theorien in einer Studie mit 504 Gymnasiast*innen im Fach Mathematik mit der response surface analysis (RSA). Zu *t*1 wurden die selbsteingeschätzte Mathematikkompetenz, die tatsächliche Mathematikkompetenz (erfasst mit einem Kompetenztest) und die akademische Leistung in Mathematik (Mathematiknoten) erhoben. Zu *t*2 wurden erneut die Mathematiknoten erfasst. Das Modell, welches die Daten am besten repräsentierte, war das Modell, welches lediglich positive lineare Zusammenhänge der Selbsteinschätzung und der tatsächlichen Kompetenz mit der Veränderung in den Noten beinhaltete, aber keine SE Bias Effekte. Verschiedene Kontrollvariablen (Geschlecht, Bildungshintergrund der Eltern, Anzahl Bücher im Haushalt) wurden kontrolliert.

Stichwörter: Selbsteinschätzung, Selbsteinschätzungsbias, response surface analysis, Mathematiknoten, Akademische Leistung

1. Linear and nonlinear relationships between self-evaluation and self-evaluation bias with grades.

Thinking highly of own competences and the own capacity to achieve good results in the academic context is beneficial for various measures of students' academic achievement, such as grades, standardized test scores, or attainment (e.g., Lee & Kung, 2018; Trautwein & Möller, 2016; Valentine et al., 2004). We call these effects self-evaluation effects. However, little is known about how the discrepancy between a self-evaluation of competence and objective competence, the so-called self-evaluation bias (SE bias), affects academic achievement. We call these effects SE bias effects. Different theoretical accounts on SE bias effects have been suggested. Some authors suppose that having a higher SE bias is beneficial for academic achievement (e.g., Bonneville-Roussy et al., 2017; Chung et al., 2016; Leduc & Bouffard, 2017), while others argue against such effects (e.g., Paschke et al., 2020; Robins & Beer, 2001). In general, studies on SE bias effects on academic achievement have yielded mixed results and do not yet allow drawing an unequivocal conclusion on the influence of SE bias on academic achievement (Bonneville-Roussy et al., 2017; Paschke et al., 2020; Trautwein & Möller, 2016). Additionally and most importantly, approaches on this topic are often limited by theoretical and methodological problems. On a theoretical level, arguments for SE bias effects have been intermingled or confused with arguments for self-evaluation effects. On a methodological level, methods were used that confound SE bias effects with self-evaluation effects (see Humberg et al., 2018; Humberg et al., 2019 for a discussion). In the present study, we used models from response surface analysis (RSA; Edwards, 2002) developed by Humberg et al. (2019) in order compare different rationales about SE bias effects on academic achievement in two high-tracking schools ("Gymnasium") in math. Thus, we use a novel methodological approach to address a long lasting question in educational science (e.g., Helmke, 1998; Lopez et al., 1998; Taylor & Brown 1988).

1.1 Self-evaluation effects and SE bias effects on academic achievement

Self-evaluation effects are documented in several meta-analyses (e.g., Huang, 2011; Talsma et al., 2018; Valentine et al., 2004; Wu et al., 2021). Well-established theoretical models posit that self-evaluations and academic achievement mutually affect each other (reciprocal effects model; e.g., Arens et al., 2017; Lee & Kung, 2018; Preckel et al., 2017; Seaton et al., 2015; Wu et al., 2021; but see Burns et al., 2020) and that the effect of self-evaluation on academic achievement is mediated by expectancies of success and task values (Eccles & Wigfield, 2020; Wigfield et al., 2020), as well as other motivational variables (e.g., Helmke, 1990; Trautwein et al., 2009; Wigfield & Eccles, 2000; see Trautwein & Möller, 2016). Contrary to self-evaluation effects, SE bias effects and their underlying mechanisms are not yet well understood. An SE bias can be positive (the self-evaluation exceeds the competence; overestimation), negative (the competence exceeds the self-evaluation; underestimation) or zero (the self-evaluation equals the competence; perfect accuracy of selfevaluation). In the following, we will summarize and discuss five different rationales about SE bias effects that can be derived from theoretical considerations in the literature. These rationales can be categorized into positing linear, nonlinear, or no relations between SE bias and academic achievement (see also Humberg et al., 2019).

1.1.1 Linear SE bias effects on academic achievement

1.1.1.1 Beneficial SE bias hypothesis. The beneficial SE bias hypothesis posits a positive linear effect of SE bias on academic achievement. Several studies have found that having a higher SE bias predicts increased future academic achievement. In these studies, academic achievement was most commonly operationalized by grade scores (e.g., Bouffard et al., 2011; Chung et al., 2016; Côte et al., 2014; Dupeyrat et al., 2011; Gramzow et al., 2008; Leduc & Bouffard, 2017; Lopez et al., 1998; Willard & Gramzow, 2009; Wright, 2000). However, other measures such as teacher ratings (Bonneville-Roussy et al., 2017; Leduc & Bouffard, 2017), SAT test scores (Chung et al., 2016), or exam results (Martin & Debus,

1998) have also been used occasionally.

The idea that unrealistically positive self-evaluations can be beneficial has been established by the influential paper on *positive illusions* (Taylor & Brown, 1988). These authors suggest that positive illusions can "foster motivation, persistence at tasks, and ultimately, more effective performance" (Taylor & Brown, 1988, p. 199). The concept of positive illusions in the academic context has since been adopted by other researchers (e.g., Bonneville-Roussy et al., 2017; Dufner et al., 2015; Wright, 2000). Likewise, Helmke (1998) considers overestimation as an "auxiliary motor" (p. 130) which positively affects task choice, task persistence, motivation, and performance. Indeed, for the math domain, it has been shown that both, an inflated ability self-concept (Martin & Debus, 1998) and inflated perceptions of past performance (Bench et al., 2015), lead to an increased intent to engage in math in the future. Additionally to the pathways from SE bias via motivation and persistence, it has also been suggested that an SE bias can affect academic achievement positively through increased engagement, external attributions of failure, and a tendency to interpret difficulties as challenges (Bonneville-Roussy et al., 2017). Albeit not testing mediation effects directly, several studies have found that having a higher SE bias predicts higher future academic achievement as well as other academically relevant variables including higher motivation (Willard & Gramzow, 2009; Martin & Debus, 1998), psychological adjustment in the academic context (Bouffard et al., 2011; Chung et al., 2016; Dufner et al., 2015), better self-regulation and less school alienation (Leduc & Bouffard, 2017), and behavioral composure and less anxiety (Gramzow et al., 2008; Willard & Gramzow, 2009). However, some authors argue that the positive link between a higher SE bias and academic achievement is not causal (Gramzow et al., 2003; Willard & Gramzow, 2009) and thus, the aforementioned findings of SE biases predicting increased future academic achievement might not reflect actual SE bias effects.

1.1.1.2 Detrimental SE bias hypothesis. The much less frequently discussed detrimental SE bias hypothesis posits the opposite of the beneficial SE bias hypothesis: a negative linear effect of SE bias on academic achievement. To our knowledge, the only rationale of that kind is that students can use underestimation as a technique to increase their motivation, while overestimation hinders learning because it leads to unrealistic study patterns (Rohr & Ayers, 1973).

1.1.2 Nonlinear SE bias effects on academic achievement

1.1.2.1 Optimal margin hypothesis. The optimal margin hypothesis posits that there is a certain positive level of SE bias, which is optimal for future academic achievement, while increasing the SE bias even further is detrimental. The idea of an optimal level of self-overestimation has been described in the theory of the *optimal margin of illusion* (Baumeister, 1989). In regard to effects on achievement, according to this theory, illusions can be harmful because they can lead to self-handicapping behavior, for example learning less for an upcoming test or investing fruitless efforts on unsolvable tasks, but also helpful as they prevent nervousness and insecurity (Baumeister, 1989; see also Helmke, 1998; Taylor, 1989, p. 240; Taylor & Brown, 1994). However, to our knowledge, no study has yet found an optimal margin of illusion for an effect of SE bias on academic achievement (Lopez et al., 1998; Praetorius et al., 2016; Wright, 2000).

1.1.2.2 Self-knowledge hypothesis. The self-knowledge hypothesis posits that an SE bias of zero is optimal and thus, both overestimation and underestimation are detrimental for academic achievement. To our knowledge, researchers who investigated SE bias effects on academic achievement have not explicitly proposed this hypothesis. However, in a different research field, metacognitive learning, a large body of evidence suggests that accurate self-evaluations might be important in the academic context (see below).

Contrary to the theories of positive illusions and optimal margins of illusions, researchers from the field of metacognitive learning stress the importance of accurate selfevaluations in order to motivate and guide effort and learning behavior (e.g., Buckelew, et al., 2013; Dunlosky & Rawson, 2012; Hacker & Bol, 2019; Händel & Fritzsche, 2016; Kim et al., 2010; Nietfeld & Schraw, 2002; Thiede et al., 2003). Thus, both overestimation and underestimation should be detrimental for academic performance and achievement. Indeed, research suggests that students monitor their learning and use the obtained information to control their learning behavior (e.g., Boekaertes, 1997; Dunlosky et al., 2005; Hong et al., 2020). For example, knowing which contents have been well-learned and which have not, helps students decide where to apply additional learning efforts and how to set their learning goals (Dunning et al., 2004; Dunlosky, et al., 2005; Dunlosky & Rawson, 2012). Thus, inaccurate judgements can impair learning success (Cogliano et al., 2020; Dunlosky & Rawson, 2012; Ehrlinger & Shain, 2014). Following these considerations, it seems plausible that both self-overestimation and self-underestimation could be detrimental for academic achievement.

However, because these studies stem from a different research field than SE bias research, they typically have a different focus and methodology. SE bias research, at least in the academic context, is usually concerned with effects of discrepancies between relatively general self-evaluations (e.g., math ability self-concept) and relatively general competences (e.g., math competence) on relatively general outcomes (e.g., math grades). On the other hand, metacognitive learning is more concerned with the more fine-grained perspective of the learning process itself. Thus, competence is usually assessed with more specific and narrow tasks and self-evaluations are usually assessed as predictions or postdictions of how well the participant will solve or did solve that particular task, or how sure the participant is that they can or could solve the task (e.g., Dunlosky & Rawson, 2012; Ehrlinger & Shain, 2014; Jang et al., 2020; Temelman-Yogev et al., 2020; see Schraw, 2009). Because of this different focus and methodology, we cannot and do not aim to test particular theories or hypotheses of metacognitive learning. Instead, we adopt the reasoning that accurate self-knowledge might be important for learning success and thus academic achievement to SE bias research and compare it to other hypotheses from this field. Thus, we try to answer the question, whether having an accurate relatively general self-evaluation of own competences is beneficial for academic achievement.

1.1.3 Rationales positing no SE bias effects

1.1.3.1 Beneficial self-evaluation and competence hypothesis. The beneficial selfevaluation and competence hypothesis posits that both self-evaluation and competence are beneficial for future academic achievement and that there are no substantial linear or nonlinear relations between SE bias and future academic achievement. It is based on the facts that 1) other authors have already pointed out important problems of studies on SE bias effects – both on a theoretical and on a methodological level – which compromises their interpretability. 2) To our knowledge, the only study on SE bias effects on academic achievement that avoided these problems did not find any SE bias effects (Paschke et al., 2020). Regarding theoretical problems, some authors have criticized a lack of theoretical distinctions between SE bias effects and self-evaluation effects in the literature (Colvin & Block, 1994; Humberg et al., 2018; Humberg et al., 2019; Paschke et al., 2020). For example, as an explanation for beneficial SE bias effects on academic achievement, Bonneville-Roussy et al. (2017) suggest that self-overestimating children more often interpret difficulties as challenges and make external attributions to failure. However, it is not made clear why it should be the overestimation rather than the high self-evaluation per se that positively affects the children's interpretations and attributions in achievement situations (see Paschke et al., 2020).

Additionally, many studies on SE bias effects suffer from applying methods that confound SE bias effects with the main effects of the self-evaluation and objective competence. Both traditional methods of computing SE biases, algebraic difference scores and residual scores, inherently confound the self-evaluation alone with the SE bias (Asendorpf & Ostendorf, 1998; Edwards & Parry, 1993; Griffin et al., 1999; Humberg et al., 2018; Humberg et al., 2019). Thus, it cannot be ruled out that any observed association between an SE bias score and academic achievement is an artifact of the association of the self-evaluation and academic achievement. A recent study (Paschke et al., 2020) made use of a novel method called condition-based regression analysis (CRA; Humberg et al., 2018), which was developed in order to disentangle SE bias and self-evaluation effects. The authors analyzed SE bias effects in 10th grade students in math and German and only found positive effects of the self-evaluation and competence on grades in the same domain, but no SE bias effects. However, since the CRA only allows testing for linear SE bias effects, this result only contradicts the beneficial and detrimental SE bias hypotheses, but not the optimal margin and self-knowledge hypotheses.

1.2 The present research

In the present study, we replicate and extend the findings by Paschke et al. (2020) to evaluate linear and nonlinear associations between SE bias in math competence and academic achievement in math. Thus, our study is the first to empirically compare the five different hypotheses about linear and nonlinear SE bias effects on academic achievement outlined above. To do so, we compared self-evaluated math competence (for brevity simply called *self-evaluation* in the following) to math competence as scored on a standardized math competence test (called *objective competence* in the following). We included several covariates (prior math grades, gender, parental educational level, and number of books at the students' home) in our main analyses. Including prior grades is especially important in our view because grades, standardized test scores and self-evaluations in the same domain are positively correlated (Möller et al., 2009; 2014). Therefore, without including prior grades as a predictor, any observed regression weight from the regression of grades on self-evaluation or objective competence might not reflect a possible causal link but simply a correlation between these variables, which is relatively stable over time. The other covariates were included for similar reasons. Gender correlates with self-evaluations and grades in math (Heyder et al., 2019; Skaalvik & Skaalvik, 2004; Steinmayr & Spinath, 2008; Wilgenbusch & Merrell, 1999). Parental educational level and number of books at the students' home are related to objective math competence (Stubbe et al., 2016; OECD, 2016; Schwippert et al., 2020). Additionally, there is a correlation between students' socioeconomic background and the grades assigned by teachers beyond the effect of the students' objective competence (Lauermann et al., 2020; Maaz et al., 2013; Steinmayr et al., 2012). Thus, since gender and socioeconomic variables are related to both the predictors in our regression analyses (self-evaluation and objective competence) as well as the criterion (grades), not controlling for them could lead to part of the observed predictive value of self-evaluation and objective competence being an artifact of the effect of these variables.

We explain the statistical procedure in detail in the method section. However, in order to derive our specific research questions, a brief introduction to the main method we used is necessary. This is the response surface analysis (RSA; Edwards, 2002). The RSA is essentially a multiple polynomial regression of second order, which we explain in the following: The criterion (i.e., math grades) is predicted by five predictors: two linear terms (i.e., linear self-evaluation and linear objective competence), two quadratic terms (i.e., squared self-evaluation and squared objective competence), and an interaction term (self-evaluation multiplied by objective competence). This is called the full model. Models that correspond to certain theoretical ideas about the relation between the criterion and the predictors can then be specified through parameter restrictions. For example, the beneficial self-evaluation and competence hypothesis posits that there are only positive linear effects of self-evaluation and objective competence on grades. Thus, the regression weights of linear self-evaluation and linear objective competence are fixed to be positive and all other regression weights are fixed to be zero. Finally, the different models can be compared regarding their fit to the data to determine the best fitting model(s). Because self-evaluation effects on academic achievement are theoretically well-founded, we hypothesized that there would be a positive effect of self-evaluation on math grades. More specifically, we hypothesized that in the full model, in which all parameters are estimated freely, the linear self-evaluation term positively predicts math grades beyond the effects of all other predictor variables (Hypothesis 1). By contrast, due to the contradictory theoretical assumptions and methodological problems in the majority of studies on SE bias effects, we did not specify statistical hypotheses for the five aforementioned theoretical hypotheses about SE bias effects. Instead, we followed the above-mentioned approach and compared the different models underlying the five rationales on SE bias effects outlined above.

2. Method

2.1. Participants and procedure

Participants were 504 students from two cohorts in two schools in Germany. Data stem from a larger longitudinal project investigating the psychosocial scholastic adjustment of students in the last three years of school. We considered data from the first measurement occasion (*t*1) which took place when students were at the beginning of Grade 10 (October/November 2015 for Cohort 1, October/November 2016 for Cohort 2) and on average 15.28 years old (SD = 0.60). There were 264 (52.4%) female and 240 (47.6%) male students. Students were tested in classrooms in groups of about 20 students by trained researchers. Personality tests were administered before the performance tests. At that time, we also received the students' last report card grades from the schools. These were the grades from the report cards at the end of Grade 9, and thus from July 2015 for Cohort 1, and from July 2016 for Cohort 2. Later, we received the students' report card grades from the end of Grade 10 (*t*2) and thus from July 2016 for Cohort 1, and from July 2017 for Cohort 2.

In the federal state in which the present study was conducted, after four years of elementary education, students are split into different secondary school types that vary in their academic demands. All students in the present investigation attended a "Gymnasium", the most academically advanced of the three school types and the only one besides the integrated school that leads to the "Abitur", a degree that qualifies students to study at a university.

The present project is in line with ethical guidelines for human subject research. The participants were not misled about the purpose of the study, there were no exclusion or inclusion criteria, treatments or questions that could cause mental or physical harm and the participants were not part of a vulnerable group. Participation in the study was voluntary and approved by the school administrations beforehand. Approval by an ethics committee was not required in the federal state in which the study was conducted. About 85% of the basic population participated at *t*1. Students not participating were ill, took part in a different extracurricular activity, or spent the semester abroad. We received informed consent forms from the parents of all participating students.

2.2. Measures

***2.2.1. Self-evaluation of math competence at t*1**

At *t*1, we assessed students' self-evaluations of math competence with four subjectspecific items from the absolute school self-concept scale of the SESSKO (German Scales for the Assessment of School-Related Competence Beliefs; Schöne et al., 2002, p. 26). The same four items have successfully been used to assess self-evaluation of math competence before (e.g., Steinmayr et al., 2018). All items consist of self-referencing statements about the students' competence in math, which are rated on a scale from 1 (totally disagree) to 5 (totally agree). English translations of the items are "I am good at math", "It is easy to for me to learn in math", "In math, I know a lot", and "Most assignments in math are easy for me" (see Spinath & Steinmayr, 2012). The scale's internal consistency (Cronbach's Alpha, α) was high (α = .95) in this study.

*2.2.2. Objective math competence at t*1

Students' objective math competence was assessed with the DEMAT 9 at *t*1 (German Mathematics Test for Ninth Grade; Schmidt et al., 2013). The DEMAT 9 consists of 43 open-ended and semi-open tasks organized in nine task types (e.g., linear equations, Pythagoras' theorem). We chose the DEMAT 9 as the math competence test because of its curricular validity (Schmidt et al., 2013) and the fact that it covers a broad range of different task types, making it suitable to assess general math competence and thus well comparable to our also general self-evaluation measure. The internal consistency of the DEMAT 9 was high (α = .84).

*2.2.3. Math grades at t*1 **and** ***t*2**

We chose grades as the measure of academic achievement because they are the most common measure in former studies on this topic and because of their high practical relevance for students (Brookhart et al., 2016). We assessed students' math grades on the last report card they received, which was the report card at the end of Grade 9 at *t*1 and the report card at the end of Grade 10 at *t*2. At the time the present study was conducted, the grading system of the examined schools changed between Grade 9 and Grade 10. Up to Grade 9, grades ranged from 1 (*very good*) to 6 (*insufficient*). From Grade 10 onwards, grades ranged from 0 to 15 with higher scores indicating better grades. We reversed the grades from Grade 9, so that higher scores indicate better grades at both measurement time points.

*2.2.4. Sociodemographic variables at t*1

Parental educational level was assessed at *t*1 with the question "Which qualification do your parents have? Please select only the highest qualification". Students selected the highest qualification for each parent and we dichotomized the answers into "Abitur or Fachabitur" (coded as 1) and "lower qualifications" (coded as 0). The Fachabitur is a degree that qualifies students to enter a university of applied sciences, but not a university. The higher of the two scores for a student's parents was treated as the parental educational level. Number of books at the students' home was assessed with the question "How many books are present at your home? *Approximately 40 books fit on one bookshelf. Magazines, newspapers, and your school books are not included*". Students selected one of six response choices ranging from "fewer than 10" to "more than 500". We dichotomized the answers into "100 or less (coded as 0) and "more than 100" (coded 1). This is the common procedure in large scale assessments such as PISA, TIMSS, and IGLU (Hußmann et al., 2017; OECD, 2016; Schwippert et al., 2020).

2.3. Statistical analysis

Analyses were computed in R 4.0.0 with the packages lavaan, AICmodavg, and RSA and the R script provided by Humberg et al. (2019).

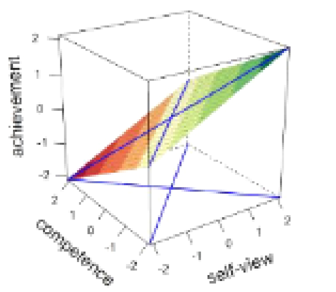
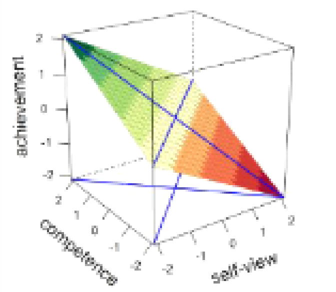
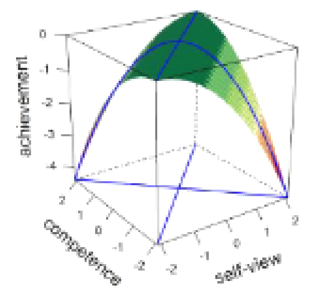
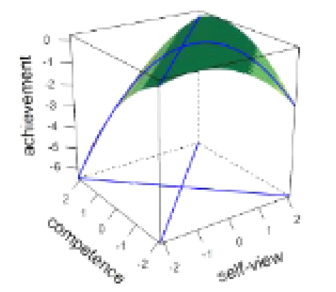
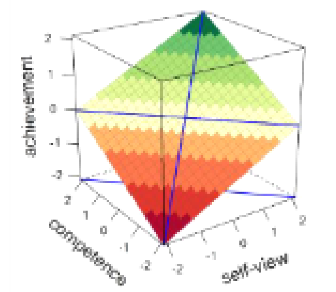
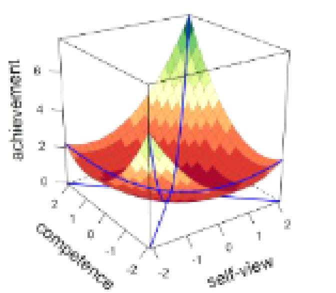
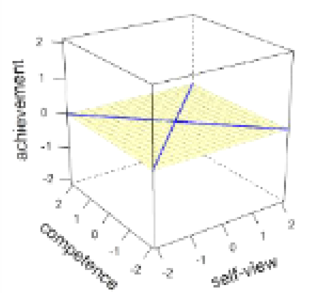
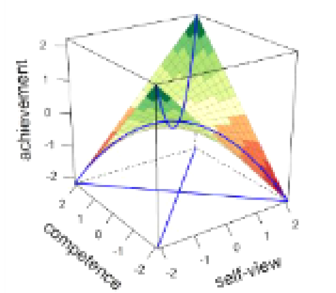
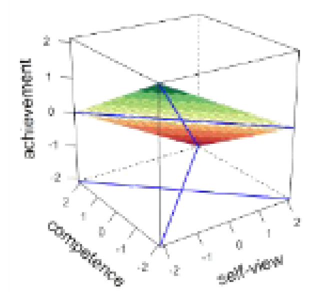
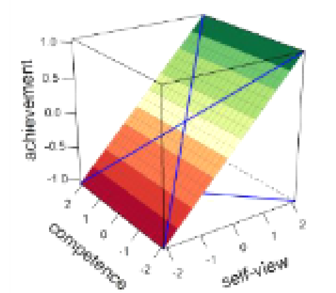
2.3.1. Preliminary analyses

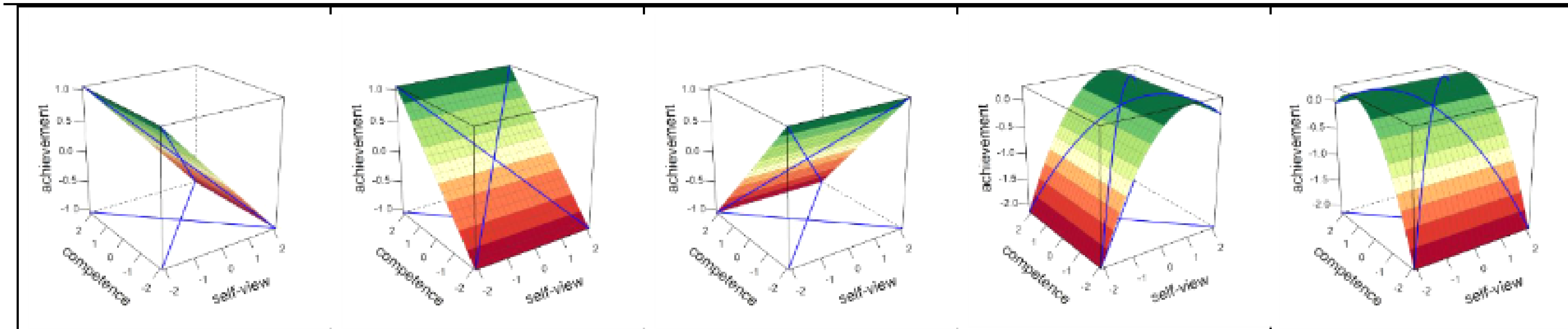
In our models, we included additional predictors (*t*1 math grades, gender, parental educational level, number of books hat home). We did so in order to control for the effects of these additional predictors as has been discussed above. However, the inclusion of additional predictors can potentially change the interpretation of the other predictors (i.e., self-evaluation and objective competence terms; Lynam et al., 2006; Vize et al., 2018). In other words, it is not obvious what, for example, self-evaluation of math competence after controlling for gender and all the other predictors actually is, in terms of its content. Therefore, we examined the extent to which the content of the linear self-evaluation and objective competence terms has been altered by the inclusion of the additional predictors. We did so by computing double-entry intraclass correlations, which can serve as a similarity index between a variable and its residual after partialling out other variables (Vize et al., 2018). For details on the computation of the double-entry intraclass correlations and the reasoning behind it, see online supplemental material 1 (OSM 1; see OSM 2 for detailed results). These additional predictors have been added in the following order: gender, parental educational level, number of books. After computing double-entry intraclass correlations, we largely followed the procedure outlined by Humberg et al. (2019). First, we z-standardized self-evaluation and objective competence before computing quadratic and interaction terms for the regression in order to reduce non-essential multicollinearity (Salmerón-Gómez et al., 2020). Next, we inspected the remaining multicollinearity between predictors and inspected whether there was enough variation between students with higher z-standardized self-evaluations or objective competence scores (Shanock et al., 2010). We tested for outliers according to the three conditions for multivariate outliers described in detail by Humberg et al. (2019, p. 847; see also OSM 3). Finally, before specifying the models corresponding to the different SE bias hypotheses, we tested whether the full model, in which all regression weights are estimated freely, explained a significant amount of variance in the *t*2 math grade scores. Since this was the case, we then specified the constrained models.

2.3.2. Constrained models and information-theoretic model comparison

As described above, we specified five models that correspond to the five hypotheses derived from the literature on SE bias effects on academic achievement. For an overview of these models, including their parameter constraints, visual representations and short descriptions, see the top row of Figure 1. In OSM 4, the methodologically interested readers can find a detailed description why these statistical models correspond to the respective hypotheses.

We used the information-theoretic (IT) approach for model comparison. In the IT approach, the likelihood for each model to be the best model to explain the data, given the competing models, can be estimated. Thus, if certain plausible competing models are omitted, the likelihood of the remaining models to best explain the data can be overestimated. Therefore, the IT approach requires that all models nested within the full model that could plausibly represent the mechanisms by which the outcome variable is affected must be taken into account (Burnham & Anderson, 2002; Symonds & Moussalli, 2011). Thus, we specified 10 additional models that could theoretically explain the significant associations with grade scores observed in the full model. Therefore, our *initial model set* comprised 15 models. The

				
Beneficial SE Bias model	Detrimental SE Bias Model	Self-Knowledge model	Optimal Margin Model	Beneficial self-evaluation and competence model
The higher the discrepancy between S and C, the higher is A.	The higher the discrepancy between S and C, the lower is A.	The higher the absolute discrepancy between S and C, the lower is A.	A is highest for a certain positive discrepancy between S and C but decreases for even larger discrepancies.	The higher S and the higher C, the higher is A.
$b_1 > 0$; $b_2 < 0$; $b_3 = b_4 = b_5 = 0$	$b_1 < 0$; $b_2 > 0$; $b_3 = b_4 = b_5 = 0$	$b_1 = b_2 = 0$; $b_3 = b_5 < 0$; $b_3 + b_4 + b_5 = 0$	$b_1 + b_2 = 0$; $b_1 - b_2 > 0$; $b_3 = b_5 < 0$; $b_3 + b_4 + b_5 = 0$	$b_1 > 0$; $b_2 > 0$; $b_3 = b_4 = b_5 = 0$
				
Full Model	Null model	Interaction model	Detrimental self-evaluation and competence model	Beneficial self-evaluation model
There are linear as well as quadratic and interactive associations of S and C with A.	There are no associations of S or C with A.	The association of C with A is more positive or less negative at higher values of S.	The higher S and the higher C, the lower is A.	The higher S, the higher is A.
No parameter constraints	$b_1 = b_2 = b_3 = b_4 = b_5 = 0$	$b_3 = b_5 = 0$; $b_4 > 0$	$b_1 < 0$; $b_2 < 0$; $b_3 = b_4 = b_5 = 0$	$b_1 > 0$; $b_2 = b_3 = b_4 = b_5 = 0$



Detrimental self-evaluation model	Beneficial competence model	Detrimental competence model	Curvilinear self-evaluation model	Curvilinear competence model
The higher S, the lower is A.	The higher C, the higher is A.	The higher C, the lower is A.	The association between S and A decreases or becomes negative with increasing values of S.	The association between C and A decreases or becomes negative with increasing values of C.
$b1 < 0$; $b2 = b3 = b4 = b5 = 0$	$b2 > 0$; $b1 = b3 = b4 = b5 = 0$	$b2 < 0$; $b1 = b3 = b4 = b5 = 0$	$b2 = b4 = b5 = 0$; $b3 < 0$	$b1 = b3 = b4 = 0$; $b5 < 0$

Figure 1 - continued. Response surfaces, short descriptions, and parametric definitions of models about the associations between self-evaluation (S) and competence (C) with academic achievement (A). The parametric definitions are based on a regression of the form $A = b0 + b1*S + b2*C + b3*S^2 + b4*SC + b5*C^2$. The five models derived from theoretical rationales are depictured in the top row of this Figure.

parameter constraints, visual representations and short descriptions of the 10 additional models are reported in the two bottom rows of Figure 1 and discussed in detail in OSM 4. Following the procedure by Humberg et al. (2019) we computed the second-order Akaike Information Criterion (*AICc*) for each model as well as Akaike weights, which are based on the *AICc* values of the models in a given model set. Akaike weights can be interpreted as the likelihood of a model to best explain the data given the competing models. Therefore, Akaike weights can be used for model selection among nested or non-nested models and to estimate model selection uncertainty. For example, in case of two models, one would be relatively certain to select model A over a competing model B if model A had a likelihood of 99% to best explain the data while model B had a likelihood of 1%. If model A had a likelihood of 60% and model B had a likelihood of 40%, that would indicate that there is some evidence for both models and one should not readily discard model B. After testing the initial model set, we excluded models that were virtually redundant as their log-likelihood was essentially the same as a model nested within them, creating a *reduced model set*. In such a case, the more parsimonious model should be selected and the larger model discarded since it requires the estimation and interpretation of additional parameters, which do not offer a substantial improvement in the approximation to the data. Thus, if the difference in the log-likelihood of two nested models was smaller than 1, we excluded the larger model (Humberg et al., 2019, p. 847; Symonds & Moussalli, 2011). We then again tested all remaining models and computed their Akaike weights. Finally, we sorted these models by their Akaike weights and assembled a *confidence set,* which included all models with the highest Akaike weights whose combined Akaike weights exceeded 95%. Thus, with a likelihood of more than 95%, the best model to explain the data among all tested models is one of the models in the confidence set. If there was only one model in the confidence set, that is, a single model had a likelihood of over 95% to be the best model, we considered this clear evidence for the respective model. Otherwise, there would be some evidence for multiple models.

2.3.3. Missing data

Missing data in repeated measures designs can be categorized as resulting from attrition or nonresponse (Little, 2013). Regarding attrition, 6.3% of the *t*2 grade scores were missing. Due to nonresponse at *t*1, 0.6% of the scale level self-evaluation scores, 0.8% of the math grade scores, 1.6% of the scores for number of books at home, and 7.5% of the scores for parental educational level were missing. There were no missing data for the objective competence scores. Missing data in all regression models was treated with full information maximum likelihood (FIML) estimation.

3. Results

3.1. Preliminary results

Descriptive statistics of all non-dichotomous analysis variables are reported in Table 1. For detailed results on the analysis of multicollinearity, see Table 2 that shows the correlations between the variables. The correlations between the predictor variables ranged from $r = -.42$ ($p < .001$) between the linear and quadratic self-evaluation terms to $r = .64$ ($p < .001$) between the linear self-evaluation term and the *t*1 math grades. Inspecting how many students had higher z-standardized self-evaluations vs. objective competence scores revealed that the difference between self-evaluation and objective competence was below -.50 for 27% of the participants, between -.50 and .50 for 45%, and above .50 for 28%. Thus, there was enough variation between students with higher z-standardized self-evaluations vs. objective competences. An inspection of the double-entry intraclass correlations revealed that they were very high for linear self-evaluation ($r_{ICC} = .99$, $p < .001$) and high for linear objective competence ($r_{ICC} = .83$, $p < .001$), indicating that the inclusion of gender, parental educational level, and number of books at home did not have a large influence on the content of these variables (for detailed results and discussions, see OSM 2). Testing the full model revealed that it explained a significant and substantial amount of variance in the *t*2 math grade scores ($R^2 = .45$, $p < .001$). The addition of the covariates had little impact on the amount of explained variance (model with no covariates: $R^2 = .42$; model also including gender: $R^2 = .44$; model also including parental educational level: $R^2 = .45$; model also including number of books: $R^2 = .45$) There were no outliers according to our criteria (see OSM 3 for details on the criteria).

Table 1

Sample sizes (n), means (M), standard deviations (SD), skewness, kurtosis, and Cronbach's α for all nondichotomous analyses variables in math

	n	*M*	*SD*	Skewness	Kurtosis	α
Self-evaluation $t1^a$	501	3.41	1.13	-0.44	-0.73	.95
Objective competence $t1^b$	504	24.88	6.36	-0.47	0.06	.84
Grade $t1^c$	500	4.27	1.01	0.03	-0.69	-
Grade $t2^d$	473	8.93	3.16	0.02	-0.81	-

[a]Possible Range 1 to 5; [b]Possible Range 0 to 43; [c]Possible Range 1 to 6; [d]Possible Range 0 to 15.

3.2 Self-evaluation effects and initial model set

Table 3 shows the detailed results for the full model. In the full model, the linear self-evaluation term positively and significantly predicted *t*2 math grades (β = .26, $p < .001$), supporting Hypothesis 1. Additionally, the linear objective competence term (β = .21, $p < .001$), the *t*1 math grades (β = .33, $p < .001$), gender (β = -.18, $p < .001$), and parental educational level (β = .09, $p = .014$) significantly predicted *t*2 math grades. By contrast, the quadratic self-evaluation term (β = -.04, $p = .333$), the quadratic objective competence term (β = .04, $p = .320$), the self-evaluation and objective competence interaction term (β = .03, $p = .484$), and the number of books at home (β = -.05, $p = .177$) did not significantly predict *t*2 grades. In Table 4, the results for the initial model set including *AICc*s, Akaike weights (*w*), and log-likelihoods are reported in detail. An inspection of the log-likelihoods revealed that

Table 2

Correlations between the model variables

	2	3	4	5	6	7	8	9	10
1 Self-evaluation linear term	.54***	-.42***	-.16**	-.08	.64***	.56***	.26***	.06	.10*
2 Objective comp. linear term	-	-.13**	-.07*	-.27***	.48***	.44**	.34***	.12**	.14***
3 Self-evaluation quadratic term		-	.51***	.23***	-.16**	-.20***	-.14**	.13**	.01
4 Self-evaluation and objective comp. interaction term			-	.46***	-.04	-.02	-.06	.09	.05
5 Objective comp. quadratic term				-	.02	-.01	-.07	.07	.05
6 *t*1 grade					-	.61***	.06	.11*	.10**
7 *t*2 grade						-	-.03	.05	.15**
8 Gender							-	-.05	.07
9 Number of books								-	.25***
10 Parental educational level									-

Note. N = 439 – 504; Objective comp. = objective competence; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100; 2 = more than 100.

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

Table 3

Detailed results for the full model predicting t2 grades in math

Predictor	β	*SE*	*95% CI*	*p*
(Intercept)	1.75	0.284	[1.19, 2.31]	< .001
Self-evaluation linear term	.26	.048	[0.16, 0.35]	< .001
Objective comp. linear term	.21	.050	[0.11, 0.31]	< .001
Self-evaluation quadratic term	-.04	.044	[-0.13, 0.04]	.333
Self-evaluation and objective comp. interaction term	.03	.044	[-0.06, 0.12]	.484
Objective comp. quadratic term	.04	.042	[-0.04, 0.13]	.320
*t*1 grade	.33	.046	[0.24, 0.42]	< .001
Gender	-.18	.036	[-0.25, -0.11]	< .001
Number of books	-.05	.037	[-0.12, 0.02]	.177
Parental educational level	.09	0.35	[0.02, 0.16]	.014

Note. N = 504. 95% *CI* = 95% confidence interval; Objective comp. = objective competence; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 = .45.

there were several models of the overall 15 models that did not offer any substantial advantage over more parsimonious models (difference in log-likelihood between nested models < 1) and should thus be removed. In summary, only the null model (11), the self-knowledge model (10), the beneficial competence model (7), the beneficial self-evaluation model (4), and the beneficial self-evaluation and competence model (1) remained in the analysis.

3.3 Reduced model set and confidence set

In Table 5, the results for the reduced model set are reported. Detailed results for the model in the confidence set (i.e., the beneficial self-evaluation and competence model) are reported in Table 6. Visual representations of the empirical results for this model and the full model based on the observed relations in this study are depicted in Figure 2. The Akaike

Table 4

Results for the initial model set sorted by Akaike weights (w)

Model	*AICc*	*w*	*w cumul.*	*LL*
1 Beneficial self-evaluation and competence[4,7,11]	12343.60	68.9%	68.9%	-6100.96
2 Interaction[1,4,6,7,9,11,13,14,15]	12345.52	26.4%	95.3%	-6100.62
3 Full[1,2,4,5,6,7,8,9,10,11,12,13,14,15]	12348.99	4.7%	> 99.9%	-6099.72
4 Beneficial self-evaluation[11]	12359.47	< 0.1%	> 99.9%	-6110.20
5 Curvilinear self-evaluation[4,11,13]	12362.06	< 0.1%	> 99.9%	-6110.19
6 Beneficial SEBias[4,11,14]	2362.07	< 0.1%	> 99.9%	-6110.20
7 Beneficial competence[11]	12373.56	< 0.1%	> 99.9%	-6117.24
8 Curvilinear competence[7,11,14]	12376.16	< 0.1%	> 99.9%	-6117.24
9 Detrimental SE Bias[7,11,13]	12376.16	< 0.1%	> 99.9%	-6117.24
10 Self-Knowledge[11]	12402.16	< 0.1%	> 99.9%	-6131.54
11 Null	12404.07	< 0.1%	> 99.9%	-6133.79
12 Optimal Margin[10,11]	12404.47	< 0.1%	> 99.9%	-6131.40
13 Detrimental self-evaluation[11]	12406.65	< 0.1%	> 99.9%	-6133.79
14 Detrimental competence[11]	12406.65	< 0.1%	> 99.9%	-6133.79
15 Detrimental self-evaluation and comp.[11,13,14]	12409.25	< 0.1%	> 99.9%	-6133.79

Note. N = 504; comp. = competence; *AICc* = second order Akaike Information Criterion; *w* = Akaike weight; *w* cumul. = cumulated Akaike weights; *LL* = log-likelihood. Superscript numbers indicate that another model is nested within this model, e.g., "Beneficial self-evaluation and competence[4,7,11]" denotes that models 4, 7 and 11 are nested within the beneficial self-evaluation and competence model. Models in bold are based on theoretical considerations in the literature, while models in normal font have been added for methodological reasons; models in italics have been selected for the reduced model sets based on *LL*s (see Table 5).

Table 5

Results for the reduced model set sorted by Akaike weights (w)

	w	*w cumul.*
Model		
Beneficial self-evaluation and competence[a]	> 99.9%	> 99.9%
Beneficial self-evaluation	< 0.1%	> 99.9%
Beneficial competence	< 0.1%	> 99.9%
Self-Knowledge	< 0.1%	> 99.9%
Null	< 0.1%	> 99.9%

Table 6

Detailed results for the model in the confidence set: The beneficial self-evaluation and competence model

	Beneficial self-evaluation and competence model			
Predictor	*β*	*SE*	*95% CI*	*p*
(Intercept)	1.72	0.278	[1.18, 2.27]	< .001
Self-evaluation linear term	.28	.045	[0.19, 0.37]	< .001
Objective comp. linear term	.19	.045	[0.10, 0.28]	< .001
Self-evaluation quadratic term	Fixed to 0	-	-	-
Self-evaluation and objective comp. interaction term	Fixed to 0	-	-	-
Objective comp. quadratic term	Fixed to 0	-	-	-
*t*1 grade	.33	.045	[0.25, 0.42]	< .001
Gender	-.17	.036	[-0.25, -0.10]	< .001
Number of books	-.05	.036	[-0.12, 0.02]	.181
Parental educational level	.09	.035	[0.02, 0.16]	.011

Note. $N = 504$; 95% CI = 95% confidence interval; Objective comp. = objective competence; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur; 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100; 2 = more than 100; $R^2 = .45$.

weights clearly indicated that the beneficial self-evaluation and competence model fit the data best ($w > 99.9\%$, see Table 5). A closer inspection of this model (see Table 6) revealed that both the linear self-evaluation term ($\beta = .28$, $p < .001$) and the linear objective competence term ($\beta = .19$, $p < .001$) significantly predicted $t2$ grade scores. Additionally, $t1$ grades ($\beta = .33$, $p < .001$), gender ($\beta = -.17$, $p < .001$), and parental educational level ($\beta = .09$, $p = .011$) significantly predicted $t2$ grades, while the number of books at home did not ($\beta = -.05$, $p = .181$). The model explained a substantial amount of variance in the $t2$ grade scores ($R^2 = .45$). In fact, it explained the same amount of variance as the full model up to the second decimal place.

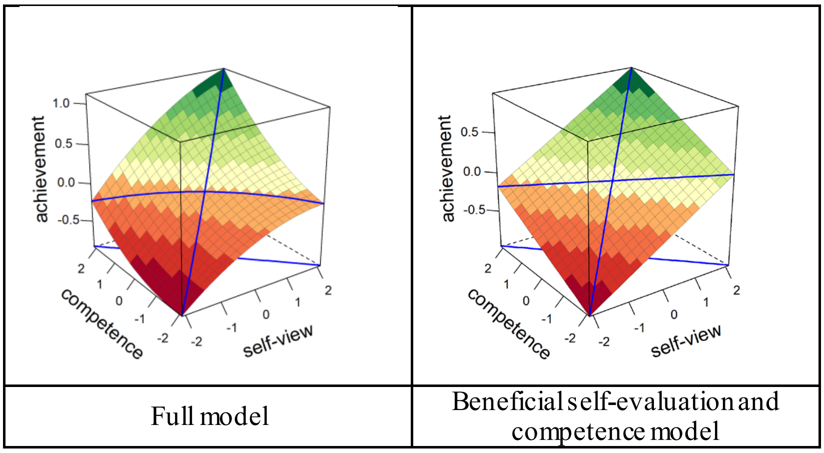


Figure 2. Empirical response surfaces for the full model and the model in the confidence set. The models depict the empirical association between self-evaluation, competence and academic achievement observed in the present study.

4. Discussion

We used a novel methodological approach to test different models on the effects of self-evaluation, objective competence, and self-evaluation bias on grades in math (see Humberg et. al., 2019). Our findings indicate that the beneficial self-evaluation and competence model had by far the greatest likelihood to be the best fitting model, given the data and the competing models. Since neither the beneficial SE bias model, the detrimental

SE bias model, the optimal margin model, nor the self-knowledge model was part of the confidence set, we did not find any evidence for linear or nonlinear SE bias effects beyond the control variables.

4.1 Self-evaluation effects on grades

Our findings support the beneficial self-evaluation and competence model and thus confirmed that the students' self-evaluation and objective competence in math had positive and significant linear effects on math grades beyond the effects of the control variables ($t1$ math grades, gender, parental educational level, number of books at home) and in case of the full model, beyond the nonlinear terms. These findings support Hypothesis 1 and are in line with previous studies in this field (e.g., Arens et al., 2017; Lee & Kung, 2018; Preckel et al., 2017; Seaton et al., 2015). Thus, our results highlight the importance of having a high self-evaluation of own competence for students' future academic achievement in math. Most importantly, our findings extend previous evidence on the important role of self-evaluations of competence for academic achievement because we not only tested linear but also quadratic effects of the self-evaluation and were able to show that there are indeed no quadratic effects in the present study.

4.2 Self-evaluation bias effects on grades

Since the beneficial self-evaluation and competence model was by far the best fitting model, we did not find evidence for linear or nonlinear SE bias effects beyond the control variables. Thus, overall, our results indicate that it is indeed the main effects of self-evaluation and objective competence rather than linear or nonlinear SE bias effects that influence grades beyond the included control variables. This means that both, high self-evaluations and high objective competence, positively affect math grades. On the other hand, the match between self-evaluation and objective competence does not have an influence. Irrespective of whether a student's objective competence is low or high, they benefit equally (or nigh equally) from having a high self-evaluation. Likewise, students benefit from higher competence, irrespective of their level of self-evaluation. These results support the findings by Paschke et al. (2020) for linear SE bias effects and extent them to nonlinear effects. We argue that the significant beneficial SE bias effects on academic achievement found in many studies (e.g. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côte et al., 2014; Leduc & Bouffard, 2017) are likely, at least to some degree, artifacts of self-evaluation effects. This seems plausible because self-evaluation effects on academic achievement are well-documented (e.g. Huang, 2011; Trautwein & Möller, 2016; Valentine et al., 2004) and common methods for computing SE biases confound the effect of the SE bias with the effect of the self-evaluation, as has been pointed out by several authors (e.g., Asendorpf & Ostendorf, 1998; Edwards & Parry, 1993; Griffin et al., 1999; Humberg et al., 2018). Additionally, while the mechanisms behind self-evaluation effects are relatively well understood (e.g., Areepattamannil, 2012; Guay et al., 2010, Helmke, 1990; Trautwein et al., 2009; Wigfield et al., 2020; see Trautwein & Möller, 2016), the theoretical rationales for linear SE bias effects and optimal margin effects remain largely ambiguous (see Colvin & Block, 1994; see Humberg et al., 2018; see Paschke et al., 2020). The main problem of these latter rationales is that they often do not make it clear why it should be the SE bias rather than the self-evaluation per se that affects academic achievement. For example, the idea that a positive SE bias makes children perceive academic difficulties as challenges (Bonneville-Roussy et al., 2017) is not plausible for different levels of the self-evaluation. Why should a student with low competence and average self-evaluation (positive SE bias) show a greater increase in achievement than a student with high competence and high self-evaluation (no SE bias)? Even when disregarding the effect of competence, it would seem more plausible that the second student is more inclined to interpret difficulties as challenges than the first because the self-evaluation of the second student is higher. Thus, it does not become clear, why there should be SE bias effects rather than self-evaluation effects on academic achievement. Our empirical results reflect these considerations on theoretical problems of arguments for SE bias effects by showing the absence of SE bias effects when they are disentangled from self-evaluation effects.

4.3 Limitations and future directions

The first and in our opinion most relevant limitation of the present study is that we assessed SE bias in only one way, by comparing a relatively general measure of self-evaluated math competence to objective math competence. This approach is suitable to investigate how discrepancies in relatively general self-evaluations affect academic achievement, which was the aim of the present study. However, one could also extend this scope to consider the more fine-grained perspective of effects of misjudging success on particular tasks. Based on the results of the present study, we cannot rule out that misjudging success on specific tasks (task-specific self-efficacy) during self-regulated learning is indeed detrimental as posited by researchers from metacognitive learning (e.g., Cogliano et al., 2020; Dunlosky & Rawson, 2012; Hong et al., 2020), while misjudging one's own general competence in the respective field is not. Appraisal of task-specific self-efficacy is largely based on prior experiences of success or failure with similar tasks because these are the most reliable sources of information for judging the chance of success in the future (Bandura, 1986; 1997; Bong et al., 2012; Bong & Skaalvik, 2003). While prior experiences are also relevant in the appraisal of general self-evaluations, frames of references, such as social or dimensional comparisons, and causal attributions of the prior experiences are more important than the prior experiences themselves (Bong & Clark, 1999; Bong et al., 2012; Bong & Skaalvik, 2003; Zimmermann, 1995). Studies on causal attributions of success and failures show that high achieving students more often use personally controllable attributions, for example, on effort, than personally uncontrollable attributions, for example, on abilities (Dong et al., 2015; Cheng & Chiou, 2010; Cortes-Suarez & Sandiford, 2008; Hsieh & Schallert, 2008). Thus, students who correctly identify their own mistakes and inabilities to solve specific tasks might still overestimate their own general competence because they attribute their mistakes and inabilities on low effort rather than low competence. It is possible that it is most beneficial for students to correctly recognize their mistakes on particular tasks (accurate task-specific self-efficacy) while still maintaining a high self-evaluation, so they can benefit from both, well-guided metacognitive learning and the motivating influence of a high self-evaluation. Future studies could test these assumptions by comparing SE bias effects based on general self-evaluations and task-specific self-efficacy. Additionally, it would be interesting to examine the more fine-grained perspective of how particular attributions of successes and failures (e.g. on an inherent ability or effort) affect SE biases based on task-specific self-efficacy as well the SE biases' effects on academic achievement. Second, the interpretation of our results hinges on the DEMAT 9 measuring "objective" math competence. However, like any psychometric test, the DEMAT 9 is not a perfect measurement instrument and the students' scores might have been affected by sources other than competence, such as fatigue or motivation. Third, we analyzed a homogenous sample of students from academic track schools in 10th grade, which constrains the generalizability of our results. However, at least for effects of self-evaluation on academic achievement, previous studies provide evidence that this effect holds equally for students in mixed-ability and academically selective groups (Seaton et al., 2015; Preckel et al., 2017). Fourth, we analyzed self-evaluation and SE bias effects in only one domain, namely math. Thus, future studies should test the generalizability of the present results to other subjects. Fifth, since the present study followed a predictive design instead of an experimental one, we cannot unequivocally draw conclusions about causal relations. We aimed to reduce this problem, which is shared by virtually any other study on self-evaluation or SE bias effects, through the inclusion of control variables that could influence the effects of self-evaluation and objective competence on grades. Sixth, by using the RSA, which is a form of regression analysis, we did not model the latent structure of our data (self-evaluation and objective competence).

4.4 Conclusion

Overall, our results contradict the supposition that a bias in general self-evaluated competence has a substantial impact on academic achievement. Instead, our findings indicated that absolute main effects of self-evaluation and objective competence have greater effects on academic achievement. Thus, both high self-evaluations and high objective competences are beneficial for academic achievement in math, while it does not seem to matter much, how closely these factors match. However, additional studies are needed to test whether the present results can be generalized to students in different grades, less academically selective schools, and in different school subjects, and most importantly, whether they differ when a different self-evaluation construct, such as task-specific self-efficacy, is assessed. The discrepancy between our results on linear SE bias effects and the results of studies using algebraic difference scores or residual scores highlights the importance of employing methods that allow disentangling SE bias effects from self-evaluation effects.

References

Areepattamannil, S. (2012). Mediational role of academic motivation in the association between school self-concept and school achievement among Indian adolescents in Canada and India. *Social Psychology of Education: An International Journal, 15*(3), 367–386. https://doi.org/10.1007/s11218-012-9187-1

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*(5), 621–634. https://doi.org/10.1037/edu0000163

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966.https:// doi.org/10.1037//0022-3514.74.4.955

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Baumeister, R. F. (1989). The optimal margin of illusion. *Journal of Social and Clinical Psychology, 8*(2), 176-189. https://doi.org/10.1521/jscp.1989.8.2.176

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles, 72*(11-12), 536-546. https://doi.org/10.1007/s11199-015-0486-9

Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*(2), 161–186. https://doi.org/10.1016/S0959-4752(96)00015-1

Bong, M., Cho, C., Ahn, H. S., & Kim, H. J. (2012). Comparison of self-beliefs for predicting student motivation and achievement. *The Journal of Educational Research, 105*(5), 336–352. https://doi.org/10.1080/00220671.2011.627401

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*, 139–154. https://doi.org/10.1207/s15326985ep3403_1

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. https://doi.org/10.1016/j.jsp.2017.02.002

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*, 221-229. https://doi.org/10.1016/j.ijer.2011.08.003

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H, Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://doi.org/10.3102/0034654316672069

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*(2), 134-138. https://doi.org/10.1177/0098628312475034

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.)*. Springer.

Burns, R. A., Crisp, D. A., & Burns, R. B. (2020). Re-examining the reciprocal effects model of self-concept, self-efficacy, and academic achievement in a comparison of the cross-lagged panel and random-intercept cross-lagged panel frameworks. *British Journal of Educational Psychology, 90*(1), 77–91. https://doi.org/10.1111/bjep.12265

Cheng, P.-Y., & Chiou, W.-B. (2010). Achievement, attributions, self-efficacy, and goal setting by accounting undergraduates. *Psychological Reports, 106*(1), 54–64. https://doi.org/10.2466/PR0.106.1.54-64

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: A longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*(10), 1384-1401. https://doi.org/10.1177/0146167216662866

Cogliano, M., Bernacki, M. L., & Kardash, C. M. (2020). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000624.supp (Supplemental)

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116*, 3–20. https://doi.org/10.1037/0033-2909.116.1.3

Cortés-Suárez, G., & Sandiford, J. (2008). Causal Attributions for Success or Failure. Of Students in College Algebra. *Community College Journal of Research and Practice, 32*(4-6), 325-346. https://doi.org/ 10.1080/10668920701884414

Côté, S., Bouffard, T., & Vezeau, C. (2014). The mediating effect of self-evaluation bias of competence on the relationship between parental emotional support and children's academic functioning. *British Journal of Educational Psychology, 84*(3), 415–434. https://doi.org10.1111/bjep.12045

Dong, Y., Stupnisky, R. H., Obade, M., Gerszewski, T., & Ruthig, J. C. (2015). Value of college education mediating the predictive effects of causal attributions on academic success. *Social Psychology of Education: An International Journal, 18*(3), 531–546. https://doi.org/10.1007/s11218-015-9299-5

Dufner, M., Reiz, A. K., & Zander, L. (2015). Antecedents, consequences, and mechanisms: On the longitudinal interplay between academic self-enhancement and psychological adjustment. *Journal of Personality, 83*(5), 511-522. https://doi.org10.1111/jopy.12128

Dunlosky, J., Hertzog, C., Kennedy, M. R. F., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10*(1), 4–11.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https//doi.org10.1016/j.learninstruc.2011.08.003

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250. https://doi.org/10.1016/j.ijer.2011.08.005

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61.* https://doi.org/10.1016/j.cedpsych.2020.101859

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613. https://doi.org/10.2307/256822

Ehrlinger, J., & Shain, E. A. (2014). How accuracy in students' self perceptions relates to success in learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum.* (pp. 142–151). Society for the Teaching of Psychology.

Gramzow, R. H., Elliot, A. J., Asher, E., & McGregor, H. A. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality, 37*(2), 41–61. https://doi.org/10.1016/S0092-6566(02)00535-4

Gramzow, R. H., Willard, G., & Mendes, W. B. (2008). Big tales and cool heads: Academic exaggeration is related to cardiac vagal reactivity. *Emotion, 8*(1), 138–144. https://doi.org/10.1037/1528-3542.8.1.138

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6*, 505-518. https://doi.org/10.1111/j.1475-6811.1999.tb00206.x

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*(6), 644–653. https://doi.org/10.1016/j.lindif.2010.08.001

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education*. (pp. 647–677). Cambridge University Press. https://doi.org/10.1017/9781108235631.026

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition, 44*(2), 229–241. https://doi.org/10.3758/s13421-015-0552-0

Helmke, A. (1990). Mediating processes between children's self-concept of ability and mathematical achievement: A longitudinal study. In H. Mandl, E. de Corte, S. N.

Bennett, & H. F. Friedrich (Eds.), *Learning and instruction. European research in an international context. Volume 2.2: Analysis of complex skills and complex knowledge domains* (pp. 537–549). Pergamon Press.

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe [From an optimist to a realist? On the development of the ability self-concept from kindergarten to 6th grade]. In F. E. Weinert (Ed.), *Entwicklung im Kindesalter* (pp. 115-132). Beltz.

Heyder, A., Steinmayr, R., & Kessels, U. (2019). Do Teachers' Beliefs About Math Aptitude and Brilliance Explain Gender Differences in Children's Math Ability Self-Concept? *Frontiers in Education, 4*. https://doi.org/10.3389/feduc.2019.00034

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology, 112*(7), 1409–1430. https://doi.org/10.1037/edu0000445

Hsieh, P.-H. P., & Schallert, D. L. (2008). Implications from self-efficacy and attribution theories for an understanding of undergraduates' motivation in a foreign language course. *Contemporary Educational Psychology, 33*(4), 513–532. https://doi.org/10.1016/j.cedpsych.2008.01.003

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*, 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859. https://doi.org/10.1037/pspp0000204

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology, 114*(2), 303–322. https://doi.org/10.1037/pspp0000134

Hußmann, A., Stubbe, T. C., & Kasper, D. (2017). Kapitel VI. Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, & R. Valtin (Eds.), *IGLU 2016. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich [IGLU 2016. Reading competence of grade schoolers in Germany by international comparison]*. (S. 195-218). Münster, DE: Waxmann.

Jang, Y., Lee, H., Kim, Y., & Min, K. (2020). The relationship between metacognitive ability and metacognitive accuracy. *Metacognition and Learning*. https://doi.org/10.1007/s11409-020-09232-w

Kim, Y.-H., Chiu, C., & Zou, Z. (2010). Know thyself: Misperceptions of actual performance undermine achievement motivation, future performance, and subjective well-being. *Journal of Personality and Social Psychology, 99*(3), 395–409. https://doi.org/10.1037/a0020555

Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology, 112*(2), 364–383. https://doi.org/10.1037/edu0000377.supp (Supplemental)

Leduc, C., & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences, 55*, 193–201. https://doi.org/10.1016/j.lindif.2017.04.006

Lee, C.-Y., & Kung, H.-Y. (2018). Math self-concept and mathematics achievement: Examining gender variation and reciprocal relations among junior high school students in Taiwan. *Eurasia Journal of Mathematics, Science & Technology Education, 14*(4), 1239–1252.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Lopez, D. F., Little, T. D., Oettingen, G., & Baltes, P. B. (1998). Self-regulation and school performance: Is there optimal level of action-control? *Journal of Experimental Child Psychology, 70*, 54-74. https://doi.org/10.1006/jecp.1998.2446

Lynam, D. R., Hoyle, R. H., & Newman, J. P. (2006). The perils of partialing: Cautionary tales from aggression and psychopathy. *Assessment, 13*, 328–341. https://doi.org/10.1177/1073191106290562

Maaz, K., Baeriswyl, F., & Trautwein, U. (2013). II. Studie: „Herkunft zensiert?" Leistungsdiagnostik und soziale Ungleichheiten in der Schule. In Deißner, D. (Ed.), *Chancen bilden [Building chances]*. (pp. 185–334). Springer.

Martin, A. J., & Debus, R. L. (1998). Self-reports of mathematics self-concept and educational outcomes: The roles of ego-dimensions and self-consciousness. *British Journal of Educational Psychology, 68*(4), 517–535. https://doi.org/10.1111/j.2044-8279.1998.tb01309.x

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*(3), 1129-1167. https://doi.org/10.3102/0034654309337522

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*(4), 591–611. https://doi.org/10.1111/bjep.12047

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research, 95*(3), 131–142. https://doi.org/10.1080/00220670209596583

OECD (2016). *PISA 2015 Ergebnisse (Band 1): Exzellenz und Chancengerechtigkeit in der Bildung [PISA 2015 Results (Volume 1): Excellence and Equity in Education]*. Bertelsmann.

Paschke, P., Weidinger, A. & Steinmayr, R. (2020). Separating the Effects of Self-Evaluation Bias and Self-View on Grades. *Learning & Individual Differences*. Advance online publication. https://doi.org/10.1016/j.lindif.2020.101940

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern [Are students with optimistic self-concepts one step ahead? Relations between optimistic, realistic, and pessimistic self-concepts and the achievement development of primary school children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. https://doi.org/10.1177/0016986216687824

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*(2), 340-352.

https://doi.org/10.1037/0022-3514.80.2.340

Rohr, M. E., & Ayers, J. B. (1973). Relationship of student grade expectations, selected characteristics, and academic performance. *Journal of Experimental Education, 41*(3), 58-62. https://doi.org/10.1080/00220973.1973.11011410

Salmerón-Gómez, R., Rodríguez-Sánchez, A. & García-García, C. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational statistics, 35*, 647-666. https://doi.org/10.1007/s00180-019-00922-x

Schmidt, S., Ennemoser, M., & Krajewski, K. (2013). *Deutscher Mathematiktest für neunte Klassen (DEMAT 9) [German Mathematics Test for Ninth Grade (DEMAT 9)].* Hogrefe.

Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzeptes SESSKO [Scales for the assessment of the school self-concept SESSKO].* Hogrefe.

Schraw, G. (2009). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 415–429). Routledge/Taylor & Francis Group.

Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M., & Wendt, H. (2020). *TIMSS 2019 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland und im internationalen Vergleich [TIMSS 2019 grade school students' mathematical and science competences in Germany and by international comparison].* Waxmann.

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly, 59*(3), 143–156.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25*(4), 543–554. https://doi.org/10.1007/s10869-010-9183-4

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles: A Journal of Research, 50*(3–4), 241–252. https://doi.org/10.1023/B:SERS.0000015555.40976.e6

Spinath, B., & Steinmayr, R. (2012). The role of competence beliefs and goal orientations for change in intrinsic motivation. *Journal of Educational Psychology, 104*(4), 1135-1148. https://doi.org/10.1037/a0028115

Steinmayr, R., Dinger, F. C., & Spinath, B. (2012). Motivation as a mediator of social disparities in academic achievement. *European Journal of Personality, 26*(3), 335–349. https://doi.org/10.1002/per.842

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality, 22*(3), 185–209. https://doi.org/10.1002/per.676.

Steinmayr, R., Weidinger, A. F., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology, 53*, 106-122. https://doi.org/10.1016/j.cedpsych.2018.02.004

Stubbe, T. C., Schwippert, K., & Wendt, H. (2016). Kapitel X. Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich [TIMSS 2015. Mathematical and science competences of grade schoolers in Germany by international comparison].* (S. 299-316). Waxmann.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65*, 13–21. https://doi.org/10.1007/s00265-010-1037-6

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind.* Basic Books/Hachette Book Group.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210. https://doi.org/10.1037/0033-2909.103.2.193

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21–27. https://doi.org/10.1037/0033-2909.116.1.21

Temelman-Yogev, L., Katzir, T., & Prior, A. (2020). Monitoring comprehension in a foreign language: Trait or skill? *Metacognition and Learning*. https://doi.org/10.1007/s11409-020-09245-5

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97*(6), 1115–1128. https://doi.org/10.1037/a0017048

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D.

Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187-214). Springer International Publishing.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. do: 10.1002/per.2137

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Ed.), *Advances in motivation science Vol. 7* (pp. 161–198). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.05.002

Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly, 14*(2), 101–120. https://doi.org/10.1037/h0089000

Willard, G., & Gramzow, R. H. (2009). Beyond oversights, lies, and pies in the sky: Exaggeration as goal projection. *Personality and Social Psychology Bulletin, 35*(4), 477–492. https://doi.org/10.1177/0146167208329631

Wright, S. (2000). Looking at the self in a rose-colored mirror: Unrealistically positive self-views and academic performance. *Journal of Social and Clinical Psychology, 19*(4), 451-462. https://doi.org/10.1521/jscp.2000.19.4.451

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09600-1

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). Cambridge University Press.

Supplemental Material

Overview

Supplement 1. Computation of double-entry intraclass correlations

Supplement 2. Results and discussion of double-entry intraclass correlations

Supplement 3. Criteria for outliers

Supplement 4. Detailed explanations of the 15 analyzed models

Supplement 5. R syntax

Supplement 1. Computation of double-entry intraclass correlations

The goal of the analysis of double-entry intraclass correlations was to examine whether the inclusion of gender, parental educational level, and number of books hat home substantially altered the content of the self-evaluation and objective competence terms in our models. If that were the case, it would be difficult to interpret what construct these variables actually reflect and thus, their effects on *t*2 grades could not readily be interpreted as effects of self-evaluation and objective competence. In that case, the models would not reflect the respective SE bias rationales anymore. Thus, we computed double-entry intraclass correlations (r_{ICC}; Vize et al., 2018). First, we computed two residual scores ("residual 1" and "residual 2") each for the linear self-evaluation and the linear objective competence term, resulting in four different residual scores ("self-evaluation residual 1", "self-evaluation residual 2", " objective competence residual 1", and "objective competence residual 2"). Self-evaluation residual 1 was computed by partialling out all other model predictors from linear self-evaluation (quadratic self-evaluation, self-evaluation and objective competence interaction, linear objective competence, quadratic objecetive competence, *t*1 grades, gender, parental educational level, and number of books hat home). Likewise, objective competence residual 1 was computed by partialling out all other predictors from linear objective competence. Self-evaluation residual 2 and objective competence residual 2 were computed in the same way, except that gender, parental educational level, and number of books at home were not partialled out. Therefore, the only difference between the two residuals of the same construct is whether these three variables have been partialled out. Second, we computed the correlations of all four residual scores with other variables in the data set representing other constructs (e.g., neuroticism). These variables have been assessed in the same research project the present study stems from. For a list of these variables and the measurement instruments used for their assessment, see Table S1. Third, we computed the double-entry intraclass correlations between the correlations of the different residual scores with the other variables.

Thus, the double-entry intraclass correlations reflect how similar the correlations, for example, of self-evaluation residual 1 and self-evaluation residual 2, with other variables are. If these two residuals show highly similar correlations with other variables in the data set, it indicates that they are embedded in a similar nomological network and can be interpreted in the same or at least a similar way. If the correlations of self-evaluation residual 1 and self-evaluation residual 2 were dissimilar, that would limit the interpretability of our results and, depending on the severity of the problem, we might have to exclude gender, parental educational level, and/or number of books at home from the analysis. The double-entry intraclass correlations provide an overall estimation of the similarity between these correlations in a single score. Thus, high double-entry intraclass correlations for the residuals of the same construct are an indicator that they reflect the same content.

Table S1

Constructs correlated with self-evaluation and objective competence residuals and their measurement instruments

Construct	Measurement instrument
Neuroticism Extraversion Openness Agreeableness Conscientiousness	NEO five factor inventory (NEO-FFI; Borkenau & Ostendorf, 2008)
Math intrinsic value Math utility value Math attainment value Math expectancy of success	Scale assessing subjective educational task values (SESSW; Steinmayr & Spinath, 2010)
Math test anxiety agitation Math test anxiety worry	Short version (Schwarzer & Jerusalem, 1999) of the German test anxiety inventory (TAI-G; Hodapp,1991,1996)
Practical/technical orientation Intellectual/researching orientation Artistic/linguistic orientation Social orientation Business orientation Conventional orientation	General interest structure test with environment structure test (AIST-R/UIST-R; Bergmann & Eder, 2005)
Basic Arithmetic skills	Additional test for assessing basic arithmetic skills (KRW) published together with the DEMAT 9 (Schmidt et al., 2013)
General Intelligence	CFT 20-R with WS/ZF-R. General Intelligence scale 2 – Revision (CFT 20-R) with vocabulary (WS) and numerical order test – Revision (WS/ZF-R) (CFT20-R; Weiß, 2006)
General Intelligence	PSB-R 6-13 – Inspection system for school and education consultation for Grades 6 to 13 – revised version (PSB 6-13; Horn et al., 2003)

Table S1 - continued

Construct	Measurement instrument
Perceptual Speed	The Connecting numbers test (ZVT). A language free intelligence test for measuring "cognitive speed". Manual instruction, 2. revised version (ZVT; Oswald & Roth, 1987)

Supplement 2. Results and discussion of double-entry intraclass correlations

See Table S2 for the correlations of the two self-evaluation and objective competence residuals with other variables. Inspecting the correlations of residuals 1 and 2 of the same construct with other variables revealed that these correlations were similar (all $\Delta r \leq .17$). The double-entry intraclass correlation for the correlations of the self-evaluation residuals was very high ($r_{ICC} = .99$) and the one for the correlations of the objective competence residuals was high ($r_{ICC} = .83$). Therefore, the results suggest that partialling out gender, parental educational level, and number of books hat home did not substantially alter the content of the linear self-evaluation and competence terms and thus, including them as predictors in our models seems unproblematic. The finding that the double-entry intraclass correlation for self-evaluation was substantially higher than for objective competence was not unexpected. Of the 21 additional variables correlated with the residuals, 17 reflected personality traits like self-evaluation and only four reflected performance constructs like objective math competence. The 17 personality variables generally correlated more strongly with self-evaluation than objective competence. Thus, it seems plausible that the double-entry intraclass correlation for self-evaluation was higher.

Table S2

Correlations between residualized linear self-evluation and objective competence scores and related variables and their double-entry intraclass correlations (r_{ICC})

	Linear self-evaluation		Linear objective competence	
	Residual 1	Residual 2	Residual 1	Residual 2
Neuroticism	-.11*	-.14**	.01	-.09*
Extraversion	.04	.02	-.06	-.02
Openness	-.03	-.07	-.06	-.06
Agreeableness	-.06	-.05	-.07	-.09*
Conscientiousness	.06	.05	-.07	-.11*
Math intrinsic value	.53***	.55***	.01	-.01
Math utility value	.38***	.39***	-.07	-.04
Math attainment value	.35***	.32***	.00	-.04
Math expectancy of success	.41***	.42***	.02	-.00
Math test anxiety agitation	-.29***	-.32***	-.02	-.10*
Math test anxiety worry	-.23***	-.29***	-.04	-.12**
Practical/technical orientation	.20***	.26***	-.02	.15**
Intellectual/researching orientation	.16**	.21***	-.02	.09*
Artistic/linguistic orientation	-.14**	-.21***	-.04	-.08
Social orientation	-.12**	-.14**	-.02	-.15**
Business orientation	-.04	-.06	.02	-.02
Conventional orientation	.04	.05	.04	.05
Basic arithmetic	.13**	.16***	.32***	.33***
General intelligence (CFT)	.07	.06	.24***	.24***
General intelligence (PSB)	.06	.02	.29***	.23***
Perceptual speed	.13**	.09	.11*	.03

Table S2 - continued

	Linear self-evaluation		Linear objective competence	
	Residual 1	Residual 2		Residual 1
Similarity indices (r_{ICC})				
Linear self-evaluation residual 2	.99***	-		
Linear objective competence residual 1	.06	.07	-	
Linear objective competence residual 2	.24	.27	.83***	-

Note. N = 504; Residual 1 = residual after partialling out all other model variables (linear term of the respective other construct, e.g., self-evluation from objective competence and vice versa, quadratic self-evaluation, quadratic objective competence, self-evaluation and objective competence interaction, t1 grades, gender, parental educational level, number of books hat home); Residual 2 = residual after partialling out all other model variables except gender, parental educational level, and number of books hat home.

Supplement 3. Criteria for outliers

We used the same criteria for outliers as Humberg et al. (2019; p. 847). Thus, we considered three indicators, Cook's distance (D), the dfFit value, and the hat value. A multivariate outlier was defined as a case in which all three of the following conditions applied:

(1) D > 50th percentile of F(k, n - k)

(2) $|dfFit| > 3\sqrt{\frac{k}{n-k}}$

(3) $hat > \frac{3k}{n}$

Here, n denotes the sample size and k denotes the number of estimated parameters in the full regression model. According to these criteria, there were no multivariate outliers in our analyses. While the hat value is a measure of leverage, Cook's distance and the dfFit value are global indicators that measure the effect of deleting a certain observation (Cohen et al., 2003).

Supplement 4. Detailed descriptions of the 15 models in the initial model set

Using response surface analyses (RSA; Edwards, 2002), we have analyzed five models that correspond to five different hypotheses about SE bias effects (called *theory-derived models* in the following) as well as 10 *additional models* that have been added for methodological reasons (i.e., to avoid overestimating the likelihood of the five theory-derived models to be the best model given the data and the competing models). One of these models is the full model; each of the other 14 models has been constructed through parameter constraints of the full model. The parameter constraints that we will discuss in the following have also been presented in Figure 1 in the main text alongside visual representations of the models. In the following, we will explain the logic behind the parameter constraints of each model in more depth, with the aim to provide methodologically interested readers with detailed background information.

The full model is the following polynomial regression model:

(1) $A = b_0 + b_1S + b_2C + b_3S^2 + b_4SC + b_5C^2 + b_6(t1\text{ math grades}) + b_7(\text{gender}) + b_8(\text{parental educational level}) + b_9(\text{number of books at home}) + \varepsilon$

Here, A denotes the student's academic achievement, S denotes the self-evaluation score (i.e., self-evaluation of math competence), and C denotes the objective competence score (i.e., score on the standardized math competence test DEMAT 9). Therefore, academic achievement is predicted by a polynomial regression onto two linear terms (S and C), two quadratic terms (S^2 and C^2) and an interaction term between S and C (SC), as well as several additional linear predictors (*t*1 math grades, gender, parental educational level, number of books at home). We added these additional predictors to control for their influence on the other five predictors predictive value on A. They have been added in all constrained models, whereby their regression weights have been freely estimated in all models. However, for the sake of brevity, we will exclude the additional predictors from the equations in the models presented below because they are not vital for the understanding of the logic of RSA. As has been shown in OSM 1 and OSM 2, adding these additional predictors has not substantially altered the interpretation of S an C. Also for the sake of brevity, we will exclude the intercept, b_0, and the error term, ε, from the following equations. At the end of the chapter, we will provide a brief explanation why this does not affect the following argumentation.

Thus, we will explain the parameter constraints in the constrained models with respect to the following full model:

(2) $A = b_1S + b_2C + b_3S^2 + b_4SC + b_5C^2$

This equation will be the starting point from which all other models are derived. In the following sections, we explain the 15 models in detail. First, we will explain why the five theory-derived models actually do reflect what the respective theories posit. Second, we will explain the logic of the 10 additional models.

S4.1 Models reflecting theoretical rationales

S4.1.1 Beneficial SE bias model

The beneficial SE bias hypothesis posits a positive linear effect of SE bias on academic achievement. The respective model is derived from the full model through the following parameter restrictions made in equation (2):

(3) $b_1 > 0$; $b_2 < 0$; $b_3 = b_4 = b_5 = 0$

Thus, the regression weight of the linear self-evaluation term (S) is constrained to be positive, the regression weight of the linear objective competence term (C) is constrained to be negative, and the regression weights of the nonlinear terms are constrained to be zero. Why does this constrained model represent the beneficial SE bias hypothesis? The beneficial SE bias hypothesis posits that an increase in the difference S – C is associated with an increase in academic achievement (A). The difference S – C increases if either S increases or C decreases. In other words, when S increases, S – C increases and when C decreases, S – C increases. Therefore, if the difference S – C is positively associated with A, it follows that an increase in S and a decrease in C must be positively associated with A. Thus, b_1, the regression weight of S, is constrained to be positive and b_2, the regression weight of C, is constrained to be negative. Since the beneficial SE bias hypothesis does not posit any nonlinear effects on A, the nonlinear terms are constrained to be zero. This results in a regression surface as seen in Figure 1 in the main text where higher values of S and lower values of C are positively associated with A.

S4.1.2 Detrimental SE bias model

The detrimental SE bias hypothesis posits a negative linear effect of SE bias on academic achievement and the respective model can be derived from equation (2) through the following parameter restrictions:

(4) $b_1 < 0$; $b_2 > 0$; $b_3 = b_4 = b_5 = 0$

Thus, the regression weight of the linear self-evaluation term (S) is constrained to be negative, the regression weight of the linear objective competence term (C) is constrained to be positive and the regression weights of the nonlinear terms are constrained to be zero. In this model, the signs of b_1 and b_2 have been swapped compared to the beneficial SE bias model. According to the detrimental SE bias hypothesis, an increase in the difference S – C is associated with a decrease in academic achievement (A) which is the same as saying that a decrease in S – C is associated with an increase in A. S – C decreases if S decreases or C increases. Thus, with the same logic as for the beneficial SE bias hypothesis, it follows that a decrease in S and increase in C must be associated with an increase in A and therefore, b_1 is constrained to be negative and b_2 is constrained to be positive, while the other predictors are constrained to be zero.

S4.1.3 Self-Knowledge model

The self-knowledge hypothesis posits that both a positive and a negative SE bias is detrimental for academic achievement and thus, an SE bias of zero is optimal. The respective model can be derived from equation (2) through the following parameter restrictions:

(5) $b_1 = b_2 = 0$; $b_3 = b_5 < 0$; $b_3 + b_4 + b_5 = 0$

Thus, the regression weights of the linear self-evaluation term (S) and the linear objective competence term (C) are constrained to be zero, the regression weights of the quadratic self-evaluation and objective competence terms (S^2 and C^2) are constrained to be equal and negative and the sum of all nonlinear terms (S^2, C^2, and SC) is constrained to be zero. In the following, we will show why these constraints are in accordance with the self-knowledge hypothesis. The last constraint, $b_3 + b_4 + b_5 = 0$, can be transformed into the following equation:

(6) $b_4 = -(b_3 + b_5)$

By applying the first constraint, $b_1 = b_2 = 0$ to the full model (see equation (2)) and substituting $-(b_3 + b_5)$ for b_4 we get the following model:

(7) $A = 0S + 0C + b_3S^2 – (b_3 + b_5)SC + b_5C^2$

Since b_3 and b_5 are constrained to be equal, we can substitute b_5 for b_3 and factorize b_3:

(8) $A = b_3(S^2 – 2SC + C^2)$

Using the binomial theorem, we can rewrite equation (8):

(9) $A = b_3(S – C)^2$

Thus, in the self-knowledge model, A is a product of two factors, b_3 and $(S – C)^2$. Since b_3 is constrained to be negative and $(S – C)^2$ cannot have negative values, A cannot be positive. The largest value A can take is zero, which is reached when S is equal to C. Additionally, because b_3 is negative, A necessarily decreases whenever $(S – C)^2$ increases. In other words, this regression model implies that academic achievement is maximal when the self-evaluation equals the objective competence and decreases with increasing values in the squared difference between them. Squaring the difference is important because the self-knowledge model posits that A decreases with an increasing absolute difference of S and C no matter if the self-evaluation is larger than the objective competence or vice versa. Thus, this model produces a response surface as depicted in Figure 1 in the main text, where the maximum is reached for all points at which S and C are equal and decreases quadratically with an increasing absolute difference S – C.

S4.1.4 Optimal margin model

The optimal margin hypothesis posits that a certain positive value of SE bias is optimal for academic achievement but increasing the SE bias even further is detrimental. The respective model can be derived from equation (2) through the following parameter restrictions:

(10) $b_1 + b_2 = 0; b_1 - b_2 > 0; b_3 = b_5 < 0; b_3 + b_4 + b_5 = 0$

Thus, the only differences between the self-knowledge model and the optimal margin model are the constraints regarding b_1 and b_2 (see equation (5)). In the self-knowledge model, b_1 and b_2 are constrained to be zero. This is not the case in the optimal margin model. Applying only the constraints regarding b_3 to b_5, which are equal across both models, to the full model and making the same transformations as for the self-knowledge model, we get:

(11) $A = b_1S + b_2C + b_3(S - C)^2$

This equation is the same as equation (9) for the self-knowledge model, except that b_1 and b_2 are not constrained to be zero. The first constraint of the optimal margin model, $b_1 + b_2 = 0$ can be rewritten as $b_2 = -b_1$. Substituting $-b_1$ for b_2 in equation (11) and factorizing b_1 we get:

(12) $A = b_1(S - C) + b_3(S - C)^2$

While b_3 is constrained to be negative, the second constraint of the optimal margin model, $b_1 - b_2 > 0$, guarantees that b_1 is positive because otherwise it would not be possible for their sum to be zero as specified in the first constraint. The optimal margin hypothesis posits that there is a certain positive SE bias for which academic achievement is maximal. Thus, in order to show that the self-knowledge model actually reflects this hypothesis, we need to proof that the global maximum for A occurs at a certain positive difference S – C. Since calculating the global maximum for equation (12) exactly requires advanced methods such as partial derivatives, it will be omitted here. However, it is easy to show that the global maximum must indeed occur at some positive difference S – C. If S is equal to C, A in equation (12) is zero. If S – C is negative, A is smaller than zero because b_1 is positive and b_3 is negative and thus, both summands in equation (12) are negative since for both summands a negative value is multiplied with a positive value. Therefore, the global maximum for A cannot occur at a negative difference S – C. If S – C is positive, $b_1(S – C)$ will also be positive and $b_3(S – C)^2$ will be negative. Thus, A can be positive only if $b_1(S – C)$ is larger than the absolute value of $b_3(S – C)^2$. Since $b_3(S – C)^2$ is negative, its absolute value can be computed by multiplying with -1 and therefore:

(13) $b_1(S – C) > -b_3(S – C)^2$

Thus, A can only have positive values at the points at which equation (13) is true. Equation (13) can be transformed into the following equation:

(14) $–b_1/b_3 > S – C$

Since we know that A can also only have positive values for positive differences S – C, we can conclude that A has positive values exactly at all points at which the following equation is true:

(15) $–b_1/b_3 > S – C > 0$

Since b_1 is positive and b_3 is negative, $–b_1/b_3$ is positive. In other words, all positive values of A lie within a certain range for the difference S – C that is given by equation (15) and this range only includes positive values of S – C. Therefore, the global maximum of A must occur at a positive difference S – C. Additionally, for differences S – C greater than $–b_1/b_3$, A becomes negative again. Therefore, A must decrease again, after a certain “optimal” value of S – C, which is what the self-knowledge hypothesis posits. This is also why the optimal margin model produces a response surface as depicted in Figure 1 in the main text where all highest values of A lie on a line that consists of all points with a certain positive difference S – C.

S4.1.5 Beneficial self-evaluation and competence model

The beneficial self-evaluation and competence hypothesis posits that there are positive linear effects of the self-evaluation and objective competence on academic achievement but no SE bias effects. The respective model can be derived from equation (2) through the following parameter restrictions:

(16) $b_1 > 0$; $b_2 > 0$; $b_3 = b_4 = b_5 = 0$

It is easy to see, why this model reflects the beneficial self-evaluation and competence hypothesis. Because this hypothesis posits that there are only positive linear effects of self-evaluation and objective competence on academic achievement, b_1 and b_2 are constrained to be positive, while all other regression weights are constrained to be zero. This results in a model, in which A increases with increasing values of S and C, while the difference S – C has no impact on A.

S.4.2 Models added for methodological reasons

To our knowledge, the models addressed in this section do not reflect any theoretical rationales on SE bias effects on academic achievement in the literature. Therefore, unlike for the models discussed above, we do not explain, why these models represent a certain rationale. However, we still explain, what kind of association of the self-evaluation and competence with academic achievement these models imply.

S4.2.1 Beneficial or detrimental self-evaluation or competence models

In this section, we summarize four models: the beneficial self-evaluation model, the detrimental self-evaluation model, the beneficial competence model, and the detrimental competence model. In these models, the regression weight of either the self-evaluation or objective competence is constrained to be either positive (beneficial models) or negative (detrimental models) and all other regression weights are constrained to be zero.

Specifically, the constraints for the beneficial self-evaluation model are:

(17) $b_1 > 0;\ b_2 = b_3 = b_4 = b_5 = 0$

The constraints for the detrimental self-evaluation model are:

(18) $b_1 < 0;\ b_2 = b_3 = b_4 = b_5 = 0$

The constraints for the beneficial competence model are:

(19) $b_2 > 0;\ b_1 = b_3 = b_4 = b_5 = 0$

The constraints for the detrimental competence model are:

(20) $b_2 < 0;\ b_1 = b_3 = b_4 = b_5 = 0$

Thus, these models each imply, that there is only one linear association of either the self-evaluation or objective competence with academic achievement (either positive or negative), while the respective other variable does not have any association with academic achievement. There are also no nonlinear associations of the self-evaluation or objective competence with academic achievement in these models.

S4.2.2 Detrimental self-evaluation and competence model

In this model, there are negative linear effects of the self-evaluation and objective competence on academic achievement but no SE bias effects. It can be derived from equation (2) through the following parameter restrictions:

(21) $b_1 < 0;\ b_2 < 0;\ b_3 = b_4 = b_5 = 0$

Thus, this model can colloquially be understood as the "opposite" of the beneficial self-evaluation and competence model, where b_1 and b_2 are constrained to be negative instead of positive and b_3, b_4, and b_5 are still constrained to be zero.

S4.2.3 Curvilinear self-evaluation or competence models

In this section, we summarize the curvilinear self-evaluation model and the curvilinear competence model and explain them by the example of the curvilinear self-evaluation model, which is derived from equation (2) through the following parameter restrictions:

(22) $b_2 = b_4 = b_5 = 0;\ b_3 < 0$

Thus, the regression weights of the linear objective competence term, the quadratic objective competence term and the self-evaluation and objective competence interaction term are constrained to be zero. Therefore, there are no associations, linear or nonlinear, of objective competence with academic achievement. Additionally, the regression weight of the quadratic self-evaluation term is constrained to be negative. This means that while the regression weight b_1 of the linear self-evaluation term S, is estimated freely, the association of S and A will decrease or become (more) negative with increasing values of S because the regression weight b_3 of the quadratic term is constrained to be negative and grows faster than the linear term. The curvilinear competence model is derived from equation (2) through the following parameter restrictions.

(23) $b_1 = b_3 = b_4 = 0;\ b_5 < 0$

Thus, the association of C and A decreases or becomes (more) negative with increasing values of C, as it is the case for the association of S and A in the curvilinear self-evaluation model.

S.4.2.4 Interaction model

The interaction model is derived from equation (2) by the following parameter restrictions:

(24) $b_3 = b_5 = 0; b_4 > 0$

Thus, the regression weights of the linear terms are estimated freely, while the regression weights of the quadratic terms are constrained to be zero and the regression weight of the interaction term is constrained to be positive. This means that the association of C with A is more positive or less negative at higher values of S and likewise the association of S with A is more positive or less negative at higher values of C. We explain these associations in the following. Applying the parameter constraints in equation (24) to equation (2) we get:

(25) $A = b_1S + b_2C + b_4SC$ with $b_4 > 0$

We can rewrite equation (25) like this:

(26) $A = b_1S + (b_2 + b_4S)C$ with $b_4 > 0$

Thus, the slope of the linear association of C with A at each specific value of S is b_2+b_4S. Because b_4 is constrained to be positive, this means that the larger S, the larger (more positive or less negative) is the association of C with A. Equation (25) can also be rewritten in a different way:

(27) $A = b_2C + (b_1 + b_4C)S$ with $b_4 > 0$

Thus, following the same argument as for the association of C with A, the slope of the linear association of S with A increases with increasing values of C.

S.4.2.5 Null Model

In the null model, all regression weights in equation (2) are constrained to be zero and thus, there are no associations of S or C with A:

(28) $b_1 = b_2 = b_3 = b_4 = b_5 = 0$

S4.2.6 Full model

In the full model no parameter constraints are applied to equation (2). Thus, all quadratic, linear and interactive associations of S and C with A are possible, including but not limited to the associations described in any of the above models.

S4.3 Validity of the models with additional predictors

In S4.1 and S4.2, we have explained why each of the 15 models included in our analyses represents a certain association of S and/or C with A. However, the validity of some statements made in S4.1 and S4.2 (e.g., that in the self-knowledge model, A cannot be positive) hinges on the exclusion of some terms from equation (1) for the sake of brevity. These terms are the intercept b_0, the additional predictors b_6(t1 math grades), b_7(gender), b_8(parental educational level), and b_9(number of books at home), and the error term ε. Adding these terms means that, for example, A in the self-knowledge model *can* be positive. In other words, we have only shown why these models represent certain associations of S and/or C with A in the specific case where these additional terms are omitted. We chose this approach, because it made the necessary discussion much shorter and simpler. In this section, we will explain why all models described above indeed describe the same associations of S and/or C with A, even if the additional terms are not zero.

We chose equation (9) of the self-knowledge model as an example to illustrate this. In S4.1.3 we argued that A in equation (9) and therefore in the self-knowledge model cannot have positive values. If we include all additional terms, equation (9) becomes:

(29) $A = b_0 + b_3(S - C)^2 + b_6(t1\text{ math grades}) + b_7(\text{gender}) + b_8(\text{parental educational level}) + b_9(\text{number of books at home}) + \varepsilon$

Therefore, A *can* actually be positive because of the addition of the intercept and the additional predictors. However, this does not affect the logic of our argumentation. Since the intercept is a constant, it only shifts the regression surface, and thus A, by a certain constant but it does not affect the associations between the other variables. By including b_0 (while ignoring the additional predictors) the maximum of A becomes b_0 instead of zero but all regression weights will be equal to the specific case of $b_0 = 0$. In other words, the shape of the response surface is the same. It still has its maximum at all points at which S and C are equal and decreases quadratically with the absolute difference between S and C. A is just shifted by a certain constant.

On the other hand, including the additional predictors can also change the regression weights of S and C because of the shared variance between the additional predictors and S, C, and A. However, while the regression weights of S and C can change, the parameter constraints made in the respective models still apply and thus, the regression weights cannot take values that are prohibited by these constraints (e.g., when considering equation (9), even after adding the additional predictors as in equation (29), b_3 will always be negative). Another aspect that is important to consider is that there are no interactions between the additional predictors and S and/or C in any of our models. Adding four additional predictors changes what has been a two-dimensional response surface of the association of S and C with A two a six-dimensional response space of the association of S, C, and the four additional predictors with A. By setting each of the additional predictors to a certain value after the computation of the regression, we again obtain a response surface for the association of S, C, and A. Thus, for each of the infinitely many combinations of values of the additional predictors, we obtain one response surface of S, C, and A. However, because the additional predictors do not interact with S or C, the shape of all the infinitely many response surfaces will be the same. They are only shifted by different values, because different values of the additional predictors (each multiplied by its specific regression weight) are added to all points of the response surface.

To summarize how including the additional predictors in the regression model can change the response surface of S, C, and A compared to the case where the additional parameters are omitted: 1) It *can* change the value of the regression weights of S and/or C. 2) It *cannot* change the regression weights to values that are not allowed by the parameter constraints made in the respective models. 3) It *cannot* result in response surfaces of the association of S and C with A whose shapes differ at different values of the additional predictors. Therefore, the only potential problem we have to worry about is whether the change in regression weights of S and C changes the interpretation of the underlying theoretical effects of self-evaluation and objective competence on academic achievement. However, since the inclusion of additional predictors does not seem to have changed the content of the self-evaluation and objective competence scores substantially (see OSM 1 and OSM 2), this does not seem to a problem in the present study. Therefore, despite the fact that we included additional predictors in our models, the resulting response surfaces of S, C, and A can still be interpreted with the logic of RSA. For example, the beneficial self-evaluation and competence model can be interpreted as a model of positive linear effects of the self-evaluation and objective competence on academic achievement after controlling for the effects of the *t*1 grades, gender, parental educational level and number of books at home.

Supplement 5. R syntax

The R syntax for the statistical analyses in the present study are freely available at: osf.io/faz3y

References

Bergmann, C., & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (AIST-R/UST-R) [General Interest Structure Test with Environment Structure Test (AIST-R/UST-R)].* Beltz Test.

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI) - Manual [NEO Five Factor Inventory (NEO-FFI) by Costa and McCrae - Manual], 2nd ed. (rev).* Hogrefe.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.).* Erlbaum Publishers.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). Jossey-Bass.

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The test anxiety inventory TAI-G: An extended and modified version with four components]. *Zeitschrift für Pädagogische Psychologie, 5*, 121–130.

Hodapp, V. (1996). The TAI-G: A multidimensional approach to the assessment of test anxiety. In C. Schwarzer, & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 95-130). Francke.

Horn, W., Lukesch, H., Mayrhofer, S., & Kormann, A. (2003). *PSB 6-13 – Prüfsystem für Schul- und Bildungsberatung für 6. Bis 13. Klassen – revidierte Fassung [PSB-R 6-13 – Inspection system for school and education consultation for Grades 6 to 13 – revised* *version].* Hogrefe.

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859. https://doi.org/10.1037/pspp0000204

Oswald, W. D., & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT). Ein sprachfreier Intelligenz-Test zur Messung der „kognitiven Leistungsgeschwindigkeit". Handanweisung. 2., überarbeitete und erweiterte Auflage [The Connecting numbers test (ZVT). A language free intelligence test for measuring "cognitive speed". Manual instruction, 2. revised version].* Hogrefe.

Schmidt, S., Ennemoser, M., & Krajewski, K. (2013). *Deutscher Mathematiktest für neunte Klassen (DEMAT 9) [German Mathematics Test for Ninth Grade (DEMAT 9)].* Hogrefe.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for the assessment of teacher and student attributes. Documentation of psychometric instruments as part of the scientific support of the pilot project self-efficient schools].* Freie Universität Berlin.

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and first validation of a scale assessing subjective educational task values (SESSW)]. *Diagnostica, 56*(4), 195-

211. doi:10.1026/0012-1924/a000023

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. do: 10.1002/per.2137

Weiß, R. H. (2006). *CFT 20-R mit WS/ZF-R Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R) [CFT 20-R with WS/ZF-R. General Intelligence scale 2 – Revision (CFT 20-R) with vocabulary (WS) and numerical order test – Revision (WS/ZF-R)]*. Hogrefe.

3. Gesamtdiskussion

3.1 Weiterführende Analyse (Beitrag III)

The effects of self-evaluation, competence, and their discrepancy on academic achievement in math mediated by expectancy of success and subjective task values

Paschke, P. & Steinmayr, R. (2023). *The Effects of Self-Evaluation, Competence, and their Discrepancy on Academic Achievement in Math Mediated by Expectancy of Success and Subjective Task Values [Die Effekte von Selbsteinschätzung, Kompetenz und ihrer Diskrepanz auf akademische Leistung in Mathematik mediiert durch Erfolgserwartung und subjektive Werte].* [Manuscript in preperation for publication]. Institut für Psychologie, TU Dortmund University

Acknowledgements: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Note: This article is not the copy of record and may not exactly replicate the final, authoritative version of the article.

Abstract

According to situated expectancy-value theory (SEVT), ability self-concepts (ASC) affect academic achievement positively through pathways of expectancy of success and subjective task values. However, this proposition has not been tested longitudinally. Additionally, little is known about the influence of the discrepancy between ASC and actual competence (self-estimation bias) on academic achievement beyond the main effect of the ASC per se. In a longitudinal study with 504 10th grade students (*t*0: second term 9th grade, *t*1: first term 10th grade, *t*2: first term 11th grade, *t*3: second term 11th grade), we tested whether math ASC, actual math competence (assessed at *t*1), and the discrepancy between the two affect subsequent math academic achievement (operationalized as math report card grades assessed at *t*0 and *t*3) directly and via the pathway of expectancy of success and subjective task values in math (assessed at *t*2). Results indicated that competence but not ASC affected subsequent grades. The effect of competence was partially mediated by expectancy of success. There were no self-estimation bias effects on grades.

Keywords: Ability-self concept, academic achievement, expectancy-value theory, self-evaluation bias, response surface analysis

Zusammenfassung

Nach der Erwartungs-Wert-Theorie (SEVT) beeinflusst ein hohes Fähigkeitsselbstkonzept akademische Leistung positiv, wobei dieser Effekt durch höhere Erfolgserwartung und subjektive Werte mediiert wird. Allerdings wurde diese Mediationsannahme bisher nicht längsschnittlich untersucht. Zudem ist wenig darüber bekannt, wie sich die Diskrepanz zwischen dem Fähigkeitsselbstkonzept und der tatsächlichen Kompetenz (Selbsteinschätzungsbias) über den Haupteffekt eines hohen Fähigkeitsselbstkonzepts hinaus auf akademische Leistung auswirkt. Wir untersuchten in einer Längsschnittstudie mit 504 Zehntklässler*innen (*t*0: Zweites Halbjahr neunte Klasse, *t*1: Erstes Halbjahr 10. Klasse, *t*2: Erstes Halbjahr 11. Klasse, *t*3: Zweites Halbjahr 11. Klasse), ob sich das Fähigkeitsselbstkonzept und die Kompetenz in Mathematik (erfasst zu *t*1) sowie deren Diskrepanz auf die akademische Leistung der Schüler*innen auswirkte (operationalisiert als Zeugnisnoten in Mathematik zu *t*0 and *t*3). Untersucht wurden direkte Effekte sowie indirekte Effekte vermittelt über die Erfolgserwartung und subjektiven Werte in Mathematik zu *t*2. Die Mathematikkompetenz nicht aber das Fähigkeitsselbstkonzept in Mathematik hatte einen Effekt auf die nachfolgenden Mathematiknoten. Dieser Kompetenzeffekt wurde partiell mediiert durch die Erfolgserwartung. Es wurden keine SE Bias Effekte auf Noten gefunden.

Stichwörter: Fähigkeitsselbstkonzept, akademische Leistung, Erwartungs-Wert-Theorie, Selbsteinschätzungsbias, response surface analysis

The Effects of Self-Evaluation, Competence, and their Discrepancy on Academic Achievement in Math Mediated by Expectancy of Success and Subjective Task Values

Thinking highly of own competences and abilities has positive consequences for students' academic achievement (e.g., grades, test results, attainment; e.g., Marsh et al., 2022; Valentine et al., 2004; Wu et al., 2021). Meta-analyses suggest that these effects hold for various different forms of self-evaluations (e.g., ability self-concept, self-efficacy, self-esteem; Talsma et al., 2018; Valentine et al., 2004; Wu et al., 2021). According to *situated expectancy-value theory* (SEVT; Eccles & Wigfield, 2020; Wigfield & Eccles, 2020), this positive influence is mediated by students' expectancy of success and subjective task values. Specifically, the model posits that ability self-concepts (ASCs) positively affect expectancy of success and subjective task values, which in turn positively affect academic choices and achievement. However, while the effects of expectancy of success and subjective task values on academic achievement are well-established (e.g., Brown & Putwain, 2022; Steinmayr et al., 2019), their mediating role between ASC and academic achievement has not been tested longitudinally. Additionally, whereas positive effects of high ASCs are well-established, there is an ongoing discussion if the bias between self-evaluated and actual competence (self-evaluation bias; SE bias) also has an effect on academic achievement (e.g., Leduc & Bouffard, 2017; Lee, 2021). Moreover, proponents of SE bias effects often argue that these effects are mediated by motivational variables such as intrinsic motivation as well (e.g., Bonneville-Roussy et al., 2017; Lee, 2021), although this mediation has not been tested empirically. Summing up, we know that high ASCs positively affect academic achievement, but we still do not know for sure 1) whether the bias between ASC and actual competence also has an effect on academic achievement (for first studies see authors, 2020; 2023) and 2) whether the effects of ASC and potentially SE bias are mediated by motivational variables such as expectancy of success and subjective task values. Thus, the present study has the central goal to help answer the two aforementioned research questions.

ASC Effects on Academic Achievement and Expectancy Value Theory

Evidence from research on the reciprocal effects model (Marsh et al., 2022; Marsh & Yeung, 1997) shows that the link between ASC and academic achievement stems from the fact that the two mutually affect each other (e.g., Bakadorova & Raufelder, 2020; Marsh, 2022; Marsh et al., 2022; Preckel et al., 2017; Seaton et al., 2015; Sewasew & Koester, 2019; Sewasew & Schroeders, 2019; Weidinger et al., 2018; Wu et al., 2021). Meta-analyses of longitudinal studies have found a small beneficial effect of ASC on academic achievement (Valentine et al., 2004; Wu et al., 2021). Additionally, there is evidence that this effect holds even when competence measured on a standardized test is controlled (authors, 2020; 2023). Thus, there is ample evidence that high ASCs positively affect academic achievement.

According to SEVT (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020), a person's ASC affects both, the person's expectancy of success at tasks for which the respective competence is necessary, as well the subjective values the person assigns to those tasks. Expectancy of success and subjective task values are considered highly subject specific (Eccles et al., 1993; Gaspard et al., 2018; Wigfield & Eccles, 2000; Wigfield et al., 2020). The authors also distinguish between four subjective task value components (Eccles & Wigfield, 2020; Wigfield & Eccles, 2020; Wigfield et al., 2020): intrinsic value (subjective interest and enjoyment of the task), utility value (subjective relevance of the task for future goals), attainment value (subjective importance of doing well on a task), and costs (emotional and other psychological costs, opportunity costs, as well as the time and effort required to perform a task). However, most studies do not consider costs (but see Flake et al., 2015) and only concentrate on the other value components as we do in the present article (e.g., Weidinger et al., 2020). According to SEVT, expectancy of success and subjective task values in turn affect academic choices and achievement. The positive effect of expectancy of success and subjective task values on academic achievement is supported by empirical results (e.g., Brown & Putwain, 2022; Froiland & Davison, 2016; Geng et al., 2022; Steinmayr et al., 2019; Steinmayr & Spinath, 2009; Wang, 2012; Weidinger et al., 2020). However, to our knowledge, the mediating role of expectancy of success and subjective task values for the effect of self-concept on academic achievement has not been tested longitudinally. One reason for the low number of studies that consider ASC and expectancy of success simultaneously is that these constructs, even though conceptually distinct, are empirically strongly related (see Wigfield et al., 2020). Older and newer studies could not empirically distinguish them (Eccles et al., 1993; Eccles & Wigfield, 1995; Marsh et al., 2019). Consequentially, in more recent studies, many researchers treat these constructs interchangeably, often assessing expectancy of success by a measure of ASC (Gaspard et al., 2018; Guo et al., 2017; Jansen et al., 2021; Nagengast et al., 2011; Trautwein et al., 2012) or combining measures for the two constructs into a single score (Wang, 2012). Exceptions to this are the studies by Priess-Groben & Hyde (2017) and Mac Iver (1991). Priess-Groben & Hyde (2017) tested a mediation by expectancy of success for the effect of ASC on course-taking intentions instead of achievement. They found a significant mediation. However, the fit of the model was poor (p. 1327). Similarly, Mac Iver et al. (1991) analyzed the mediating role of subjective task values for the effect of ASC on academic effort. This study did not find a mediation. Whereas ASC affected subjective task values and effort directly, subjective task values did not affect effort in turn. Thus, despite the proposition in SEVT, there is little research and empirical evidence for expectancy of success and subjective task values mediating the effects of ASC on academic outcomes. By now, the authors of SEVT consider it "probably a mistake" (Eccles & Wigfield, 2020, p. 3) that the two constructs have been combined instead of focusing on ways to separate them empirically. Thus, in the present study, we attempted to assess ASC and expectancy of success as separate constructs and test the proposed mediating role of expectancy of success and subjective task values for the effect of ASC on academic achievement.

Self-Evaluation Bias Effects on Academic Achievement

In the previous sections, we cited evidence that high self-evaluations of competence are beneficial for academic achievement. However, all of those studies only considered linear main effects of the self-evaluation but not possible effects of the bias between self-evaluated and actual competence (SE bias). There is an ongoing discussion whether an SE bias has an effect on academic achievement beyond the main effect of the self-evaluation alone (e.g., authors, 2020; 2023; Bonneville-Roussy et al., 2017; Leduc & Bouffard, 2017; Lee, 2021; see Trautwein & Möller, 2016). Thus, whereas high ASCs are generally beneficial, it cannot be ruled out that this relation is more complex than a simple linear main effect. For example, the beneficial effect of high ASC might become weaker or even negative (and thus turn into a detrimental effect) if the ASC exceeds the actual competence too much (high positive SE bias). At least five competing hypotheses regarding SE bias effects on academic achievement can be derived from considerations in the literature. Four of these hypotheses propose different kinds of SE bias effects, whereas the fifth suggests that there might be no SE bias effects on academic achievement at all. In the following, we briefly discuss the four hypotheses positing different SE bias effects. Then, we turn to motivation as a possible mediator of these effects. Finally, we discuss the fifth hypothesis positing that there are no SE bias effects on academic achievement at all.

Self-Evaluation Bias Effects Hypotheses

Beneficial SE Bias Hypothesis. The most common hypothesis among researchers on SE bias effects is that having a high SE bias is beneficial for academic achievement. Indeed, a relatively large number of empirical studies found beneficial SE bias effects (e.g., Bonneville-Roussy et al., 2017; Bouffard et al., 2011; Chung et al., 2016; Côte et al., 2014; Dupeyrat et al., 2011; Helmke, 1998; Leduc & Bouffard, 2017; Lee, 2021; Lopez et al., 1998; Martin & Debus, 1998; Taylor & Brown, 1988; Wright, 2000). In other words, a larger difference of self-evaluation minus competence would lead to greater academic achievement. This implies that overestimation (self-evaluation > competence) is beneficial for academic achievement, while underestimation (self-evaluation < competence) is detrimental. Authors supporting this position often argue that a high SE bias positively affects students' motivation and thus ultimately academic achievement and performance (e.g., Bonneville-Roussy et al., 2017; Helmke, 1998; Lee, 2021; Martin & Debus, 1998; Taylor & Brown, 1988; Wright, 2000).

Optimal Margin Hypothesis. Some researchers who argue for generally beneficial effects of SE bias on academic achievement suppose that there might also be negative effects if the SE bias is very large (e.g., Baumeister et al., 1989; Helmke et al., 1998; Taylor & Brown, 1994). In other words, there would be a certain (positive) level of SE bias which is optimal for the development of academic achievement, whereas increasing the SE bias even further is detrimental. This view is based on the same general assumption as the beneficial SE bias hypothesis: that having a high SE bias fosters motivation and effort and thus academic achievement. However, it also encompasses the idea that very high levels of SE bias (and thus very unrealistically positive assumptions about own competences) can hinder motivation and learning efforts because students assume they have already learned enough (Baumeister et al., 1989; Taylor & Brown, 1994). Arguments for the beneficial SE bias hypothesis and optimal margin hypothesis often overlap since both positions stress a generally beneficial influence of SE bias. Few studies have tested this hypothesis and those that did, did not find any evidence for optimal margin effects on academic achievement (Lee, 2021; Lopez et al., 1998; Praetorius et al., 2016; Wright, 2000).

Self-Knowledge Hypothesis. Researchers from the field of metacognitive self-regulated learning stress the importance of knowledge about one's own competences in order to motivate and guide learning behavior (e.g., Buckelew et al., 2013; Dunlsoky & Rawson, 2012; Hacker & Bol, 2019; Händel & Fritzsche, 2016; Kim et al., 2010; Nietfeld & Schraw, 2002; Thiede et al., 2003). Thus, it seems plausible that accurate self-estimations might be most beneficial for academic achievement. However, it should be noted that research in this field is typically concerned with the more fine-grained perspective of the bias in judgements about learning success at particular tasks rather than the bias in more global self-evaluations of competence (i.e., ASC). Therefore, we do not aim to test the assumptions of meta-cognitive self-regulated learning. Rather, we test whether an accurate ASC is beneficial for academic achievement.

Detrimental SE Bias Hypothesis. Very few researchers have also suggested generally detrimental SE bias effects on academic achievement. Rohr and Ayers (1973) suggested that underestimation improves learning motivation, especially in already high achieving students, while overestimation leads to unrealistic study patterns, which hinders academic achievement. Thus, this position might be considered as the opposite of the beneficial SE bias hypothesis. Here, underestimation is beneficial, while overestimation is detrimental.

Summary. There are four hypotheses positing SE bias effects on academic achievement – a positive effect (beneficial SE bias hypothesis), a generally positive effect that turns negative if the SE bias is too large (optimal margin hypotheses), a reverse u-shaped effect where an SE bias of zero is optimal (self-knowledge hypothesis), and a negative effect (detrimental SE bias hypothesis). In all these hypotheses, motivational variables are considered as mediators between the SE bias and academic achievement.

Self-Estimation Bias and Motivation

To our knowledge, the mediating role of motivation for SE bias effects on academic achievement has not been tested empirically. Additionally, despite the common assumption that having a higher SE bias should be beneficial for motivation (e.g., Bonneville-Roussy et al., 2017; Bouffard et al., 2011; Côté et al., 2014; Helmke, 1998; Lee, 2021; Taylor & Brown, 1988), there are relatively few studies on this topic (see Lee, 2021).

Interest. Some researchers found that a positive SE bias in math is associated with an increased interest in math (Bench et al., 2015; Lee, 2021), whereas other authors could not replicate that finding but found the same effect for language (Gonida & Leondari, 2011). However, only the study by Lee (2021), which found a positive association, is longitudinal. The author also found that the positive effect of SE bias flattened out at very high levels of SE bias, especially for low-achieving students, constituting an optimal margin effect on interest instead of academic achievement.

Engagement and effort. Similar to the results for interest, Lee (2021) found that an SE bias in math positively predicts subsequent engagement in math and that this effect flattened out at very high levels of SE bias, especially for low-achieving students. Murphy et al. (2018) analyzed two different SE bias effects: an effect of SE bias in intelligence on academic effort in school and an effect of SE bias in sporting ability on sporting effort in co-curricular activities. Only the latter effect was significant.

Goal orientations. The two studies that investigated the association between SE bias and goal orientations found that SE bias is positively associated with performance-approach goals (Gonida & Leondari, 2011; Dupeyrat et al., 2011), whereas for mastery and performance-avoidance goals, some authors found positive (Gonida & Leondari, 2011) and some found no associations (Dupeyrat et al., 2011).

Persistence. Bonneville-Roussy et al. (2017) found positive SE bias effects on persistence in school in general, whereas Gonida and Leondari (2011) did not find an association of SE bias with persistence in math or language. However, persistence was operationalized differently in these studies. Gonida and Leondari (2011) operationalized persistence with self-report items assessing students' willingness to persist in learning a topic in the face of difficulties, whereas Bonneville-Roussy et al. (2017) used attitude towards dropout, actual dropout, and the highest sought degree as indicators of persistence.

Summary. Overall, a small number of studies with heterogeneous methods found either positive effects or no effects but not negative effects of SE bias on motivational outcomes. Thus, we cannot yet draw a conclusion on SE bias effects on motivation or the mediating role of motivation for SE bias effects on academic achievement.

Beneficial Self-Evaluation and Competence Hypothesis

Before, we have discussed four hypotheses positing different kinds of SE bias effects and the possible mediating role of motivation for these effects. However, we argue that there are also theoretical considerations as well as empirical results which suggest that there might be no SE bias effects on academic achievement at all. In other words, according to this hypothesis, having a high self-evaluation is beneficial for academic achievement irrespective of the person's actual competence and likewise, being highly competent is beneficial irrespective of the level of the self-evaluation. This position is based on the following considerations. Most theoretical arguments for SE bias effects in the literature fall short of explaining why it should be the bias in the self-evaluation rather than the self-evaluation per se that affects academic achievement. For example, Bonneville-Roussy et al. (2017) suggest that a high SE bias makes students perceive academic difficulties as challenges and persist longer, which would then positively affect academic achievement. However, by itself, this argument does not clarify why it should be the bias rather than the high self-evaluation per se that affects academic achievement. Imagine two students, Sarah and Julia. Sarah has a high self-evaluation of math competence and high actual math competence (SE bias near zero). Julia has an average self-evaluation of math competence and low actual math competence (positive SE bias). If the above argument by Bonneville-Roussy et al. (2017) is correct, Julia should benefit from her higher SE bias compared to Sarah. However, it does not become clear why this should be the case. Why should it not be Sarah who persists longer when facing difficulties because her self-evaluation is higher? It seems plausible that high self-evaluations make students persist at tasks because they believe they have the necessary competence to complete it successfully . We do not argue that such an SE bias effect cannot exist. However, we argue that so far there is hardly any convincing theoretical rationale for such an effect.

Other authors voiced similar criticism on SE bias theories (Colvin & Block 1994; Humberg et al., 2018; 2019; Robins & Beer, 2001). As discussed in the previous section, Bonneville-Roussy et al. (2017) did indeed find an SE bias effect on persistence and likewise, a relatively large number of studies found positive SE bias effects on academic achievement. These results appear to support the beneficial SE bias hypothesis. However, almost all of the studies made use of algebraic difference scores or residual scores to quantify the SE bias (e.g., Bonneville-Roussy et al., 2017; Chung et al., 2016; Dupeyrat et al., 2011; Gonida & Leondari, 2011; Leduc & Bouffard, 2017; Lee, 2021; Lopez et al., 1998; Praetorius et al., 2016; Wright, 2000). Both algebraic difference scores and residual scores confound the level of the SE bias with the level of the self-evaluation alone and algebraic difference scores are also confounded with the competence score as well (see Edwards & Parry, 2018; Humberg et al., 2018; 2019). Thus, since the methods used in former studies confound SE bias effects with self-evaluation effects and self-evaluation effects are well-established, any observed SE bias effect in these studies could be an artifact of a self-evaluation effect (see authors, 2020; 2023; Humberg et al., 2018; 2019). In other words, it is not clear whether the discrepancy between a self-evaluation of competence and actual competence has any effect on academic achievement beyond the main effect of the self-evaluation alone. This problem is long known (e.g., Asendorpf & Ostendorf, 1998; Colvin & Block, 1994; Edwards & Parry, 1993; Edwards, 2002) but has so far rarely been addressed in this particular research field. First studies that used different methods, which do not confound SE bias effects with self-evaluation effects, found evidence only for positive main effects of the self-evaluation and the competence but not for an SE bias (authors, 2020; 2022) and thus support the beneficial self-evaluation and competence hypothesis.

The Present Study

The aim of the present study is twofold. First, we aimed to comparatively evaluate the different hypotheses about self-estimation effects and SE bias effects on academic achievement (Research Question 1, RQ1). Second, we aimed to test whether the potential self-estimation effects and SE bias effects on academic achievement are mediated by expectancy of success and subjective task values (Research Question 2, RQ2).

We assessed students' ASCs in math. To assess math competence we used a standardized math competence test and we operationalized the outcome variable (math academic achievement) as later math grades. In-between, we assessed students' expectancies of success and their intrinsic, utility, and attainment values in math as possible mediators. Thus, we investigated how math ASC, math competence and the discrepancy between the two (SE bias) affect math grades directly and via the pathways of expectancy of success and subjective task values. Using response surface analysis (RSA; Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2019), we estimated models that correspond to the five aforementioned hypotheses about SE bias effects in order to test these hypotheses against each other. In each model, we also incorporated a set of control variables: prior math grades, prior scores in the respective mediator, gender, as well as two indicators of socio-economic status (SES): parental education and number of books at the students' home. Since math grades as well as the mediators are endogenous variables in our models, we controlled for their baseline levels at the previous measurement occasions. Additionally, gender (Heyder et al., 2019; Skaalvik & Skaalvik, 2004; Steinmayr & Spinath, 2008; Wilgenbusch & Merrell, 1999) and SES (OECD, 2016; Stubbe et al., 2016; Schwippert et al., 2020; Steinmayr et al., 2012) correlate with both math ASC and math academic achievement. Thus, we controlled for these variables to mitigate the possibility of observing a spurious effect of math ASC on math academic achievement due to the common variance with these variables. For example, it is possible that prior math ASC has a positive correlation with subsequent math academic achievement because a lower SES leads to both, lower math ASC and lower math academic achievement. Thus, the influence of gender and SES should be controlled.

Method

Participants and procedure

The present project is in line with ethical guidelines for human subject research. We did not mislead participants about the purpose of the study, there were no exclusion or inclusion criteria, treatments or questions that could cause mental or physical harm and the participants were not part of a vulnerable group. Participation in the study was voluntary and approved by the school administrations beforehand. Approval by an ethics committee was not required in the federal state in which the study was conducted. About 90% of the basic population participated at *t*1. Students not participating were ill, took part in a different extracurricular activity, or spent the semester abroad. We received informed consent forms from the parents of all participating students.

Participants were students from four 10th grades in academic-track schools (Gymnasium) located in the federal state North Rhine-Westphalia. The first measurement occasion took place in fall 2015 and fall 2016, respectively, when students were in 10th grade. The measurements were part of a larger project to help students make up their professional plans. Only the data relevant for the present study are reported below. At *t*1, we assessed students' math ASC, math competence, subjective task values in math, math expectancy of success, as well as sociodemographic variables. Additionally, schools provided students' math grades on their last report card (the report card at the end of 9th grade, *t*0). *T*2 took place one year later when students were in 11th grade. At *t*2, we again assessed students' subjective task values in math and math expectancy of success. Finally at *t*3, we received the students' math grades from the report cards at the end of 11th grade. Course system changed from *t*1 to *t*2 and *t*3, respectively. In 11th grade students had to choose either basic math classes or advanced math classes. Thus, students were not nested in the same classes in *t*1 as in *t*2 and *t*3.

The sample size was $N = 504$ (264 female, 240 male). The average age of the students at *t*1 was $M = 15.28$ years ($SD = 0.60$). Overall, 484 (96.0%) students were born in Germany, 17 (3.4%) were born in another country, and 3 (0.6%) did not provide information on their country of birth. Additionally, 441 (87.5%) students spoke German as a first language, while 31 (6.2%) did not, and 32 (6.3%) did not provide information on their first languages. All students attended a “Gymnasium” which is the most academically advanced secondary school type in the German school system.

Measures

Math Ability Self-Concept at T1

We assessed math ASC with four items from the absolute school self-concept scale of the SESSKO (German Scales for the Assessment of School-Related Competence Beliefs; Schöne et al., 2002, p. 26) at *t*1. The same items had successfully been used before to assess math ASC (Steinmayr et al., 2018). English versions of the original German items are “I am good at math”, “It is easy to for me to learn in math”, “In math, I know a lot”, and “Most assignments in math are easy for me”. Participants rate them on a five-point Likert scale from “totally disagree” to “totally agree”. The scale’s internal consistency (Cronbach’s α) was high ($\alpha = .95$).

Math Competence at T1

We assessed math competence with the DEMAT 9 (German Mathematics Test for Ninth Grade; Schmidt et al., 2013) at *t*1. The DEMAT 9 is curricular valid for math in ninth grade. Since at *t*1 the students in our sample were at the beginning of 10th grade, this seemed appropriate to ensure that all test contents had already been covered in the students’ courses. The DEMAT 9 consists of 43 individual tasks which are categorized into nine task types: linear equations, number puzzles, percentage and interest calculation, rule of three, theorem of Pythagoras, geometric surfaces, geometric figures, data in diagrams, and data in tables. The DEMAT 9 had high internal consistency ($\alpha = .84$).

Math Expectancy of Success at T1 and T2

We assessed expectancy of success in math with adaptations of three items reported by Eccles & Wigfield (1995). The items were rated on a five-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). An example item is “I will perform very well in math this year”. The internal consistency of the scale was high at $t1$ ($\alpha = .93$) as well as $t2$ ($\alpha = .95$).

Subjective Task Values in Math at T1 and T2

We assessed intrinsic values, utility values, and attainment values in math with the scale for the assessment of subjective school values (SESSW; Steinmayr & Spinath, 2010). Each subjective task value component was assessed with three items that are rated on a five-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). Example items are “I like doing math” (intrinsic value), “Math is useful for my future” (utility value), and “Being good in math is important for me” (attainment value). The internal consistency of all scales was high at $t1$ (intrinsic value: $\alpha = .93$; utility value: $\alpha = .86$; attainment value: $\alpha = .92$) as well as $t2$ (intrinsic value: $\alpha = .92$; utility value: $\alpha = .89$; attainment value: $\alpha = .94$).

Math Academic Achievement at T0 and T3

We chose report card grades as a measure of academic achievement in math because of grades’ high practical relevance for students (Brookhart et al., 2016). Additionally, most studies on SE bias effects on academic achievement have focused on grades as the dependent variable (e.g., Bouffard et al., 2011; Chung et al., 2016; Côte et al., 2014; Dupeyrat et al., 2011; Leduc & Bouffard, 2017; Willard & Gramzow, 2009), increasing the comparability with these studies. Math grades at the investigated schools ranged from 1 (very good) to 6 (insufficient) in Grade 9 ($t0$). In Grade 11 ($t3$), they ranged from 0 to 15 with higher scores indicating better grades. We recoded the grades from Grade 9 so that higher scores indicate better grades at both time points.

Socioeconomic Status at T1

We assessed two indicators of socioeconomic status (SES): parental educational level and number of books at home. Parental educational level was assessed with the question "Which qualification do your parents have? Please select only the highest qualification". Students selected one option per parent from a list of qualifications obtainable in the German school system. We dichotomized the answers into 1 (Abitur or Fachabitur) and 0 (no Abitur or Fachabitur). The "Fachabitur" is a degree obtainable in the German school system, which qualifies students to enter a "Fachhochschule", which is comparable to a university of applied sciences, but not a university. The higher of the scores for the two parents was treated as the parental educational level.

Number of books at home was assessed with the question "How many books are present at your home? *Approximately 40 books fit on one bookshelf. Magazines, newspapers, and your school books are not included"*. Students could choose options ranging from "fewer than 10" to "more than 500". We dichotomized the answers into 0 (up to 100) and 1 (more than 100). This is the common procedure in large scale assessments such as TIMSS and IGLU (e.g., Stubbe et al., 2016; Hußmann et al., 2017). Number of books at home is considered a "classic among the instruments to assess SES" (Stubbe et al., 2016, p. 301) and is still one of the most common and useful indicators of SES today (Schwippert, 2019; Schwippert et al., 2020). It represents a different aspect of SES than parental educational level, namely the cultural and educational resources present at a student's home.

Statistical Analysis

Analyses were computed in R 4.1.3 with the packages lavaan, AICmodavg, and RSA and the R script provided by Humberg et al. (2019). We largely followed the analysis framework by Humberg et al. (2019), using a combination of response surface analysis and information theoretic model comparison. In this approach, no SE bias score (algebraic difference score, residual score, or any other) is computed. Instead, a polynomial regression model of second order (i.e., including quadratic and interactive effects) with the self-evaluation (i.e., *t*1 math ASC) and a related reality criterion (i.e., *t*1 math competence) as predictors is computed. These variables predict the *t*3 math grades directly as well as by the pathway of the mediators (*t*2 expectancy of success, *t*2 intrinsic value, *t*2 utility value, *t*2 attainment value) controlling for prior grades (*t*0). We computed separate models for each mediator because the mediators were highly intercorrelated (see Table 3 in the results section). As pointed out above, we also incorporated a set of control variables (prior scores in the mediators, gender, parental education, and number of books) as covariates. Exemplarily, the model for expectancy of success is depicted in Figure 1. The three models for the subjective task value components are analogous to this model.

To answer our research questions, we computed constrained models that represent the different hypotheses about effects of self-evaluation and SE bias on academic achievement in the literature. For example, the beneficial self-evaluation and competence hypothesis posits that there are only positive linear effects of ASC and competence on grades but no nonlinear effects or SE bias effects. The effect of ASC is thought to be (partially) mediated by expectancy of success and subjective task values according to SEVT. Thus, in the model representing this hypothesis, the *total effects* of *t*1 linear ASC and *t*1 linear competence on *t*3 grades are constrained to be positive. Meanwhile, the *direct* and *indirect effects* are freely estimated so we could analyze whether the effect of ASC (and the effect of competence) on *t*3 grades is mediated by expectancy of success or subjective task values. Additionally, since the beneficial self-evaluation and competence hypothesis does not posit any nonlinear effects or SE bias effects, the total effects of the quadratic and interactive terms on *t*3 grades are constrained to be zero. Likewise, the other hypotheses (i.e., beneficial SE bias hypothesis, optimal margin hypothesis, self-knowledge hypothesis, detrimental SE bias hypothesis) are represented by models with different constraints. The models representing the hypotheses are largely based on those published by Humberg et al. (2019) and have been expanded to include possible mediators of the presumed effects.

We used information theoretic model comparison to compare the aforementioned models against each other. In this approach it is conservative to, when in doubt, include a model in the analysis. More specifically, any model nested within the full polynomial regression model that could plausibly represent the mechanism that produced the data should be included (Burnham & Anderson, 2002; Humberg et al., 2019; Symonds & Moussalli, 2011). Thus, for the sake of completeness, we added 11 additional models, also largely based on the ones published by Humberg et al. (2019), for which we had no theoretical assumptions, to the analysis resulting in 16 models overall. In Table 1, we provide an overview of these 16 models and their parameter constraints. We then compared these 16 models against each other in each of the four analyses sets (one for each mediator). After fitting all models in this initial model set to the data, we excluded models that were virtually redundant as their log-likelihood differed by less than 1 from a more parsimonious model (Humberg et al., 2019; Symonds & Moussalli, 2011). We then computed Akaike weights for the models in this reduced model set. Akaike weights are based on the second-order Akaike information criterion and can be interpreted as a model's likelihood to be the best model given the data and the competing models (Burnham & Anderson, 2002; Humberg et al., 2019). We sorted the models by their Akaike weights and assembled a confidence set which included the models with the highest Akaike weights whose combined Akaike weights exceeded 95%. Thus, analogous to a 95% confidence interval, with an a priori likelihood of over 95% the best model to explain the data given the competing models is one of the models in the confidence set.

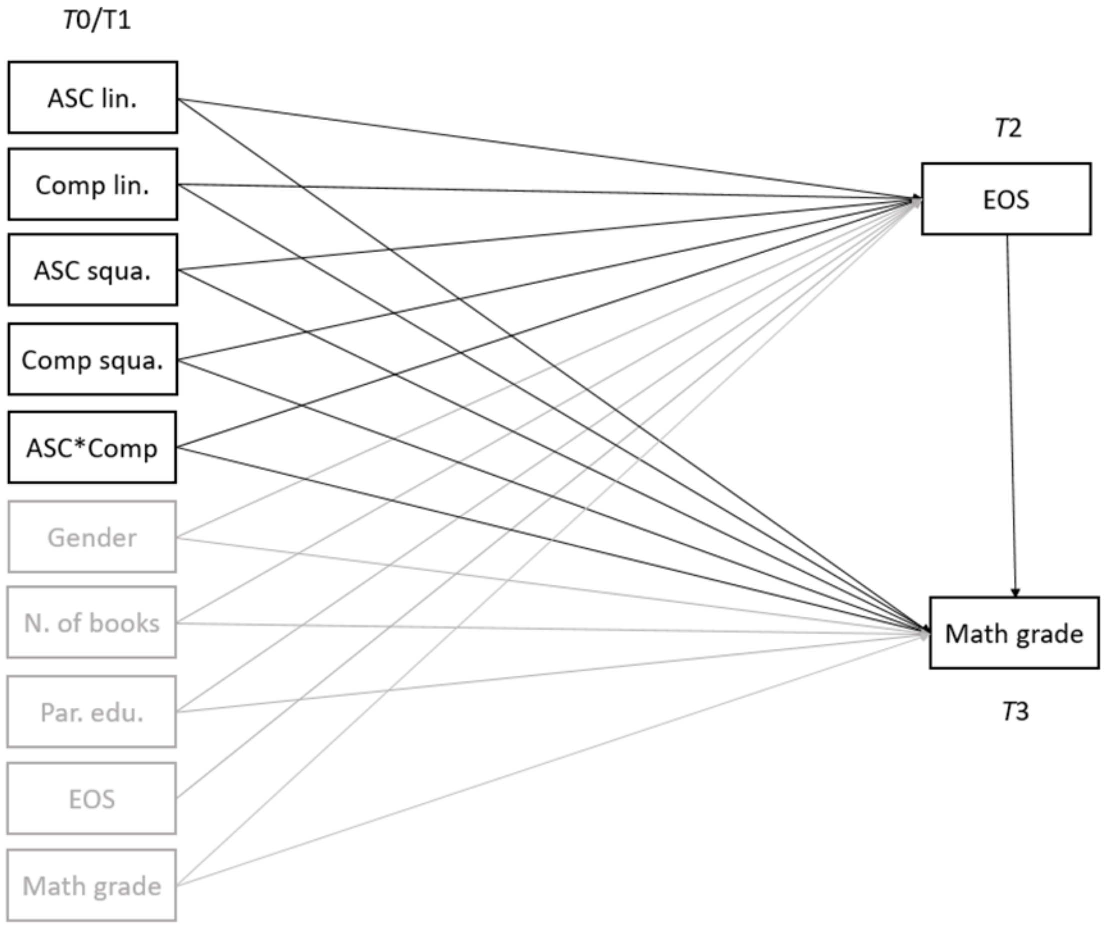


Figure 1. Exemplary full polynomial regression model of *t*3 math grades on *t*1 ASC and competence, *t*0 math grades and selected covariates (printed in grey) mediated by expectancy of success. ASC = ability self-concept; lin. = linear term; squa. = squared term; N. of Books = number of books at the students' home; Par. edu. = Parental education; EOS = expectancy of success.

Analysis Summary

We answer Research Question 1 by empirically comparing models that represent different hypotheses about self-evaluation effects and SE bias effects in the literature. We use Akaike weights as the criterion for model comparison.

We then answer Research Question 2 by inspecting the best model(s). We test the direct, indirect, and total effects in these models for significance. For example, if the beneficial self-evaluation and competence model would be the best model, we test whether the positive linear effects of ASC (and competence) that this model posits are, for example, mediated by expectancy of success.

Supplementary Analyses

Before conducting the analyses described above, we z-standardized the linear ASC and competence scores to reduce nonessential multicollinearity with the nonlinear terms (Salmerón-Gomez et al., 2020). Additionally, we tested to what extent the inclusion of the covariates (gender, number of books, parental education) changed the interpretability of the other predictors (ASC and competence) and their effects by computing double-entry intraclass correlations between the correlations of the predictors and mediators with other constructs (e.g., intelligence) before and after partialling out the covariates (Vize et al., 2018; for details see online supplemental material 1; OSM 1). Finally, since some former studies could not empirically distinguish between ASC and expectancy of success (Eccles et al., 1993; Eccles & Wigfield, 1995) and because ASC, expectancy of success, and intrinsic value were highly correlated in the present study (see Table 3 in the results section), we computed confirmatory factor analyses to test whether these constructs could be empirically distinguished in the present study.

Table 1

Short descriptions and parametric definitions of the analyzed models

Model	Description	Parameter constraints[a]
Full	All effects are freely estimated.	None
Beneficial SE bias	The larger the difference between *t*1 ASC and *t*1 competence, the better *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects.	$b_1 + b_6*c > 0$; $b_2 + b_7*c < 0$; $b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Optimal margin	There is a certain positive value of the difference between *t*1 ASC and *t*1 competence, which is optimal for *t*3 math grades. This effect can be partially mediated by the respective mediator.	$b_1 + b_6*c = -b_2 - b_7*c$; $b_1 + b_6*c - b_2 - b_7*c > 0$; $b_3 + b_8*c < 0$; $b_3 + b_8*c = b_5 + b_{10}*c$; $b_3 + b_8*c + b_4 + b_9*c + b_5 + b_{10}*c = 0$
Self-knowledge	A difference of 0 between *t*1 ASC and *t*1 competence is optimal for *t*3 math grades. This effect can be partially mediated by the respective mediator.	$b_1 + b_6*c = b_2 + b_7*c = 0$; $b_3 + b_8*c < 0$; $b_3 + b_8*c = b_5 + b_{10}*c$; $b_3 + b_8*c + b_4 + b_9*c + b_5 + b_{10}*c = 0$
Detrimental SE bias	The larger the difference between *t*1 ASC and *t*1 competence, the worse *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects.	$b_1 + b_6*c < 0$; $b_2 + b_7*c > 0$; $b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Beneficial self-evaluation and competence	There are positive linear effects of *t*1 ASC and *t*1 competence on *t*3 math grades. These effects can be partially mediated by the respective mediator. There are no nonlinear effects or effects of the discrepancy between *t*1 ASC and *t*2 competence.	$b_1 + b_6*c > 0$; $b_2 + b_7*c > 0$; $b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Beneficial self-evaluation, competence, and self-knowledge	There are positive linear effects of *t*1 ASC and *t*1 competence on *t*3 math grades. Additionally, a smaller difference between *t*1 ASC and *t*1 competence is beneficial for *t*3 math grades. These effects can "counteract" each other because, for example, an increase in *t*1 ASC can also increase the difference between *t*1 ASC and *t*1 competence. These effects can be partially mediated by the respective mediator.	$b_1 + b_6*c > 0$; $b_2 + b_7*c > 0$; $b_3 + b_8*c < 0$; $b_3 + b_8*c = b_5 + b_{10}*c$; $b_3 + b_8*c + b_4 + b_9*c + b_5 + b_{10}*c = 0$
Beneficial self-evaluation	There is a positive linear effect of *t*1 ASC on *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects, effects of *t*1 competence or effects of the difference between *t*1 ASC and *t*1 competence.	$b_1 + b_6*c > 0$; $b_2 + b_7*c = b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Beneficial competence	There is a positive linear effect of *t*1 competence on *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects, effects of *t*1 ASC or effects of the difference between *t*1 ASC and *t*1 competence.	$b_2 + b_7*c > 0$; $b_1 + b_6*c = b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Detrimental self-evaluation	There is a negative linear effect of *t*1 ASC on *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects, effects of *t*1 competence or effects of the difference between *t*1 ASC and *t*1 competence.	$b_1 + b_6*c < 0$; $b_2 + b_7*c = b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Detrimental competence	There is a negative linear effect of *t*1 competence on *t*3 math grades. This effect can be partially mediated by the respective mediator. There are no nonlinear effects, effects of *t*1 ASC or effects of the difference between *t*1 ASC and *t*1 competence.	$b_2 + b_7*c < 0$; $b_1 + b_6*c = b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Detrimental self-evaluation and competence	There are negative linear effects of *t*1 ASC and *t*1 competence on *t*3 math grades. These effects can be partially mediated by the respective mediator. There are no nonlinear effects or effects of the discrepancy between *t*1 ASC and *t*1 competence.	$b_1 + b_6*c < 0$; $b_2 + b_7*c < 0$; $b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Curvilinear ASC	The effect of *t*1 ASC on *t*3 math grades becomes weaker or more negative with increasing values of *t*1 ASC (negative quadratic effect). This effect can be partially mediated by the respective mediator. There are no effects of *t*1 competence or effects of the difference between *t*1 ASC and *t*1 competence.	$b_3 + b_8*c < 0$; $b_2 + b_7*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$
Curvilinear competence	The effect of *t*1 competence on *t*3 math grades becomes weaker or more negative with increasing values of *t*1 competence (negative quadratic effect). This effect can be partially mediated by the respective mediator. There are no effects of *t*1 ASC or effects of the difference between *t*1 ASC and *t*1 competence.	$b_5 + b_{10}*c < 0$; $b_1 + b_6*c = b_3 + b_8*c = b_4 + b_9*c = 0$
Null model	There are no effects of *t*1 ASC and *t*1 competence on *t*3 math grades.	$b_1 + b_6*c = b_2 + b_7*c = b_3 + b_8*c = b_4 + b_9*c = b_5 + b_{10}*c = 0$

Note. [a]Parameter constraints are based on the full model of the form: *t*3 grade = $b_0 + c*$*t*2 mediator + b_1**t*1 ASC + b_2**t*1 competence + b_3**t*1 ASC² + b_4**t*1 ASC*competence + b_5**t*1 competence² + *t*0 grade + gender + books + parental education + ε; *t*2 mediator = $b_{11} + b_6*$*t*1 ASC + b_7**t*1 competence + b_8**t*1 ASC² + b_9**t*1 ASC*competence + $b_{10}*$*t*1 competence² + *t*1 grade + *t*1 mediator + gender + books + parental education + ε.

We compared three different CFA models: 1) a model in which ASC, expectancy of success, and intrinsic motivation are treated as separate but correlated constructs, 2) a model in which ASC and expectancy of success are treated as the same construct, 3) a model in which ASC and intrinsic value are treated as the same construct, and 4) a model in which expectancy of success and intrinsic value are treated as the same construct. We evaluated the fit of these models based on the CFI, RMSEA, and SRMR with cutoff criteria of .95, .08, and .08, respectively (see Hair et al., 2006).

Missing Data

At *t*0 and *t*1, missing were 0.2% of the math competence scores, 0.8% each of the ASC, subjective task value components, and expectancy of success scores, as well as the grades, 1.8% of the scores for number of books at home, and 3.6% of the scores for parental education. There were no missing data for gender. At *t2*, between 17.0% and 17.8% of the data were missing for the subjective task value components and expectancy of success scores, partly because students dropped out of school or changed schools between *t*1 and *t*2. At *t*3, 10.9% of the grade scores were missing. We used full information maximum likelihood (FIML) estimation in all model estimations to account for the missing data.

Results

Preliminary Results

The sample sizes, means, and standard deviations of all non-dichotomous analysis variables are reported in Table 2. With regard to the dichotomic variables, 324 (64.2%) of all participants had at least one parent with a (Fach)Abitur, whereas 163 (32.3%) did not, and 18 (3.6%) did not provide information on this variable. Additionally, 283 (56.0%) had more than 100 books at home, whereas 213 (42.2%) did not, and nine (1.8%) did not provide information on this variable.

There are many large and significant correlations between the variables (see Table 3). The highest correlations were those between linear ASC, expectancy of success and intrinsic value ($.70 \leq r \leq .82$, all $p < .001$). Thus, we used CFAs to test whether these constructs are empirically distinguishable (see Figure 2). The model with separate but correlated constructs had by far the best model fit and was the only one with acceptable fit indices (CFI = .986; RMSEA = .072; SRMR = .037). All other models had substantially worse model fits and did not meet the cutoff criteria for the CFI and the RMSEA (Hair et al., 2016). Thus, ASC, expectancy of success and intrinsic value appeared to be highly correlated but separable constructs in the present study.

Double-entry intraclass correlations were very high for all predictors. That is, the residuals of a certain predictor (e.g., linear ASC) before and after controlling for the covariates had highly similar correlations with other constructs (all $r_{ICC} \geq .97$). Thus, controlling for the covariates had little influence on the interpretability of the other predictors and their effects (for details see OSM 2).

Model comparisons (RQ 1)

In order to answer RQ1, we compared the 16 different constrained models which together make up the initial model set. In Table 4, we report detailed results for the initial model sets in all four analyses (for expectancy of success, intrinsic value, utility value, or attainment value as the mediator). In each of these analyses, several models had to be excluded from the initial model set, as their log-likelihood (*LL*) was virtually identical to a simpler model ($\Delta LL < 1$), resulting in the reduced model set. We report the results for the reduced model sets in Table 5. In each reduced model set we then constructed the 95% confidence set (also reported in Table 5). In the analyses for all four mediators, the beneficial competence model was by far the best model according to Akaike weights (*w*). This model posits that there are only positive total effects of *t*1 linear competence on *t*3 grades but no total effects of *t*1 ASC, nonlinear effects, or SE bias effects on *t*3 grades. This model also was the only model in the confidence set for the analyses of expectancy of success ($w = 96.2\%$), utility value ($w = 97.3\%$), and attainment value ($w = 96.8\%$).

Table 2

Sample sizes (N), means (M), minima, maxima, standard deviations (SD), and Cronbach's α for all non-dichotomous analysis variables

	N	*M*	Minimum	Maximum	*SD*	α
1. *T*1 linear math ASC[a]	501	3.41	1	5	1.13	.95
2. *T*1 linear math competence[a]	504	27.01	0	42	6.88	.84
3. *T*1 squared math ASC	501	1.00	0	4.58	1.13	-
4. *T*1 math ASC and competence interaction	501	0.52	-4.65	5.91	1.04	-
5. *T*1 squared math competence	504	1.00	0	15.40	1.43	-
6. *T*1 math expectancy of success	501	3.21	1	5	1.10	.93
7. *T*1 math intrinsic value	501	3.14	1	5	1.24	.93
8. *T*1 math utility value	501	3.03	1	5	1.02	.86
9. *T*1 math attainment value	501	3.83	1	5	0.96	.92
10. *T*2 math expectancy of success	419	3.19	1	5	1.20	.95
11. *T*2 math intrinsic value	418	3.27	1	5	1.22	.92
12. *T*2 math utility value	415	2.69	1	5	1.10	.89
13. *T*2 math attainment value	419	3.91	1	5	1.06	.94
14. *T*0[b] math grade	500	4.27	2	6	1.01	-
15. *T*3 math grade	450	9.01	2	15	3.04	-

Note. ASC = ability self-concept. [a]In this Table, we report the unstandardized values of all variables, including the linear ASC and competence terms. In all other analyses, z-standardized linear ASC and competence terms are used. [b]Value after recoding.

Table 3

Correlations between the analysis variables

	2.	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*	*16.*	*17.*	*18.*
1. *T*1 linear math ASC	.52***	-.39***	-.12**	-.06	.78***	.82***	.56***	.57***	.49***	.66***	.39***	.40***	.63***	.39***	.26***	.10*	.05
2. *T*1 linear math competence	-	-.12**	-.09*	-.30***	.44***	.44***	.24***	.30***	.38***	.35***	.21***	.22**	.49***	.38***	.31***	.16***	.13**
3. *T*1 squared math ASC		-	.49***	.24***	-.28***	-.30***	-.26***	-.28***	-.16***	-.25***	-.15**	-.19***	-.14**	-.13**	-.12**	.02	.12**
4. *T*1 math ASC and competence interaction			-	.52***	-.06	-.09	-.14**	-.14**	.01	-.07	-.04	-.09	-.05	.02	-.07	.03	.07
5. *T*1 squared math competence				-	-.03	-.05	-.03	-.14**	.04	-.05	.03	-.08	-.01	.03	-.08	.02	.05
6. *T*1 math expectancy of success					-	.70***	.48***	.56***	.54***	.57***	.35***	.43***	.58***	.41***	.16***	.10*	.05
7. *T*1 math intrinsic value						-	.63***	.62***	.40***	.69***	.45***	.43***	.52***	.33***	.21***	.06	.05
8. *T*1 math utility value							-	.56***	.34***	.51***	.56***	.39***	.35***	.21***	.16***	.09	.15***
9. *T*1 math attainment value								-	.33***	.48***	.37***	.57***	.42***	.27***	.03	-.02	.04
10. *T*2 math expectancy of success									-	.62***	.49***	.53***	.47***	.68***	.14**	.13**	.08
11. *T*2 math intrinsic value										-	.59***	.60***	.41***	.47***	.15**	.14**	.03
12. *T*2 math utility value											-	.49***	.24***	.30***	.21***	.05	.10*
13. *T*2 math attainment value												-	.32***	.42***	-.04	.05	.03
14. *T*0 math grade													-	.53***	.07	.14**	.10*
15. *T*3 math grade														-	.01	.15**	.09*
16. *T*1 Gender															-	.07	-.04
17. *T*1 Parental educational level																-	.27***
18. *T*1 Number of books at home																	-

Note. $N = 393 - 504$; ASC = ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = Up to 100; 2 = More than 100.

*$p < .05$, **$p < .01$, ***$p < .001$.

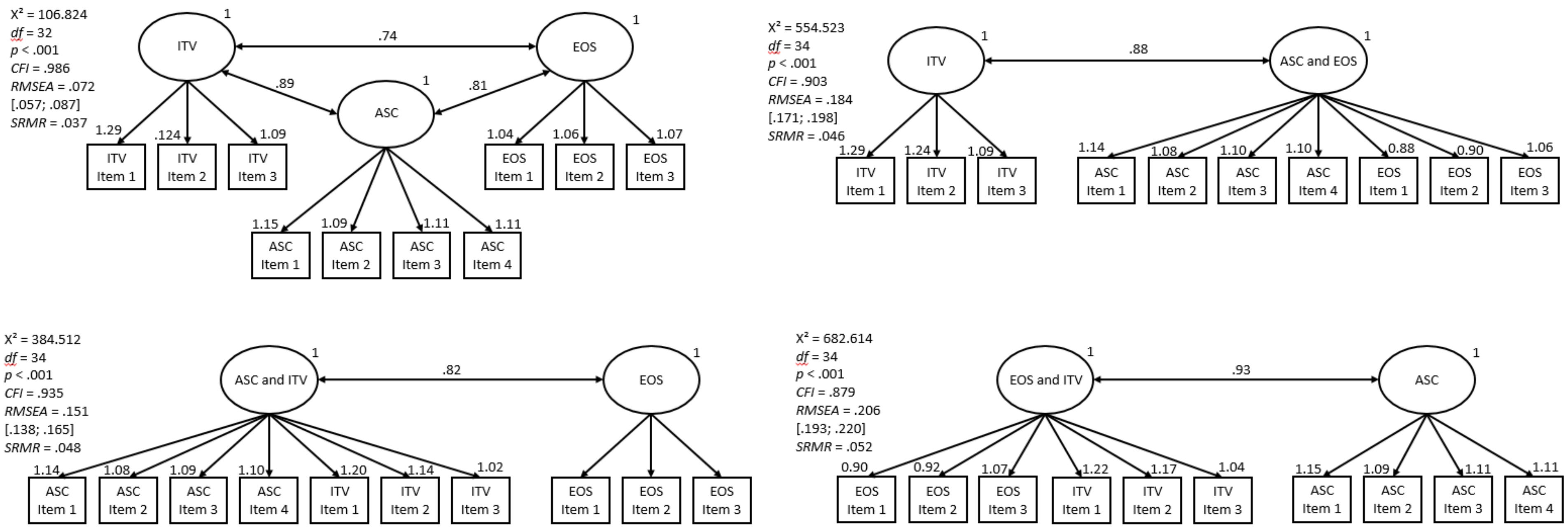


Figure 2. Confirmatory factor analyses of ability self-concept (ASC), expectancy of success (EOS), and intrinsic task value (ITV) in math. For visual clarity, error terms are not presented. All error terms were treated as uncorrelated; $N = 450$.

Thus, with a likelihood of more than 95%, the beneficial competence model was the best model when considering expectancy of success, utility value, or attainment value as mediators. For intrinsic value, the beneficial competence model narrowly missed the 95% threshold and thus, there were two models in the confidence set: The beneficial competence model (w = 93.5%) and the full model (5.9%). Therefore, while the beneficial competence model was the best model for this mediator as well, we cannot entirely reject the hypothesis of more complex effects, such as effects of $t1$ ASC, nonlinear effects or SE bias effects on $t3$ grades.

Effects of *T*1 ASC on *T*3 Grades

Since based on theoretical considerations and former results it was surprising that the best model did not include positive linear effects of $t1$ ASC on $t3$ grades, we inspected the full models to identify possible reasons for this result. Indeed, there was no significant total effect of $t1$ linear ASC on $t3$ grades in any model and the nonsignificant effects were in fact negative (expectancy of success: total effect = -.09, p = .125; intrinsic value: total effect = -.11, p = .083; attainment value: total effect = -.05, p = .442; utility value: total effect = -.03, p = .558). We suspected that the reason for this result might be suppressor effects due to high multicollinearity between predictors. Thus, we computed a regression analysis in which we predicted $t3$ grades by $t1$ linear ASC and $t1$ grades only. In this regression, the effect of ASC became positive but was only marginal significant (β = .10; p = .063). Additionally, we tested whether the regression of $t3$ grades on $t1$ ASC and $t0$ grades would yield a significant effect of ASC if we z-standardized the grade scores within students' classes and courses. This was not the case (β = .01; p = .811). Thus, there were no significant effects of $t1$ ASC on $t3$ grades.

Table 4

Log-likelihoods of the models in the initial model set

	Expectancy of success	*Intrinsic value*	*Utility value*	*Attainment value*
Model	*LL*	*LL*	*LL*	*LL*
1 Full[2,3,4,5,6,7,8,9,10,11,12,13,14,15,16]	**-7107.38**	**-7135.62**	**-7240.06**	**-7183.37**
2 Interaction[3,4,7,11,12,13,14,15,16]	-7109.46	-7138.15	-7242.31	-7185.71
3 Detrimental SE Bias[4,12,16]	-7109.46	-7138.15	-7242.32	-7185.71
4 Beneficial competence[16]	**-7109.91**	**-7138.71**	**-7242.32**	**-7185.72**
5 Curvilinear competence[4,13,16]	-7109.91	-7138.71	-7242.32	-7185.72
6 Self-knowledge plus beneficial effects[4,7,10,15,16]	-7109.91	-7138.71	-7242.28	-7185.72
7 Beneficial self-evaluation and competence[4,15,16]	-7109.91	-7138.71	-7242.31	-7185.72
8 Curvilinear self-evaluation[12,15,16]	-7115.92	**-7144.18**	-7249.90	-7192.19
9 Optimal Margin[10,16]	-7116.17	-7145.14	-7250.78	-7192.90
10 Self-Knowledge[16]	-7116.17	-7145.14	-7250.78	-7192.90
11 Detrimental self-evaluation and comp.[12,13,16]	-7116.86	-7145.29	-7251.44	-7193.07
12 Detrimental self-evaluation[16]	-7116.86	-7145.29	-7251.44	-7193.07
13 Detrimental competence[16]	-7116.87	-7145.33	-7251.44	-7193.07
14 Beneficial SE Bias[13,15,16]	-7116.87	-7145.33	-7250.73	-7192.79
15 Beneficial self-evaluation[16]	-7116.87	-7145.33	-7250.73	-7192.79
16 Null	**-7116.87**	**-7145.33**	**-7251.44**	**-7193.07**

Note. $N = 504$; comp. = competence; Superscript numbers indicate that another model is nested within this model, e.g., "Beneficial self-evaluation and competence[4,5,16]" denotes that models 4, 10 and 16 are nested within the beneficial self-evaluation and competence model. Models in *italics* are based on theoretical considerations in the literature, while models in normal font have been added for methodological reasons; *LL*s in **bold** indicate that a model has been selected for the reduced model set since there is no model nested within that model with a difference in *LL* < 1 (see Table 5 for the reduced model sets).

Table 5

Results for the models in the reduced model set sorted by Akaike weights (w)

Expectancy of success				Intrinsic value			
Model	*AICc*	*w*	*w cumul.*	Model	*AICc*	*w*	*w cumul.*
Beneficial competence	**14424.80**	**96.2%**	**96.2%**	**Beneficial competence**	**14482.40**	**93.5%**	**93.5%**
Full	14431.46	3.4%	99.6%	**Full**	**14487.93**	**5.9%**	**99.4%**
Null	14435.83	0.4%	100%	Null	14492.75	0.5%	> 99.9%
-	-	-	-	Curvilinear self-evaluation	14496.24	< 0.1%	100%
Utility value				**Attainment value**			
Model	*AICc*	*w*	*w cumul.*	Model	*AICc*	*w*	*w cumul.*
Beneficial competence	**14689.61**	**97.3%**	**97.3%**	**Beneficial competence**	**14576.41**	**96.8**	**96.8%**
Full	14696.82	2.6%	>99.9%	Full	14583.43	2.9%	99.7%
Null	14704.96	< 0.1%	100%	Null	14588.23	0.3%	100%

Note. $N = 504$; *AICc* = second order Akaike information criterion; *w* = Akaike weight; *w cumul.* = cumulated Akaike weights. Models in **bold** are part of the 95% confidence set.

Mediator analysis (RQ 2)

Since the beneficial competence model was the best model for all mediators, we inspected this model for mediator effects. For intrinsic value, we also inspected the full model since it was in the confidence set as well. Detailed model results are reported in Appendix A. Because the beneficial competence model does not include effects of t*1* ASC on t*3* grades, we did not test effects of ASC for mediations in this model. Instead, we tested the effect of *t*1 competence on *t*3 grades for mediations by expectancy of success and subjective task values.

Beneficial competence model with expectancy of success as a mediator

In the beneficial competence model with expectancy of success as the mediator (Table A1 in Appendix A), the direct effect of *t*1 linear competence on *t*3 grades was significant ($\beta = .11$, $p = .004$). Additionally, *t*1 linear competence predicted *t*2 expectancy of success ($\beta = .11$, $p = .036$) which in turn predicted *t*3 grades ($\beta = .56$, $p < .001$). This indirect effect of *t*1 linear competence was also significant ($\beta = .06$, $p = .037$), and so was the total effect ($\beta = .18$, $p < .001$). Thus, *t*1 linear competence predicted *t*3 grades and this effect was partially mediated by expectancy of success.

Beneficial competence model with subjective task values as the mediators

The results for the beneficial competence model with intrinsic value (Table A2), attainment value (Table A3), or utility value (Table A4) as the mediator were highly similar. The direct effect of *t*1 linear competence on *t*3 grades was significant in all three models (intrinsic value: $\beta = .17$, $p < .001$; attainment value: $\beta = .17$, $p < .001$; utility value: $\beta = .19$, $p < .001$). Additionally, all *t*2 subjective task value components significantly predicted *t*3 grades (intrinsic value: $\beta = .34$, $p < .001$; attainment value: $\beta = .29$, $p < .001$; utility value: $\beta = .21$, $p < .001$). *T*1 linear competence did not predict any *t*2 subjective task value components (intrinsic value: $\beta = .00$, $p = .930$; attainment value: $\beta = .03$, $p = .522$; utility value: $\beta = .05$, $p = .386$). Thus, the indirect effects of *t*1 linear competence on *t*3 grades through *t*2 subjective task values were nonsignificant as well (intrinsic value: indirect effect = .00, $p = .930$;

attainment value: indirect effect = .01, $p = .520$; utility value: indirect effect = .01, $p = .393$). The total effects of *t*1 linear competence on *t*3 grades however, were significant (intrinsic value: total effect = .17, $p < .001$; attainment value: total effect = .18, $p < .001$; utility value: total effect = .20, $p = .001$).

In summary, all subjective task value components positively predicted subsequent grades. Additionally, competence predicted subsequent grades positively, linearly, and directly, but not indirectly through any subjective task value component.

Full model with intrinsic value as the mediator

In the full model for intrinsic value (Table A5), the direct effect of *t*1 linear ASC on *t*3 grades was negative and significant ($\beta = -.20$, $p = .001$). Additionally, *t*1 linear ASC predicted *t*2 intrinsic value ($\beta = .28$, $p < .001$) which in turn predicted *t*3 grades ($\beta = .35$, $p < .001$). This indirect effect was positive and significant (indirect effect = .10, $p = .001$), resulting in a nonsignificant total effect of *t*1 ASC on *t*3 grades (total effect = -.11, $p = .083$). The direct, indirect, and total effects of *t*1 linear competence on *t*3 grades were highly similar to the same effects in the beneficial competence model (for details see Table A5), meaning that the estimation of the additional parameters in the full model compared to the beneficial competence model has not changed the estimated effects of competence much. All nonlinear effects were nonsignificant.

Summary

Overall, *t*1 competence linearly predicted *t*3 grades positively and this effect was partially mediated by *t*2 expectancy of success but not by *t*2 subjective task values. For *t*2 intrinsic value as a mediator only, there is some evidence for a negative direct and a positive indirect effect of *t*1 ASC on *t*3 grades resulting in a nonsignificant total effect. However, the evidence for this model was still much weaker than the evidence for the beneficial competence model. Thus, there is very little evidence for effects of ASC on grades in the present study.

Discussion

We analyzed how self-evaluated math competence as well as actual math competence and the discrepancy between the two (SE bias) affect math grades directly and via the pathway through expectancy of success and subjective task values in math. Results of model comparisons revealed that the beneficial competence model was the best model in all analyses (answering RQ 1). Additionally, expectancy of success as well as all subjective task value components had significant effects on grades. All these effects held while controlling for prior grades, prior scores of the respective mediator, gender, and two indicators of socioeconomic status. Thus, our results support the position of SEVT that expectancy of success and subjective task values positively affect academic achievement within the same domain, replicating the results of former studies (e.g., Brown & Putwain, 2022; Froiland & Davison, 2016; Geng et al., 2022; Steinmayr et al., 2019; Steinmayr & Spinath, 2009; Wang, 2012; Weidinger et al., 2020). However, because we did not find evidence for effects of ASC on academic achievement and therefore also no mediation of this effect by expectancy of success and subjective task values, our results contradict SEVT in that regard (answering RQ 2). In the following, we first discuss the surprising absence of ASC effects on grades in the present study before turning to implications for SE bias research.

Self-Evaluation Effects on Academic Achievement and Their Mediation by Expectancy of Success and Subjective Task Values

The result that ASC as our measure of self-evaluated math competence did not predict grades contradicts an extensive literature reporting positive effects of ASC on grades in the same domain (see Valentine et al. 2004; Wu et al., 2021 for meta-analyses). A possible explanation for the missing effect of ASC on subsequent grades in the present study is the relatively long interval between the measurement of ASC (beginning of Grade 10) and students receiving the *t*3 grades (end of Grade 11) in which the course and class system changed. However, in a meta-analysis, Wu et al. (2021) found that the time interval between the measurement of self-evaluation and the measurement of academic achievement did not moderate the size of the effect of self-evaluation on academic achievement. However, he did not consider changes in class and course systems as a potential moderator. Thus, another possible explanation is the fact that the course system of the schools changed between Grade 10 and Grade 11. Up until and including in Grade 10, students were taught in math classes that all followed the same basic curricula. From Grade 11 onwards, students could choose between basic courses and advanced courses which differed in their demands. These different demands might have obscured the effects of ASC on grades (as well as on expectancy of success and subjective task values). Particularly, as students with a high math ASC were more likely to choose an advanced math course (see Steinmayr & Spinath, 2010) and the increased difficulty of these courses could have negatively effected the students grades, expectancy of success, and subjective task values, counteracting hypothetical beneficial effects of ASC. However, after z-standardizing all grade scores within students' classes and courses, ASC still did not predict subsequent grades and the size of the effect in fact became smaller. Still, it is possible that the change of the course system made previous estimations of own competences less relevant for current motivation and grades. It is also possible that the effect of ASC was missing due to suppressor effects from high multicollinearity between predictors (Marsh et al., 2004). Expectancy of success and ASC are conceptually similar and have been found to be empiricially inseparable in some studies (Eccles et al., 1993; Eccles & Wigfield, 2002; Eccles & Wigfield, 1995; Marsh et al., 2019), to the point where most studies did not differentiate between them (e.g., Arens et al., 2019; Guo et al., 2017; Guo et al., 2015a; 2015b; 2015c; 2015d; Nagengast et al., 2011; Trautwein et al., 2012). Indeed, the correlations between ASC and expectancy of success in the present study were high for both, concurrent values and prior ASC with subsequent expectancy of success. Thus, it is possible that due to the high correlation between ASC and expectancy of success, there was not enough unique variance left in the ASC scores to significantly predict subsequent expectancy of success.

However, CFAs indicated that while highly correlated, ASC, expectancy of success, (and intrinsic value) were highly correlated but separable. Additionally, we did not observe significant effects of ASC on grades in the models with attainment value and utility despite the fact that these variables exhibited much less multicollinearity with ASC. Even when predicting grades by ASC and prior grades only, there was still no significant effect of ASC, although the effect did become larger, descriptively. Given that effects of ASC on academic achievement are typically small (Valentine et al., 2004; Wu et al., 2021), this might have been a problem of a too small sample size.

The only significant effects of ASC on grades were found in the full model for intrinsic value as a mediator. Here, ASC had a negative direct, a positive indirect and a nonsignificant total effect. Since the total effect is the sum of the direct and indirect effect (Hayes, 2013), this means that after accounting for the proportion of variance in ASC that predicts intrinsic value, the remaining variance negatively predicted grades. The positive effect of ASC on intrinsic values might be explained by the self-determination theory, according to which intrinsic motivation is facilitated when the basic needs for feelings of competence, autonomy, and social relatedness are fulfilled (Ryan & Deci, 2000; 2019). Thus, the feeling of competence associated with a high ASC might be beneficial for intrinsic motivation and thus grades. However, other studies with fewer control variables did not find ASC to predict subsequent intrinsic motivation (Spinath & Steinmayr, 2012). The negative effect of ASC on grades is probably due to suppression effects. Furthermore, the full model, even though not rejected entirely, still had much weaker evidence supporting it than the beneficial competence model, which does not include effects of ASC. Therefore, these results should be interpreted with caution.

In the present study, ASC affected intrinsic value but not attainment value, utility value or expectancy of success. Of the three subjective task value components, intrinsic value is conceptually most similar to intrinsic motivation, while utility value is most similar to extrinsic motivation (Wigfield & Eccles, 2020). As has been pointed out above, self-determination theory posits that feelings of competence should increase intrinsic motivation, which might be the reason why only intrinsic value was predicted by ASC. Additionally, empirical studies indicate that competence beliefs are most strongly related to intrinsic value (Conley, 2012; Guo et al., 2015b; Wigfield et al., 1997). In a longitudinal study, Arens et al. (2019) found that ASC predicts both intrinsic and attainment values but did not test for the effect on utility value. The effects on intrinsic value were slightly larger than the effects on attainment value, but this difference was not tested for significance. Overall, theoretical considerations and empirical results suggest that ASC might be most important for intrinsic value and least important for utility value.

SE Bias Effects

In the present study, there are no indications of SE bias effects since none of the models positing SE bias effects (beneficial SE bias model, detrimental SE bias model, self-knowledge model, optimal margin model) remained in the confidence set or even the reduced model set for any of the mediators. This means that assuming the existence of SE bias effects offers almost no improvement of the models fit to the data over simple linear main effects models. This result is in contrast to former studies, which found mostly positive SE bias effects on academic achievement (e.g., Bonneville-Roussy et al., 2017; Côté et al., 2014; Dupeyrat et al., 2011; Leduc & Bouffard, 2017; Lee, 2021; Wright, 2000). However, virtually all of these studies operationalized SE bias as an algebraic difference score or a residual score and then, in a second step, used this score to predict academic achievement. This “two-step approach” has the problem of confounding SE bias effects with simple main effects of the self-evaluation (Humberg et al., 2018; 2019). First studies using methods that allow disentangling SE bias effects from self-evaluation effects (condition-based regression analysis; RSA) did not find any SE bias effects on academic achievement (authors, 2020; 2023). Since self-evaluation effects are well-documented (e.g., Talsma et al., 2018; Valentine et al., 2004; Wu et al., 2021), the formerly found SE bias effects were likely, at least to some degree, artifacts of self-evaluation effects. Since there were no SE bias effects in the present study, SE bias effects were also not mediated by expectancy of success or subjective task values. Thus, the present study extents the findings by authors (2020; 2023) in two important ways. First, we were able to show that not only academic achievement, but also subjective task values and expectancy of success are unaffected by SE bias. Second, we were able to show that there is still no SE bias effect on academic achievement even when motivational variables are included in the model. Former results of SE bias effects on motivational variables were heterogeneous with some revealing positive effects (e.g., Bonnevile-Roussy et al., 2017; Dupeyrat et al., 2011; Gonida & Leondari, 2011; Lee, 2021; Murphy et al., 2018). However, these studies have the same limitation discussed above: Since virtually all of them used a two-step approach to analyze SE bias effects, SE bias effects on motivation cannot be disentangled from main effects of the self-evaluation. These discrepant results also reflect theoretical limitations of former studies. It has often been argued that a high SE bias should be beneficial for academic achievement because it improves motivation (e.g., Bonneville-Roussy et al., 2017; Helmke, 1998; Lee, 2021; Martin & Debus, 1998; Taylor & Brown, 1988; Wright, 2000). However, virtually none of these studies put forward arguments why the bias in the self-evaluation rather than the self-evaluation per se should affect motivation. As Griffin et al. (1999) put it: “too often, stories are told about discrepancy […] models when only simple main effect stories are there to be told” (p. 510). The strength of the present study is that by the application of RSA we were able to statistically distinguish between these discrepancy models and “simple main effect” models. However, it should be kept in mind that the results of the present study were unusual in that ASC did not predict subsequent grades. Since SE bias effects are effects of the discrepancy between a self-evaluation (e.g., ASC) and competence, it is possible that the present results for SE bias effects are also unusual and might not easily be generalizable to other populations.

Limitations

Despite the aforementioned strength of the RSA, the present study also has some noteworthy limitations. First, SE bias as a discrepancy between a self-evaluation and a related criterion can be operationalized in many different ways. We chose to compare math ASC as a measure of self-evaluation to math competence as a measure of the reality criterion. However, other measures such as math self-efficacy instead of math ASC could have been used as well. We chose ASC because according to SEVT, ASC should affect academic achievement via a pathway of expectancy of success and subjective task values. We wanted to test this assumption and further analyze whether SE bias effects (either directly or mediated by expectancy of success and subjective task values as well) affect academic achievement beyond the main effect of ASC. Thus, the measures used in the present study are suited to analyze how a general self-evaluation of own math competence and the bias in that general self-evaluation affect academic achievement. However, this also means that we cannot draw conclusions about the more fine-grained perspective of the learning process itself. Authors from the field of metacognitive self-regulated learning stress that students monitor what contents they have learned how well and use that information to motivate and guide their learning efforts (e.g., Buckelew et al., 2013; Dunlosky et al., 2005; Dunlosky & Rawson, 2012; Dunning et al., 2004; Hacker & Bol, 2019; Händel & Fritzsche, 2016; Kim et al., 2010; Thiede et al., 2003). Since we have not observed the learning process itself, we cannot and do not aim to test this assumption. Thus, it is possible that an accurate assessment of learning success is indeed beneficial for academic achievement, while the accuracy of general self-estimated competence (as measured in the present study) does not have an influence.

Second, the sample of the present study is highly homogenous and thus, it is unclear whether the results can be generalized to different grades and school types. However, it has been demonstrated that the effect of ASC on academic achievement holds equally for students in mixed or academically selective groups (Preckel et al., 2017; Seaton et al., 2015).

Third, the present study follows a predictive design instead of an experimental one. Thus, we cannot unequivocally draw conclusions about causal relations. This problem, which is shared by virtually all studies on self-evaluation effects and SE bias effects, is mitigated to some extent by the inclusion of covariates, which, as confounding variables, could have led to spurious effects of ASC, competence or the SE bias on academic achievement or the motivational variables.

Lastly, there was high multicollinearity between some predictors in our study, most notably between ASC, expectancy of success, and intrinsic value. While the CFAs indicated that these traits were indeed separable in our study, the high multicollinearity between predictors in the same model could still have affected the results.

Conclusion

Overall, the present study shows that the discrepancy between self-evaluated math competence and actual math competence (SE bias) does not appear to have an influence on either academic achievement, expectancy of success, or subjective task values. Instead, the only robust effects on academic achievement which we observed were positive effects of expectancy of success and all three subjective task value components (intrinsic, attainment, and utility) as well as a positive effect of competence which was partially mediated by expectancy of success but not subjective task values. These results highlight the importance of using methods that disentangle SE Bias effects from self-estimation effects. However, the fact that ASC did not affect academic achievement at all (neither directly nor mediated by expectancy of success or subjective task values) was surprising given the extent of former evidence for this effect. Only future research with similar samples and designs can answer whether this result was an outlier or serves as an example that self-estimation effects are not robust for certain student populations.

References

Arens, A. K., Schmidt, I., & Preckel, F. (2019). Longitudinal relations among self-concept, intrinsic value, and attainment value across secondary school years in three academic domains. *Journal of Educational Psychology, 111*(4), 663–684. https://doi.org/10.1037/edu0000313.supp (Supplemental)

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955-966.https:// doi.org/10.1037//0022-3514.74.4.955

Bakadorova, O., & Raufelder, D. (2020). The relationship of school self-concept, goal orientations and achievement during adolescence. *Self and Identity, 19*(2), 235–249. https://doi.org/10.1080/15298868.2019.1581082

Baumeister, R. F. (1989). The optimal margin of illusion. *Journal of Social and Clinical Psychology, 8*(2), 176-189. https://doi.org/10.1521/jscp.1989.8.2.176

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. *Sex Roles, 72*(11-12), 536-546. https://doi.org/10.1007/s11199-015-0486-9

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. https://doi.org/10.1016/j.jsp.2017.02.002

Bouffard, T., Vezeau, C., Roy, M., & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research, 50*, 221-229. https://doi.org/10.1016/j.ijer.2011.08.003

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H, Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://doi.org/10.3102/0034654316672069

Brown, C. & Putwain, D. W. (2022). Socio-economic status, gender and achievement: the mediating role of expectancy and subjective task value. *Educational Psychology, 42*(6), 730.748. https://doi.org/10.1080/01443410.2021.1985083

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*(2), 134-138. https://doi.org/10.1177/0098628312475034

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.)*. Springer.

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: A longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*(10), 1384-1401. https://doi.org/10.1177/0146167216662866

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin, 116*, 3–20. https://doi.org/10.1037/0033-2909.116.1.3

Conley, A. M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology, 104*, 32–47. http://dx.doi.org/10.1037/a0026042

Côté, S., Bouffard, T., & Vezeau, C. (2014). The mediating effect of self-evaluation bias of competence on the relationship between parental emotional support and children's academic functioning. *British Journal of Educational Psychology, 84*(3), 415–434. https://doi.org10.1111/bjep.12045

Dunlosky, J., Hertzog, C., Kennedy, M. R. F., & Thiede, K. W. (2005). The self-monitoring approach for effective learning. *Cognitive Technology, 10*(1), 4–11.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https//doi.org10.1016/j.learninstruc.2011.08.003

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest, 5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Dupeyrat, C., Escribe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250. https://doi.org/10.1016/j.ijer.2011.08.005

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215–225. https://doi.org/10.1177/0146167295213003

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*. https://doi.org/10.1016/j.cedpsych.2020.101859

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*(3), 830–847. https://doi.org/10.2307/1131221

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577-1613. https://doi.org/10.2307/256822

Edwards, J. R., & Parry, M. E. (2018). On the use of spline regression in the study of congruence in organizational research. *Organizational Research Methods, 21*(1), 68–110. https://doi.org/10.1177/1094428117715067

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232-244. doi: 10.1016/j.cedpsych.2015.03.002

Froiland, J. M., & Davison, M. L. (2016). The longitudinal influences of peers, parents, motivation, and mathematics course-taking on high school math achievement. *Learning and Individual Differences, 50*, 252–259. http://dx.doi.org/10.1016/j.lindif.2016.07.012

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology, 52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.00

Geng, S., Lu, Y., & Shu, H. (2022). Cross-cultural generalizability of expectancy-value theory in reading: A multilevel analysis across 80 societies. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*. https://doi.org/10.1007/s12144-022-03014-0

Gonida, E., N., & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209-220. https://doi.org/10.1016/j.ijer.2011.08.002

Griffin, D., Murray, S., & Gonzalez, R. (1999). Difference score correlations in relationship research: A conceptual primer. *Personal Relationships, 6*, 505-518. https://doi.org/10.1111/j.1475-6811.1999.tb00206.x

Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2015a). Directionality of the associations of high school expectancy value, aspirations, and attainment: A longitudinal study. *American Educational Research Journal, 52*, 371– 402. http://dx.doi.org/10.3102/0002831214565786

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction, 49*, 81–91. http://dx.doi.org/10.1016/j.learninstruc.2016.12.007

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015b). Expectancy-value in mathematics, gender and socioeconomic background as predictors of achievement and aspirations: A multi-cohort study. *Learning and Individual Differences, 37*, 161–168. http://dx.doi.org/10.1016/j.lindif.2015.01.008

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., & Trautwein, U. (2015c). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open, 2*, 1–20. http://dx.doi.org/10.1177/2332858415626884

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015d). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology, 51*, 1163–1176. http://dx.doi.org/10.1037/a0039440

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of* *cognition and education*. (pp. 647–677). Cambridge University Press. https://doi.org/10.1017/9781108235631.026

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis.* Pearson University Press.

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition, 44*(2), 229–241. https://doi.org/10.3758/s13421-015-0552-0

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. The Guilford Press. https://doi.org/10.1111/jedm.12050

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe [From an optimist to a realist? On the development of the ability self-concept from kindergarten to 6th grade]. In F. E. Weinert (Ed.), *Entwicklung im Kindesalter* (pp. 115-132). Beltz.

Heyder, A., Steinmayr, R., & Kessels, U. (2019). Do Teachers' Beliefs About Math Aptitude and Brilliance Explain Gender Differences in Children's Math Ability Self-Concept? *Frontiers in Education, 4*. https://doi.org/10.3389/feduc.2019.00034

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859. https://doi.org/10.1037/pspp0000204

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology, 114*(2), 303–322. https://doi.org/10.1037/pspp0000134

Hußmann, A., Stubbe, T. C., & Kasper, D. (2017). Kapitel VI. Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T.C. Stubbe, & R. Valtin (Eds.), *IGLU 2016. Lesekompetenz von Grundschulkindern in Deutschland im internationalen Vergleich [IGLU 2016. Reading competence of grade schoolers in Germany by international comparison]*. (S. 195-218). Münster, DE: Waxmann.

Jansen, M., Becker, M., & Neumann, M. (2021). Dimensional comparison effects on (gendered) educational choices. *Journal of Educational Psychology, 113*(2), 330–350. https://doi.org/10.1037/edu0000524.supp (Supplemental)

Kim, Y.-H., Chiu, C., & Zou, Z. (2010). Know thyself: Misperceptions of actual performance undermine achievement motivation, future performance, and subjective well-being. *Journal of Personality and Social Psychology, 99*(3), 395–409. https://doi.org/10.1037/a0020555

Leduc, C., & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences, 55*, 193–201. https://doi.org/10.1016/j.lindif.2017.04.006

Lee, E. J. (2021). Biased self-estimation of maths competence and subsequent motivation and achievement: Differential effects for high- and low-achieving students. *Educational Psychology, 41*(4), 446–466. https://doi.org/10.1080/01443410.2020.1821869

Lopez, D. F., Little, T. D., Oettingen, G., & Baltes, P. B. (1998). Self-regulation and school performance: Is there optimal level of action-control? *Journal of Experimental Child Psychology, 70*, 54-74. https://doi.org/10.1006/jecp.1998.2446

Mac Iver, D. J., Stipek, D. J., & Daniels, D. H. (1991). Explaining within-semester changes in student effort in junior high school and senior high school courses. *Journal of Educational Psychology, 83*(2), 201–211. https://doi.org/10.1037/0022-0663.83.2.201

Marsh, H. W. (2022). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000750.supp (Supplemental)

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology, 96*(3), 518–522. https://doi.org/10.1037/0022-0663.96.3.518

Marsh, H. W., Pekrun, R., & Lüdtke, O. (2022). Directional ordering of self-concept, school grades, and standardized tests over five years: New tripartite models juxtaposing within- and between-person perspectives. *Educational Psychology Review.* https://doi.org/10.1007/s10648-022-09662-9

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331-353. https://do.org/10.1037/edu0000281

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*(1), 41–54. https://doi.org/10.1037/0022-0663.89.1.41

Martin, A. J., & Debus, R. L. (1998). Self-reports of mathematics self-concept and educational outcomes: The roles of ego-dimensions and self-consciousness. *British Journal of Educational Psychology, 68*(4), 517–535. https://doi.org/10.1111/j.2044-8279.1998.tb01309.x

Murphy, S. C., Barlow, F. K., & von Hippel, W. (2018). A longitudinal test of three theories of overconfidence. *Social Psychological and Personality Science, 9*(3), 353–363. https://doi.org/10.1177/1948550617699252

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science, 22*, 1058–1066. http://dx.doi.org/10.1177/0956797611415540

Nietfeld, J. L., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research, 95*(3), 131–142. https://doi.org/10.1080/00220670209596583

OECD (2016). *PISA 2015 Ergebnisse (Band 1): Exzellenz und Chancengerechtigkeit in der Bildung [PISA 2015 Results (Volume 1): Excellence and Equity in Education]*. Bertelsmann.

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern [Are students with optimistic self-concepts one step ahead? Relations between optimistic, realistic, and pessimistic self-concepts and the achievement development of primary school children]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. https://doi.org/10.1177/0016986216687824

Priess-Groben, H. A., & Hyde, J. S. (2017). Implicit theories, expectancies, and values predict mathematics motivation and behavior across high school and college. *Journal of Youth and Adolescence, 46*(6), 1318–1332. https://doi.org/10.1007/s10964-016-0579-y

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*(2), 340-352. https://doi.org/10.1037/0022-3514.80.2.340

Rohr, M. E., & Ayers, J. B. (1973). Relationship of student grade expectations, selected characteristics, and academic performance. *Journal of Experimental Education, 41*(3), 58-62. https://doi.org/10.1080/00220973.1973.11011410

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2019). Brick by brick: The origins, development, and future of self-determination theory. In A. J. Elliot (Ed.), *Advances in motivation science.* (Vol. 6, pp. 111–156). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.01.001

Salmerón-Gómez, R., Rodríguez-Sánchez, A. & García-García, C. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational statistics, 35*, 647-666. https://doi.org/10.1007/s00180-019-00922-x

Schmidt, S., Ennemoser, M., & Krajewski, K. (2013). *Deutscher Mathematiktest für neunte Klassen (DEMAT 9) [German Mathematics Test for Ninth Grade (DEMAT 9)].* Hogrefe.

Schöne, C., Dickhäuser, O., Spinath, B., & Stiensmeier-Pelster, J. (2002). *Skalen zur Erfassung des schulischen Selbstkonzeptes SESSKO [Scales for the assessment of the school self-concept SESSKO]*. Hogrefe.

Schwippert, K. (2019). Was wird aus den Büchern? Sozialer Hintergrund von Lernenden und Bildungsungleichheit aus Sicht der international vergleichenden Erziehungswissenschaft [What happens about the books? Social background of learners and educational equity from the perspective of international comparative educational science]. *Journal for educational research online, 11*, 92-117. https://doi.org/10.25656/01:16789

Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M., & Wendt, H. (2020). *TIMSS 2019 Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland und im internationalen Vergleich [TIMSS 2019 grade school students' mathematical and science competences in Germany and by international comparison].* Waxmann.

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly, 59*(3), 143–156.

Sewasew, D., & Koester, L. S. (2019). The developmental dynamics of students' reading self-concept and reading competence: Examining reciprocal relations and ethnic-background patterns. *Learning and Individual Differences, 73*, 102–111. https://doi.org/10.1016/j.lindif.2019.05.010

Sewasew, D., & Schroeders, U. (2019). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology, 58*, 204–212. https://doi.org/10.1016/j.cedpsych.2019.03.009

Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles: A Journal of Research, 50*(3–4), 241–252. https://doi.org/10.1023/B:SERS.0000015555.40976.e6

Steinmayr, R., Dinger, F. C., & Spinath, B. (2012). Motivation as a mediator of social disparities in academic achievement. *European Journal of Personality, 26*(3), 335–349. https://doi.org/10.1002/per.842

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation? *European Journal of Personality, 22*(3), 185–209. https://doi.org/10.1002/per.676.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*(1), 80–90. https://doi.org/10.1016/j.lindif.2008.05.004

Steinmayr, R., & Spinath, B. (2010). Konstruktion und erste Validierung einer Skala zur Erfassung subjektiver schulischer Werte (SESSW) [Construction and first validation of a scale fort he assessment of subjective school-related values (SESSW)]. *Diagnostica, 56*(4), 195-211. https://doi.org/10.1026/0012-1924/a000023

Steinmayr, R., Weidinger, A. F., Schwinger, M., & Spinath, B. (2019). The importance of students' motivation for their academic achievement—Replicating and extending previous findings. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.01730

Steinmayr, R., Weidinger, A. F., & Wigfield, A. (2018). Does students' grit predict their school achievement above and beyond their personality, motivation, and engagement? *Contemporary Educational Psychology, 53*, 106-122. https://doi.org/10.1016/j.cedpsych.2018.02.004

Stubbe, T. C., Schwippert, K., & Wendt, H. (2016). Kapitel X. Soziale Disparitäten der Schülerleistungen in Mathematik und Naturwissenschaften. In H. Wendt, W. Bos, C. Selter, O. Köller, K. Schwippert, & D. Kasper (Eds.), *TIMSS 2015. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im* *internationalen Vergleich [TIMSS 2015. Mathematical and science competences of grade schoolers in Germany by international comparison].* (S. 299-316). Waxmann.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology, 65*, 13–21. https://doi.org/10.1007/s00265-010-1037-6

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences, 61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210. https://doi.org/10.1037/0033-2909.103.2.193

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21–27. https://doi.org/10.1037/0033-2909.116.1.21

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

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. http://dx.doi.org/10.1037/a0027470

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D.

Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 187-214). Springer International Publishing.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. do: 10.1002/per.2137

Wang, M.-T. (2012). Educational and career interests in math: A longitudinal examination of the links between classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*(6), 1643–1657. https://doi.org/10.1037/a0027247

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2020). The value of valuing math: Longitudinal links between students' intrinsic, attainment, and utility values and grades in math. *Motivation Science, 6*(4), 413–422. https://doi.org/10.1037/mot0000179.supp (Supplemental)

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development, 89*(2), e138–e156. https://doi.org/10.1111/cdev.12806

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Ed.), *Advances in motivation science Vol. 7* (pp. 161–198). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.05.002

Wigfield, A., Eccles, J. S., & Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance, and choice. *Educational Psychology Review, 32*(3), 657–680. https://doi.org/10.1007/s10648-020-09524-2

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across elementary school years: A 3-year study. Journal of *Educational Psychology, 89*, 451–469. http://dx.doi.org/10.1037/0022-0663.89.3.451

Wilgenbusch, T., & Merrell, K. W. (1999). Gender differences in self-concept among children and adolescents: A meta-analysis of multidimensional studies. *School Psychology Quarterly, 14*(2), 101–120. https://doi.org/10.1037/h0089000

Willard, G., & Gramzow, R. H. (2009). Beyond oversights, lies, and pies in the sky: Exaggeration as goal projection. *Personality and Social Psychology Bulletin, 35*(4), 477–492. https://doi.org/10.1177/0146167208329631

Wright, S. (2000). Looking at the self in a rose-colored mirror: Unrealistically positive self-views and academic performance. *Journal of Social and Clinical Psychology, 19*(4), 451-462. https://doi.org/10.1521/jscp.2000.19.4.451

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09600-1

Appendix A. Detailed Results of the Models in the Confidence Sets

Table A1

Detailed results for the beneficial competence model (w = 96.2%) with expectancy of success as a mediator

Parameter	β	*SE*	95% *CI*	*p*
Intercept of *t*3 grade	0.43	0.261	[-0.08, 0.94]	.098
*T*3 grade on *t*2 expectancy of success	.56	.035	[0.49, 0.63]	< .001
*T*3 grade on *t*1 linear ASC[a]	-.06	.032	[-0.13, 0.00]	.054
*T*3 grade on *t*1 linear competence	.11	.040	[0.04, 0.19]	.004
*T*3 grade on *t*1 quadratic ASC[a]	-.01	.024	[-0.06, 0.04]	.701
*T*3 grade on *t*1 ASC/competence interaction[a]	-.02	.022	[-0.06, 0.03]	.432
*T*3 grade on *t*1 quadratic competence[a]	-.02	.019	[-0.06, 0.02]	.243
*T*3 grade on Gender	-.12	.033	[-0.18, -0.05]	< .001
*T*3 grade on number of books	-.01	.035	[-0.08, 0.06]	.729
*T*3 grade on parental education	.08	0.38	[0.00, 0.15]	.044
*T*3 grade on *t*0 grade	.27	0.43	[0.18, 0.35]	< .001
Intercept of *t*2 expectancy of success	0.74	0.314	[0.13, 1.36]	.018
*T*2 expectancy of success on *t*1 linear ASC[a]	.11	.057	[-0.00, 0.22]	.051
*T*2 expectancy of success on *t*1 linear competence[a]	.11	.052	[0.01, 0.21]	.036
*T*2 expectancy of success on *t*1 quadratic ASC[a]	.02	.042	[-0.07, 0.10]	.700
*T*2 expectancy of success on *t*1 ASC/comp. interaction[a]	.03	.040	[-0.05, 0.11]	.433
*T*2 expectancy of success on *t*1 quadratic competence[a]	.04	.034	[-0.03, 0.11]	.244
*T*2 expectancy of success on *t*0 grade	.18	.051	[0.08, 0.29]	< .001
*T*2 expectancy of success on gender	.01	.043	[-0.08, 0.09]	.859
*T*2 expectancy of success on number of books	.02	.042	[-0.06, 0.10]	.652
*T*2 expectancy of success on parental education	.01	.043	[-0.07, 0.10]	.746
*T*2 expectancy of success on *t*1 expectancy of success	.30	.062	[0.18, 0.42]	< .001

Note. $N = 504$. 95% *CI* = 95% confidence interval; ASC = math ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 of *t*3 math grades = .55; R^2 of *t*2 expectancy of sucess = .35; [a]Because the total effect of this predictor was constrained to zero in this model, the regression weights of this predictor should not be interpreted.

Table A2

Detailed results for the beneficial competence model (w = 93.5%) with intrinsic value as a mediator

Parameter	β	*SE*	95% *CI*	*p*
Intercept of *t*3 grade	0.43	0.307	[-0.18, 1.03]	.166
*T*3 grade on *t*2 intrinsic value	.34	.046	[0.25, 0.43]	< .001
*T*3 grade on *t*1 linear ASC[a]	-.11	.028	[-0.16, -0.05]	< .001
*T*3 grade on *t*1 linear competence	.17	.046	[0.08, 0.26]	< .001
*T*3 grade on *t*1 quadratic ASC[a]	-.01	.012	[-0.03, 0.02]	.458
*T*3 grade on *t*1 ASC/competence interaction[a]	.00	.012	[-0.02, 0.03]	.851
*T*3 grade on *t*1 quadratic competence[a]	.01	.012	[-0.01, 0.04]	.316
*T*3 grade on Gender	-.11	.039	[-0.19, -0.03]	.005
*T*3 grade on number of books	.00	.039	[-0.07, 0.08]	.916
*T*3 grade on parental education	.06	0.41	[-0.02, 0.14]	.136
*T*3 grade on *t*0 grade	.39	0.46	[0.30, 0.48]	< .001
Intercept of *t*2 intrinsic value	1.44	0.313	[0.83, 2.06]	< .001
*T*2 intrinsic value on *t*1 linear ASC[a]	.31	.066	[0.18, 0.44]	< .001
*T*2 intrinsic value on *t*1 linear competence[a]	.00	.046	[-0.09, 0.09]	.930
*T*2 intrinsic value on *t*1 quadratic ASC[a]	.03	.036	[-0.04, 0.10]	.459
*T*2 intrinsic value on *t*1 ASC/comp. interaction[a]	-.01	.035	[-0.08, 0.06]	.850
*T*2 intrinsic value on *t*1 quadratic competence[a]	-.04	.035	[-0.11, 0.03]	.304
*T*2 intrinsic value on *t*0 grade	-.01	.044	[-0.10, 0.08]	.817
*T*2 intrinsic value on gender	-.02	.036	[-0.09, 0.05]	.529
*T*2 intrinsic value on number of books	-.03	.037	[-0.10, 0.05]	.452
*T*2 intrinsic value on parental education	.08	.039	[0.01, 0.16]	.037
*T*2 intrinsic value on *t*1 intrinsic value	.44	.063	[0.32, 0.56]	< .001

Note. *N* = 504. 95% *CI* = 95% confidence interval; ASC = math ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 of *t*3 math grades = .41; R^2 of *t*2 intrinsic value = .51; [a]Because the total effect of this predictor was constrained to zero in this model, the regression weights of this predictor should not be interpreted.

Table A3

Detailed results for the beneficial competence model (w = 96.8%) with attainment value as a mediator

Parameter	β	*SE*	*95% CI*	*p*
Intercept of $t3$ grade	0.03	0.313	[-0.58, 0.65]	.922
$T3$ grade on $t2$ attainment value	.29	.042	[0.21, 0.37]	< .001
$T3$ grade on $t1$ linear ASC[a]	-.03	.021	[-0.08, 0.01]	.106
$T3$ grade on $t1$ linear competence	.17	.046	[0.08, 0.26]	< .001
$T3$ grade on $t1$ quadratic ASC[a]	-.00	.018	[-0.04, 0.03]	.865
$T3$ grade on $t1$ ASC/competence interaction[a]	.00	.018	[-0.04, 0.03]	.981
$T3$ grade on $t1$ quadratic competence[a]	.01	.011	[-0.02, 0.03]	.571
$T3$ grade on Gender	-.06	.039	[-0.14, 0.01]	.098
$T3$ grade on number of books	-.01	.039	[-0.09, 0.07]	.814
$T3$ grade on parental education	.10	0.41	[0.02, 0.18]	.017
$T3$ grade on $t0$ grade	.39	0.46	[0.30, 0.48]	< .001
Intercept of $t2$ intrinsic value	1.81	0.454	[0.92, 2.70]	< .001
$T2$ attainment value on $t1$ linear ASC[a]	.12	.067	[-0.01, 0.25]	.080
$T2$ attainment value on $t1$ linear competence[a]	.03	.053	[-0.07, 0.14]	.522
$T2$ attainment value on $t1$ quadratic ASC[a]	.01	.063	[-0.11, 0.14]	.865
$T2$ attainment value on $t1$ ASC/comp. interaction[a]	.00	.061	[-0.12, 0.12]	.981
$T2$ attainment value on $t1$ quadratic competence[a]	-.02	.038	[-0.10, 0.05]	.566
$T2$ attainment value on $t0$ grade	.05	.055	[-0.06, 0.16]	.383
$T2$ attainment value on gender	-.11	.043	[-0.19, -0.03]	.011
$T2$ attainment value on number of books	-.01	.043	[-0.10, 0.07]	.788
$T2$ attainment value on parental education	.02	.045	[-0.07, 0.10]	.724
$T2$ attainment value on $t1$ attainment value	.48	.058	[0.37, 0.59]	< .001

Note. $N = 504$. 95% *CI* = 95% confidence interval; ASC = math ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 of $t3$ math grades = .42; R^2 of $t2$ attainment value = .35; [a]Because the total effect of this predictor was constrained to zero in this model, the regression weights of this predictor should not be interpreted.

Table A4

Detailed results for the beneficial competence model (w = 97.3%) with utility value as a mediator

Parameter	*β*	*SE*	95% *CI*	*p*
Intercept of *t*3 grade	0.77	0.300	[0.18, 1.36]	.010
*T*3 grade on *t*2 utility value	.21	.042	[0.13, 0.29]	< .001
*T*3 grade on *t*1 linear ASC[a]	-.01	.014	[-0.04, 0.01]	.323
*T*3 grade on *t*1 linear competence	.19	.046	[0.10, 0.28]	< .001
*T*3 grade on *t*1 quadratic ASC[a]	-.01	.011	[-0.03, 0.01]	.483
*T*3 grade on *t*1 ASC/competence interaction[a]	-.00	.010	[-0.02, 0.02]	.831
*T*3 grade on *t*1 quadratic competence[a]	-.01	.008	[-0.02, 0.01]	.372
*T*3 grade on Gender	-.14	.040	[-0.22, -0.06]	< .001
*T*3 grade on number of books	-.02	.040	[-0.10, 0.06]	.618
*T*3 grade on parental education	.09	0.41	[0.01, 0.17]	.032
*T*3 grade on *t*0 grade	.42	0.44	[0.33, 0.50]	< .001
Intercept of *t*2 intrinsic value	0.52	0.335	[-0.14, 1.18]	.123
*T*2 attainment value on *t*1 linear ASC[a]	.07	.064	[-0.06, 0.19]	.299
*T*2 utility value on *t*1 linear competence[a]	.05	.053	[-0.06, 0.15]	.386
*T*2 utility value on *t*1 quadratic ASC[a]	.04	.053	[-0.07, 0.14]	.477
*T*2 utility value on *t*1 ASC/comp. interaction[a]	.01	.047	[-0.08, 0.10]	.831
*T*2 utility value on *t*1 quadratic competence[a]	.03	.037	[-0.04, 0.11]	.365
*T*2 utility value on *t*0 grade	.01	.053	[-0.09, 0.11]	.841
*T*2 utility value on gender	.11	.045	[0.02, 0.19]	.018
*T*2 utility value on number of books	.03	.045	[-0.05, 0.12]	.452
*T*2 utility value on parental education	-.02	.043	[-0.11, 0.06]	.579
*T*2 utility value on *t*1 utility value	.50	.048	[0.40, 0.59]	< .001

Note. *N* = 504. 95% *CI* = 95% confidence interval; ASC = math ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 of *t*3 math grades = .39; R^2 of *t*2 utility value = .33; [a]Because the total effect of this predictor was constrained to zero in this model, the regression weights of this predictor should not be interpreted.

Table A5

Detailed results for the full model (w = 5.9%) with intrinsic value as a mediator

Parameter	*β*	*SE*	95% *CI*	*p*
Intercept of *t*3 grade	0.31	0.329	[-0.33, 0.96]	.345
*T*3 grade on *t*2 intrinsic value	.35	.047	[0.26, 0.45]	< .001
*T*3 grade on *t*1 linear ASC	-.20	.059	[-0.32, -0.09]	.001
*T*3 grade on *t*1 linear competence	.21	.050	[0.11, 0.31]	< .001
*T*3 grade on *t*1 quadratic ASC	-.09	.049	[-0.19, 0.00]	.057
*T*3 grade on *t*1 ASC/competence interaction	-.00	.053	[-0.11, 0.10]	.971
*T*3 grade on *t*1 quadratic competence	.06	.056	[-0.05, 0.17]	.266
*T*3 grade on gender	-.11	.039	[-0.18, -0.03]	.007
*T*3 grade on number of books	.01	.039	[-0.07, 0.09]	.789
*T*3 grade on parental education	.06	0.41	[-0.02, 0.14]	.151
*T*3 grade on *t*0 grade	.41	0.50	[0.32, 0.51]	< .001
Intercept of *t*2 intrinsic value	1.38	0.321	[0.75, 2.01]	< .001
*T*2 intrinsic value on *t*1 linear ASC	.28	.074	[0.13, 0.42]	< .001
*T*2 intrinsic value on *t*1 linear competence	.01	.047	[-0.08, 0.11]	.766
*T*2 intrinsic value on *t*1 quadratic ASC	.00	.038	[-0.07, 0.08]	.915
*T*2 intrinsic value on *t*1 ASC/comp. interaction	-.01	.036	[-0.08, 0.06]	.825
*T*2 intrinsic value on *t*1 quadratic competence	-.02	.035	[-0.09, 0.05]	.522
*T*2 intrinsic value on *t*0 grade	-.00	.045	[-0.09, 0.08]	.930
*T*2 intrinsic value on gender	-.02	.036	[-0.09, 0.05]	.533
*T*2 intrinsic value on number of books	-.03	.037	[-0.10, 0.05]	.470
*T*2 intrinsic value on parental education	.08	.040	[0.01, 0.16]	.037
*T*2 intrinsic value on *t*1 attainment value	.46	.066	[0.33, 0.59]	< .001

Note. *N* = 504. 95% *CI* = 95% confidence interval; ASC = math ability self-concept; Gender: 1 = female, 2 = male; Parental Educational Level: 1 = No parent with Abitur or Fachabitur, 2 = At least one parent with Abitur or Fachabitur; Number of books at home: 1 = up to 100, 2 = more than 100; R^2 of *t*3 math grades = .42; R^2 of *t*2 intrinsic value = .51.

Online Supplemental Material

Overview

Supplement 1. Computation of double-entry intraclass correlations

Supplement 2. Results and discussion of double-entry intraclass correlations

Supplement 1. Computation of double-entry intraclass correlations

In all polynomial regression models that we analyzed (the full models as well the constrained models representing different SE bias hypotheses) we incorporated a set of covariates (gender, parental education, and number of books at the students' home) as additional predictors. We did so in order to estimate the effects of *t*1 math ASC and *t*1 math competence on *t*3 math grades mediated by *t*2 math expectancy of success and subjective task values when the covariates are controlled. However, the inclusion of covariates can alter the content of the other predictors (math ASC and math competence) as well as the mediators. In other words, it is not a prior clear what the ASC, competence, expectancy of success, and subjective task value variables actually represent in terms of their content after controlling for the covariates. If the inclusion of the covariates changed the content of these variables substantially, it would also compromise the interpretability of their effects on *t*3 math grades. Thus, in order to analyze the extent of this potential problem in the present study, we computed double-entry intraclass correlations (r_{ICC}s; Vize et al., 2018). First, we computed two different residual scores for each of the linear predictors (linear ASC and linear competence) and each of the mediators (expectancy of success, intrinsic task value, attainment task value, and utility task value). The first residual (residual 1) was computed by partialling out the following variables: linear ASC, linear competence, squared ASC, squared competence, and the ASC and competence interaction term. Note that these are the predictors of interest in our models in the main analyses. Of course, a variable was not partialled out from itself and thus, for example, linear ASC residual 1 was computed by partialling out linear competence, squared ASC, squared competence, and the ASC and competence interaction term. The second residual (residual 2) was computed by partialling out the same variables but also the covariates. Thus, residual 1 represents a predictor or mediator after the other predictors except for the covariates are partialled out. Residual 2 represents a predictor or mediator after the other predictors including the covariates are partialled out. Therefore, by comparing the two residual scores of the same construct, we can estimate the extent to which the inclusion of the covariates as additional predictors has affected the content of the other predictors and the mediators. Note that we only computed residuals of linear ASC and linear competence, but not of the nonlinear terms. It was not necessary to compute residual scores for the nonlinear terms because these predictors are computed directly from the linear terms. For example, if the linear math ASC variable still represents the same construct (linear math ASC) after partialling out the covariates, then it follows that the squared math ASC variable also still represents squared math ASC. Next, we selected some variables from the dataset that were not used in the present study for other purposes. Note that the present study was part of a larger project as explained in the main article. A list of these variables and the instruments used for their assessment is reported in Table S1. We correlated these selected variables from the dataset with the different residual scores. Finally, we computed the double-entry intraclass correlations between these correlations. Double-entry intraclass correlations serve as similarity index. In this case, they represent how similar the correlations of the residual 1 scores with the other variables and the correlations between the residual 2 scores and the other variables are. If the correlations with other variables do not change much after partialling out the covariates (and thus, if the double-intraclass correlations are high), that is an indication that the original scores and the residuals are embedded in a similar nomological network and thus, can be interpreted in the same or at least a similar way. For example, a high double-entry intraclass correlation between linear ASC residual 1 and linear ASC residual 2 would imply that the correlations between linear ASC with other variables (see Table S1) are similar before and after additionally partialling out the covariates.

Table S1

Constructs correlated with the predictors and mediators in the models as well as their residuals

Construct	Measurement instrument
Neuroticism Extraversion Openness Agreeableness Conscientiousness	NEO five factor inventory (NEO-FFI; Borkenau & Ostendorf, 2008)
Math self-efficacy	
Math interest	
Math intrinsic motivation	
Math test anxiety agitation Math test anxiety worry	Short version (Schwarzer & Jerusalem, 1999) of the German test anxiety inventory (TAI-G; Hodapp,1991,1996)
Practical/technical orientation Intellectual/researching orientation Artistic/linguistic orientation Social orientation Business orientation Conventional orientation	General interest structure test with environment structure test (AIST-R/UIST-R; Bergmann & Eder, 2005)
Basic Arithmetic skills	Additional test for assessing basic arithmetic skills (KRW) published together with the DEMAT 9 (Schmidt et al., 2012)
General Intelligence	CFT 20-R with WS/ZF-R. General Intelligence scale 2 – Revision (CFT 20-R) with vocabulary (WS) and numerical order test – Revision (WS/ZF-R) (CFT20-R; Weiß, 2006)
General Intelligence	PSB-R 6-13 – Inspection system for school and education consultation for Grades 6 to 13 – revised version (PSB 6-13; Horn et al., 2003)
Perceptual Speed	The Connecting numbers test (ZVT). A language free intelligence test for measuring “cognitive speed”. Manual instruction, 2. revised version (ZVT; Oswald & Roth, 1987)

Supplement 2. Results and discussion of double-entry intraclass correlations

The correlations of the predictors and mediators with other variables were highly similar before and after additionally partialling out the covariates (all $\Delta r \leq .05$; see Table S2). Moreover, the double-entry intraclass correlations between different residuals of the same constructs were all very high (all $r_{ICC} \geq .97$; see Table S2). Therefore, additionally partialliang out the covariates in the models in the main analyses had virtually no effect on the content of the predictors (linear and nonlinear ASC and competence terms) and the mediators (expectancy of success, subjective task values). These variables are apparently still embedded in a highly similar nomological network even when the influence of the covariates is controlled. Thus, the effects of these predictors and mediators on *t*3 math grades in the main analyses is not compromised by the inclusion of the covariates.

Table S2

Correlations between residualized predictors and mediators with other constructs and their double-entry intraclass correlations (r_{ICC})

	Linear ASC		Linear comp.		EOS		ITV		ATV		UTV	
	Res. 1	Res.2	Res. 1	Res. 2	Res. 1	Res. 2	Res. 1	Res. 2	Res. 1	Res. 2	Res. 1	Res. 2
Neuroticism	-.09	-.07	.08	.05	-.01	.00	.01	.00	.03	.03	.01	.00
Extraversion	.02	.02	-.02	-.02	.05	.02	.07	.05	-.11*	-.14**	-.01	-.03
Openness	.01	.01	-.03	-.03	.03	.00	-.02	-.03	.07	.07	.00	-.01
Agreeableness	.08	.07	-.06	-.04	-.02	-.01	-.03	-.03	-.01	.00	.02	.04
Conscientiousness	.07	.07	-.01	-.01	.01	-.02	-.02	-.02	-.04	-.07	-.04	-.05
Math test anxiety agitation	-.01	.02	.03	.02	-.03	-.04	.00	.00	.05	.05	.08	.09*
Math test anxiety worry	-.03	-.01	.03	.03	-.03	-.04	-.01	-.02	-.01	.00	.02	.03
Practical/technical orientation	.01	.00	.03	.05	.03	.02	.03	.02	.03	.02	.06	.05
Intellectual/researching orientation	.03	.01	.00	.02	.01	.01	-.04	-.05	.03	.02	.01	.00
Artistic/linguistic orientation	.03	.04	-.01	.02	-.01	-.03	-.01	-.01	.02	.01	-.01	.00
Social orientation	.04	.05	-.09*	-.07	.03	.02	-.01	.00	-.01	-.01	.03	.05
Business orientation	.03	.02	-.03	-.03	.07	.06	.00	-.01	-.08	-.08	.03	.03
Conventional orientation	-.01	-.01	.02	.02	-.01	-.03	.01	.00	-.05	-.06	-.01	-.02
Math interest	.57***	.55***	-.02	-.03	.21***	.21***	.44***	.44***	.20**	.19**	.22***	.21**
Math self-efficacy	.66***	.64***	.02	.01	.21***	.20***	.14**	.14**	.12**	.12**	.10*	.08
Math intrinsic motivation	.49***	.47***	-.01	-.02	.17***	.16***	.43***	.44***	.26***	.26***	.29***	.28***
Math mastery approach goals	.36****	.38***	.00	.05	.20***	.16***	.24***	.24***	.47***	.43***	.27***	.26***
Math performance approach goals	.37***	.38***	.11*	.11	.19***	.17***	.08	.09*	.37***	.33***	.16***	.13**
Math mastery avoidance goals	.12**	.14**	.05	.08	.12**	.10*	.08	.07	.32***	.28***	.17***	.15**
Math performance avoidance goals	.28***	.30***	.12**	.14**	.18***	.17***	.06	.07	.38***	.34***	.15**	.13**
Basic arithmetic	.01	.00	-.08	-.09*	.06	.08	.02	.02	.02	.00	-.01	-.01
General intelligence (CFT)	.07	.07	-.11*	-.11*	-.01	.01	-.03	-.03	.02	.00	.05	.06
General intelligence (PSB)	.10*	.08	-.13**	-.11*	.02	.02	.07	.07	-.03	-.03	-.01	-.01
Perceptual speed	.00	.00	-.08	-.08	-.04	-.03	.04	.05	-.01	.01	.00	.02
Similarity indices (r_{ICC})												
Linear ASC Res. 1		**1.00*****		.20		.90***		.81***		.66***		.73***
Linear comp. Res. 1		.25		**.97*****		.33		.09		.52**		.30
EOS Res. 1		.90***		.39***		**.99*****		.74***		.78***		.75***
ITV Res. 1		.80***		.07		.74***		**1.00*****		.58**		.80***
ATV res. 1		.68***		.60**		.77***		.57**		**1.00*****		.84***
UTV res. 1		.78***		.39		.79***		.83***		.88***		**.99*****

Note. N = 236-504; ASC = ability self-concept; comp. = competence; EOS = expectancy of success; ITV = intrinsic task value; ATV = attainment task value; UTV = utility task value; Res. 1 = residual after partialling out all other predictors (linear ASC, linear competence, quadratic ASC, quadratic competence, ASC and competence interaction); Res. 2 = residual after partialling out all other predictors as well as gender, parental education, and number of books at home. Double-entry intraclass correlations between different residuals of the same construct are highlighted in **bold**.

References

Bergmann, C., & Eder, F. (2005). *Allgemeiner Interessen-Struktur-Test mit Umwelt-Struktur-Test (AIST-R/UST-R) [General Interest Structure Test with Environment Structure Test (AIST-R/UST-R)]*. Beltz Test.

Borkenau, P., & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI) - Manual [NEO Five Factor Inventory (NEO-FFI) by Costa and McCrae - Manual], 2nd ed. (rev)*. Hogrefe.

Hodapp, V. (1991). Das Prüfungsängstlichkeitsinventar TAI-G: Eine erweiterte und modifizierte Version mit vier Komponenten [The test anxiety inventory TAI-G: An extended and modified version with four components]. *Zeitschrift für Pädagogische Psychologie, 5*, 121–130.

Hodapp, V. (1996). The TAI-G: A multidimensional approach to the assessment of test anxiety. In C. Schwarzer, & M. Zeidner (Eds.), *Stress, anxiety, and coping in academic settings* (pp. 95-130). Francke.

Horn, W., Lukesch, H., Mayrhofer, S., & Kormann, A. (2003). *PSB 6-13 – Prüfsystem für Schul- und Bildungsberatung für 6. Bis 13. Klassen – revidierte Fassung [PSB-R 6-13 – Inspection system for school and education consultation for Grades 6 to 13 – revised version]*. Hogrefe.

Oswald, W. D., & Roth, E. (1987). *Der Zahlen-Verbindungs-Test (ZVT). Ein sprachfreier Intelligenz-Test zur Messung der „kognitiven Leistungsgeschwindigkeit“. Handanweisung. 2., überarbeitete und erweiterte Auflage [The Connecting numbers test (ZVT). A language free intelligence test for measuring "cognitive speed". Manual* *instruction, 2. revised version].* Hogrefe.

Schmidt, S., Ennemoser, M., & Krajewski, K. (2012). *Deutscher Mathematiktest für neunte Klassen (DEMAT 9) [German Mathematics Test for Ninth Grade (DEMAT 9)].* Hogrefe.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for the assessment of teacher and student attributes. Documentation of psychometric instruments as part of the scientific support of the pilot project self-efficient schools].* Freie Universität Berlin.

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2018). Examining the effects of controlling for shared variance among the dark triad using meta-analytic structural equation modelling. *European Journal of Personality, 32*, 46-61. do: 10.1002/per.2137

Weiß, R. H. (2006). *CFT 20-R mit WS/ZF-R Grundintelligenztest Skala 2 - Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest - Revision (WS/ZF-R) [CFT 20-R with WS/ZF-R. General Intelligence scale 2 – Revision (CFT 20-R) with vocabulary (WS) and numerical* order test – Revision (WS/ZF-R)]. Hogrefe.

3.2 Zusammenfassung und Vergleich der empirischen Beiträge

Das zentrale Ergebnis von Beitrag I ist die Tatsache, dass unter Verwendung eines Ansatzes, der SE Bias Effekte von Selbsteinschätzungseffekten trennt, lediglich positive lineare Selbsteinschätzungseffekte aber keine linearen SE Bias Effekte auf Noten in Mathematik und Deutsch gefunden wurden. Zudem wurde ein signifikanter Effekt der Mathematikkompetenz auf die Mathematiknote, nicht aber ein Effekt der Deutschkompetenz auf die Deutschnote gefunden. Zu bedenken ist dabei allerdings, dass der Effekt der Deutschkompetenz größer war als jener der Mathematikkompetenz, aufgrund des größeren Standardfehlers aber keine Signifikanz erreichte. Dies mag auf die im Vergleich zur Mathematikkompetenz komplexere Erfassung der Deutschkompetenz zurückzuführen sein. Da nur für die Mathematikkompetenz ein valider Test zur Verfügung stand, der das gesamte Konstrukt abbildet (Aufgaben aus TIMSS; Baumert et al., 1998), mussten zur Erfassung der Deutschkompetenz stattdessen die Ergebnisse eines Lesetests (LGVT 6-12; Schneider et al., 2007) und eines Rechtschreibtests (RT; Kersting & Althoff, 2004) zusammengefasst werden. Zu beachten ist außerdem, dass in derselben Stichprobe eines von zwei „klassischen Verfahren" zur Untersuchung von SE Bias Effekten (Residualwerte) zum Befund signifikanter positiver SE Bias Effekte führte. Aufgrund der statistischen Schwächen dieses Ansatzes (Humberg et al., 2018; 2019a) weist dieser Befund auf die Bedeutung des Einsatzes modernerer statistischer Verfahren hin.

Beitrag II stellt eine Erweiterung von Beitrag I dar, da in dieser Studie auch nonlineare Effekte des Fähigkeitsselbstkonzepts, der Kompetenz und des SE Bias untersucht werden. Dies ist von Bedeutung, da einige Autor*innen die Annahme vertreten, dass derartige nonlineare Effekte bestehen (z.B. Baumeister, 1989; Helmke, 1998; Taylor & Brown 1994) und entsprechende Hypothesen zudem aus Befunden zum selbstregulierten Lernen abgeleitet werden können (z.B. Dunlosky & Rawson, 2012; Hacker & Bol, 2019; Hadwin & Webster, 2013; van

Loon & Oeri, 2023). In Beitrag I wurden nonlineare Effekte noch nicht untersucht, da stattdessen eine zu diesem Zeitpunkt neue Methode, die condition-based regression analysis (Humberg et al., 2018), verwendet wurde, welche erstmalig erlaubte, SE Bias Effekte im Ein-Schritt-Ansatz auf Signifikanz zu prüfen, allerdings nur geeignet ist, um lineare Effekte zu untersuchen. In Beitrag II wurde dann ein ebenfalls zu diesem Zeitpunkt neues methodisches Rahmenkonzept (Humberg et al., 2019a) eingesetzt, welches durch Einsatz der response surface analysis (Edwards & Parry, 1993; Edwards, 2002) und informationstheoretischer Modellvergleiche anhand von Akaike-Gewichten erlaubte, auch nonlineare Effekte quantitativ miteinander zu vergleichen. Ein weiterer Unterschied besteht darin, dass in Beitrag II (wie auch in Beitrag III) im Gegensatz zu Beitrag I nur Effekte im Fach Mathematik, nicht aber im Fach Deutsch untersucht wurden. Der Grund dafür war die zuvor angesprochene schwierige und problematische Erfassung der Kompetenz in Deutsch. In Beitrag II wurde ursprünglich ebenfalls versucht, die Deutschkompetenz über einen Lese- und einen Rechtschreibtest zu erfassen. Allerdings führte dieses Vorgehen zu einem Messmodell der Deutschkompetenz mit inakzeptablen Fit-Indizes, weshalb auf diese Untersuchung verzichtet wurde. Die Ergebnisse der Modellvergleiche in Mathematik wiesen darauf hin, dass lediglich positive lineare Effekte sowohl des Fähigkeitsselbstkonzepts als auch der Kompetenz auf die Deutschnoten vorliegen, aber keine nonlinearen Effekte oder SE Bias Effekte. Somit bestätigen die Befunde jene aus Beitrag I und erweitern Sie zudem, da nun auch nonlineare SE Bias Effekte sowie nonlineare Effekte des Fähigkeitsselbstkonzepts und der Kompetenz ausgeschlossen werden konnten.

In der weiterführenden Analyse (im weiteren Textverlauf zwecks sprachlicher Einfachheit als Beitrag III bezeichnet) wurde dieselbe Stichprobe untersucht wie in Beitrag II allerdings mit einem zusätzlichen Messzeitpunkt und weiteren erfassten Schüler*innenmerkmalen. Vertreter*innen von SE Bias Effekten (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016;

Helmke, 1998; Leduc & Bouffard, 2017; Lee, 2021; 2022) wie auch Vertreter*innen von Selbsteinschätzungseffekten (z.B. Eccles & Wigfield, 2023; Marsh & Martin, 2011; Wigfield & Eccles, 2020) gehen davon aus, dass Selbsteinschätzungsffekte und SE Bias Effekte auf akademische Leistung von motivationalen Variablen mediiert werden. Laut dem Erwartungs-Wert-Modell (Eccles & Wigfield, 2023; Wigfield & Eccles, 2020) sind es insbesondere die Erfolgserwartungen und subjektiven Werte, die eine Rolle spielen. Daher sollten diese Annahmen in Beitrag III überprüft werden. Als abhängige Variable wurden die Mathematiknoten am Ende der 11. Klasse (*t*3) statt wie in Studie 2 am Ende der 10. Klasse gewählt. Die Erfolgserwartung und die subjektiven Werte der Schüler*innen in Mathematik wurden als Mediatorvariablen zu *t*2 erfasst. Die Ergebnisse bestätigten die Befunde aus den Beiträgen I und II, dass keine SE Bias Effekte auf akademische Leistung vorliegen. Zudem konnte gezeigt werden, dass die Effekte der Kompetenz, welche in Beitrag II und teilweise in Beitrag I nachgewiesen wurden, partiell durch Erfolgserwartungen, nicht aber durch subjektive Werte mediiert werden. Im Gegensatz zu den Beiträgen I und II, gab es in Beitrag III kaum Evidenz für Effekte des Fähigkeitsselbstkonzepts.

Insgesamt verdeutlichen die Befunde der drei Beiträge vor allem eines: Verwendet man Methoden, die eine Trennung von Selbsteinschätzungseffekten und SE Bias Effekten auf Schulnoten erlauben, zeigen sich keinerlei Hinweise auf SE Bias Effekte. Stattdessen finden sich lediglich positive lineare Effekte der Selbsteinschätzung (Beiträge I und II) und der Kompetenz (Beiträge II und III, Beitrag I teilweise).

3.3 Beantwortung der Hypothesen und Fragestellungen

In Anbetracht der oben beschriebenen Befunde können die in der vorliegenden Arbeit aufgestellten Hypothesen und Fragestellungen wie folgt evaluiert und beantwortet werden.

Hypothese 1, dass ein höheres Fähigkeitsselbstkonzept innerhalb einer Domäne (z.B. Mathematik oder Deutsch) zu besseren zukünftigen Noten führt, selbst wenn die objektiv erfasste

Kompetenz und der Ausgangswert der Note kontrolliert werden, kann durch die Beiträge I und II insgesamt bestätigt werden. Die Ergebnisse von Beitrag III widersprechen dieser Hypothese zwar, doch können diese unter Umständen auf die spezifische Methodik dieser Studie zurückgeführt werden (siehe Kapitel 3.3.1).

Die Hypothesen 2a bis 2d, dass die Erfolgserwartung sowie die subjektiven Werte den Effekt des Fähigkeitsselbstkonzepts auf die Noten mediieren, konnten in Beitrag III zurückgewiesen werden, da der Effekt des Fähigkeitsselbstkonzepts in diesem Beitrag erwartungswidrig ausblieb. Allerdings ist zu bedenken, dass dieser Befund aufgrund der Tatsache, dass zahlreiche vorige Studien Effekte des Fähigkeitsselbstkonzepts auf Noten gefunden haben, mit Vorsicht zu interpretieren ist.

Fragestellung 1, die Frage ob es SE Bias Effekte auf Schulnoten gibt, kann unter Betrachtung aller drei Studien eindeutig beantwortet werden. Es liegen in der vorliegenden Arbeit keine nachweisbaren SE Bias Effekte vor.

Da in Beitrag III keine SE Bias Effekte vorlagen, konnten zudem die Fragestellungen 2a bis 2d beantwortet werden: Es liegt in der vorliegenden Arbeit keine Mediation von SE Bias Effekten durch Erfolgserwartungen oder subjektive Werte vor.

Im Folgenden werden die einzelnen Hypothesen und Fragestellungen mit Hinblick auf die Ergebnisse der drei empirischen Beiträge näher diskutiert.

3.3.1 Hypothese 1 – Effekte des Fähigkeitsselbstkonzepts auf Schulnoten

Aufgrund einer Vielzahl an Einzelarbeiten (z.B. Arens & Niepel, 2023; Bakadorova & Raufelder, 2020; Marsh 2022; Marsh, 2023; Marsh et al., 2022; Seaton et al., 2015; Sewasew & Koester, 2019; Sewasew & Schroeders, 2019; Weidinger et al., 2018; Wolff et al., 2021a; Zhang et al., 2023) sowie zweier Metaanalysen (Valentine et al., 2004; Wu et al., 2021), in denen Effekte des Fähigkeitsselbstkonzepts auf akademische Leistung inklusive Noten nachgewiesen wurden, erschien es plausibel entsprechende Effekte auch in der vorliegenden Arbeit zu erwarten. Somit sind die Ergebnisse der ersten beiden Beiträge in diesem Punkt konform mit Hypothese 1 und der bisherigen Literatur, während die Ergebnisse von Beitrag III ihnen widersprechen.

Dies könnte zum einen auf den unterschiedlichen zeitlichen Abstand zwischen der Erfassung des Fähigkeitsselbstkonzepts und der Noten zurückzuführen sein. In den Beiträgen I und II wurde der Effekt des Fähigkeitsselbstkonzepts zu Beginn der 10. Klasse auf die Zeugnisnoten am Ende der 10. Klasse untersucht. Hingegen wurde in Beitrag III der Effekt auf die Zeugnisnoten am Ende der 11. Klasse untersucht. Es erscheint plausibel, dass die Auswirkung des zu einem bestimmten Zeitpunkt erfassten Fähigkeitsselbstkonzepts mit der Zeit abnehmen sollte. Zwar ist die Positionsstabilität akademischer Fähigkeitsselbstkonzepte relativ hoch, vergleichbar mit jener von grundlegenden Persönlichkeitseigenschaften wie jenen im Big 5 Persönlichkeitsmodell (Asendorpf & van Aken, 2003; siehe Trautwein & Möller, 2016), doch lässt sie absolut betrachtet durchaus Raum für Veränderungen. So fanden Jansen et al. (2020, S. 1622) eine Retest-Stabilität des mathematischen Fähigkeitsselbstkonzepts über einen Zeitraum von einem Jahr von $\rho = .62$ und einen geschätzten wahren Wert von $\rho_{true} = .76$. Somit wäre unter der Annahme einer im Altersbereich der Proband*innen gleich bleibenden Stabilität eine Zwei-Jahres-Stabilität von etwa $\rho_{true} = .76 * .76 = .58$ zu erwarten. Es ist somit anzunehmen, dass es innerhalb eines Intervalls von fast zwei Jahren wie in Beitrag III zu substantiellen interindividuellen Veränderungen im mathematischen Fähigkeitsselbstkonzept kommen kann. Diese interindividuellen Veränderungen dürften den Effekt des zu $t1$ gemessenen Fähigkeitsselbstkonzepts auf die fast zwei Jahre später erfassten Noten verringern und würden somit erklären, weshalb die Effekte des Fähigkeitsselbstkonzepts nur in den Beiträgen I und II, nicht aber in Beitrag III, gefunden wurden. Allerdings fanden Wu et al. (2021) in einer Metaanalyse, dass das Zeitinterval zwischen der Messung von Selbsteinschätzungen, inklusive des Fähigkeitsselbstkonzepts, und der Messung der akademischen Leistung keinen bedeutsamen Moderator des Selbsteinschätzungseffekts auf akademische Leistung darstellt.

Ein anderer mit dem Zeitintervall in Verbindung stehender Erklärungsansatz bezieht sich auf die Bildung des Fähigkeitsselbstkonzepts durch Vergleichsprozesse und die Veränderung des Klassen- und Kurssystems an Schulen in der 11. Klasse. Bei der Bildung des Fähigkeitsselbstkonzepts spielen soziale, temporale und dimensionale Vergleichsprozesse eine Rolle (Basarkod et al., 2022; Marsh et al., 2008; Möller et al., 2015; Müller-Kalthoff et al., 2017; Ross & Wilson, 2003; Trautwein & Möller, 2016; Wolff et al., 2018; Wolff & Möller, 2022). Die Schüler*innen der Stichprobe aus Beitrag III (wie auch jene aus den anderen Beiträgen) wurden nach dem G12-System unterrichtet, weshalb sie in der 11. Klassenstufe in die Qualifikationsphase der gymnasialen Oberstufe eintraten und erstmals Grund- und Leistungskurse wählten. Dieser neue Unterrichtskontext könnte sich auf alle angesprochenen Vergleichsprozesse ausgewirkt haben. Bezogen auf soziale Vergleichsprozesse ist anzunehmen, dass für Schüler*innen in Leistungskursen der Vergleich mit im Mittel mathematisch kompetenteren Mitschüler*innen salienter wurde, während für Schüler*innen in Grundkursen der Vergleich mit im Mittel weniger mathematisch kompetenten Mitschüler*innen salienter wurde. Dies könnte im Sinne eines Fischteicheffekts (z.B. Fleischmann et al., 2022; Jansen et al., 2022; Marsh et al., 2021; Parker et al., 2021; Trautwein & Möller, 2016; Wang, 2020; Werts & Watley, 1969; Wolff et al., 2021b; Zell & Lesick, 2021) zum Absinken der Fähigkeitsselbstkonzepte von Schüler*innen in Leistungskursen und zur Zunahme der Fähigkeitsselbstkonzepte von Schüler*innen in Grundkursen führen. Der Fischteicheffekt in Klassen scheint fast ausschließlich vom durchschnittlichen Leistungsstand der Klasse abzuhängen, aber unabhängig davon zu sein, mit welchen Peers Schüler*innen am häufigsten zusammenarbeiten und Zeit verbringen (Jansen et al., 2022). Ebenso scheint über den durchschnittlichen Leistungsstand der Klasse hinaus, das durchschnittliche Leistungsniveau der Schule keinen Einfluss zu haben (Fleischmann et al., 2022). Somit ist anzunehmen, dass der Fischteicheffekt innerhalb von neu gewählten Leistungs- und Grundkursen trotz potentiell konstanterer durchschnittlicher Leistungsniveaus in Peerbeziehungen und an der Schule als Ganzes einen Einfluss auf das Fähigkeitsselbstkonzept von Schüler*innen haben könnte. Eine solche Veränderung des Fähigkeitsselbstkonzepts mit Eintritt in die Grund- und Leistungskurse sollte zu einer verminderten Bedeutung des vor den Kurswahlen bestehenden Fähigkeitsselbstkonzepts führen. Temporale Vergleichsprozesse können eine Rolle spielen, wenn Schüler*innen ihre eigenen Leistungen in den Grund- und Leistungskursen mit jenen vor der 11. Klasse vergleichen. Sofern das höhere Niveau in Leistungskursen dazu führt, dass Schüler*innen bei gleicher Kompetenz und Anstrengung schlechtere Leistungen erzielen, sollte dies zu einer Abnahme des Fähigkeitsselbstkonzepts aufgrund des Aufwärtsvergleichs mit früheren höheren Leistungen führen (Theorie temporaler Vergleiche; Albert, 1977; Müller-Kalthoff et al., 2017; Wolff et al., 2018; Wolff et al., 2019; Wolff et al., 2021b; Wolff & Möller, 2022). Ebenso sollten höhere Anforderungen in Leistungs- im Vergleich zu Grundkursen dazu führen, dass sich die dimensionalen Vergleichsprozesse verschieben. Erzielt beispielsweise ein*e Schüler*in in der 10. Klasse vergleichbare Leistungen in Mathematik und Deutsch, könnte die Wahl eines Mathematikleistungskurses und eines Deutschgrundkurses dazu führen, dass die Leistungen in Mathematik aufgrund der höheren Anforderungen des Leistungskurses gegenüber den Leistungen in Deutsch abfallen. Dies sollte im Sinne der Theorie dimensionaler Vergleichsprozesse zu einer Aufwertung des Fähigkeitsselbstkonzepts in Deutsch und einer Abwertung des Fähigkeitsselbstkonzepts in Mathematik führen (Marsh et al., 2021; Möller et al., 2020; Möller & Köller, 1998; Stocker et al., 2021; van der Westhuizen et al., 2022; Wan et al., 2021; 2023; Wolff, 2022; Wolff et al., 2020; 2021a; 2021b; 2021c; Wolff & Möller, 2021; 2022). Insgesamt könnte die Wahl von Grund- und

Leistungskursen zu Beginn der 11. Klasse somit zu einer Neubewertung der eigenen domänenspezifischen Fähigkeitsselbstkonzepte der Schüler*innen geführt haben, sodass das Fähigkeitsselbstkonzept zu Beginn der 10. Klasse an Relevanz verlor. Allerdings zeigten in Beitrag III nachträglich durchgeführte Analysen, dass die Ergebnisse sich unter Kontrolle der Kurszugehörigkeit nur geringfügig änderten und weiterhin keine signifikanten Effekte des Fähigkeitsselbstkonzepts auf nachfolgende Noten gefunden wurden.

Ein weiterer Grund für das Ausbleiben eines Effekts des Fähigkeitsselbstkonzepts in Beitrag III könnten Suppressionseffekte aufgrund von Multikollinearität zwischen den Prädiktoren sein (siehe Marsh et al., 2004). Das Fähigkeitsselbstkonzept wies in Beitrag III hohe Korrelationen mit der zur selben Zeit erfassten Kompetenz ($r = .52$, $p < .001$), Erfolgserwartung ($r = .78$, $p < .001$), den intrinsischen Werten ($r = .82$, $p < .001$), Wichtigkeitswerten ($r = .57$, $p < .001$), Nützlichkeitswerten ($r = .56$, $p < .001$) und Noten ($r = .63$, $p < .001$) auf. Da viele dieser Variablen in den selben Modellen als Prädiktoren der späteren Noten eingesetzt wurden, ist denkbar, dass aufgrund der hohen Multikollinearität nicht genug unabhängige Varianz im Fähigkeitsselbstkonzept übrig blieb, um die späteren Noten vorherzusagen. Allerdings zeigte eine in Beitrag III nachträglich durchgeführte Analyse, dass der Effekt des Fähigkeitsselbstkonzepts auf die Noten am Ende der 11. Klasse zwar zunahm, wenn alle übrigen Prädiktoren außer den Noten am Ende der neunten Klasse aus der Regression entfernt wurden, der Effekt aber dennoch keine Signifikanz erreichte. Die bivariate Korrelation zwischen dem Fähigkeitsselbstkonzept zu Beginn und den Noten am Ende der 11. Klasse war hingegen signifikant ($r = .39$, $p < .001$). Außerdem ist zu bedenken, dass Suppressionseffekte nicht zwangsläufig ein statistisches Problem darstellen (Humberg et al., 2022). Ein Suppressionseffekt liegt vor, wenn das Gewicht eines Prädiktors in einem Modell durch die Hinzunahme eines weiteren Prädiktors stark verändert wird, da bestimmte Varianzanteile des ersten Prädiktors nicht mehr in die Vorhersagekraft dieses

Prädiktors eingehen (sie unterdrückt werden). Dabei kann der Effekt der ersten Variable größer werden, sofern für die Vorhersage irrelevante Aspekte unterdrückt werden, oder kleiner, sofern relevante Aspekte unterdrückt werden. Auch eine Vorzeichenumkehr ist möglich. Dies bedeutet jedoch nicht, dass die Ergebnisse fehlerhaft sind, sondern einer anderen und potentiell herausfordernderen Interpretation bedürfen (Humberg et al., 2022). Wird beispielweise, wie in Beitrag III, der direkte Effekt des Fähigkeitsselbstkonzepts auf die Noten unter Einschluss der intrinsischen Werte (sowie der weiteren zusätzlichen Prädiktoren) negativ, bedeutet dies nicht, dass ein fehlerhafter Befund vorliegt. Stattdessen bedeutet es, dass diejenigen Varianzanteile des Fähigkeitsselbstkonzepts, die nicht mit den oben genannten Variablen in Zusammenhang stehen, die akademische Leistung negativ vorhersagen. Dies kann problematisch für die Interpretation der Befunde sein, sofern nicht klar ist, beziehungsweise keine sinnvollen aus Theorien abgeleiteten Hypothesen darüber aufgestellt werden können, was der nach Kontrolle der übrigen Variablen übrig gebliebene Varianzanteil des Fähigkeitsselbstkonzepts inhaltlich repräsentiert. Falls hingegen theoretisch begründete Hypothesen abgeleitet werden können, können entsprechende Befunde Hinweise darauf geben, welche Aspekte des Fähigkeitsselbstkonzepts förderlich und welche potentiell hinderlich für akademische Leistung sind. Eine entsprechende Diskussion zu unterschiedlichen Aspekten des Fähigkeitsselbstkonzepts und ihren Einflüssen auf akademische Leistung erfolgt im folgenden Kapitel. An dieser Stelle ist lediglich relevant, dass Suppressionseffekte eine Herausforderung bei der Interpretation von Ergebnissen, aber in der Regel keine fehlerhaften Ergebnisse bedeuten. Zudem sind SE Bias Effekte nach der Logik der CRA und RSA von Suppressionseffekten zwischen der Selbsteinschätzung und Kompetenz unabhängig (Humberg et al., 2022). Dies ist darauf zurückzuführen, dass SE Bias Effekte aufgrund ihrer Konzeptionalisierung immer voraussetzen, dass Selbsteinschätzung und Kompetenz unter Kontrolle der jeweils anderen Variablen einen Effekt auf die abhängige

Variable ausüben. Beispielsweise liegt ein positiver SE Bias Effekt, wie in Abschnitt 1 gezeigt, dann und nur dann vor, wenn die Selbsteinschätzung unter Kontrolle der Kompetenz einen positiven Effekt und die Kompetenz unter Kontrolle der Selbsteinschätzung einen negativen Effekt ausübt (Humberg et al., 2018). Bei der Untersuchung eines SE Bias geht es also stets darum, die Zusammenhänge zwischen drei Variablen (Selbsteinschätzung, Kompetenz und abhängige Variable, z.B. akademische Leistung) zu beschreiben, nicht den Zusammenhang zwischen nur zwei Variablen zum Beispiel unter Wegfall der Kompetenz als zusätzlichem Prädiktor. Suppressionseffekte zwischen Fähigkeitsselbstkonzept und Kompetenz sind innerhalb der CRA und RSA also, sofern sie vorliegen, kein statistisches Problem, das behoben werden muss, sondern ein inhaltlich relevanter und zu berücksichtigender Teilaspekt von SE Bias Effekten (Humberg et al., 2022, S. 886-887).

Denkbar wäre außerdem, dass die Noten zwischen der neunten und 11. Klasse zu stabil waren, als dass das Fähigkeitsselbstkonzept die Noten der 11. Klasse bedeutsam vorhersagen konnte. Allerdings sagten andere Variablen, wie die Kompetenz, die Erfolgserwartung und alle drei subjektiven Wertekomponenten, die Noten der 11. Klasse auch über die vorigen Noten hinaus signifikant vorher. Zudem war die Korrelation zwischen den Noten der neunten und 11. Klasse zwar hoch ($r = .53$, $p < .001$), aber nicht in einem Ausmaß, das keine darüber hinausgehende Vorhersage der Noten der 11. Klasse erlauben würde.

Letztlich ist zu vermuten, dass eine Kombination der oben genannten Faktoren (längeres Intervall zwischen den Messungen, Veränderungen der Fähigkeitsselbstkonzepte durch Wahl von Leistungs- und Grundkursen, Suppressionseffekte aufgrund von Multikollinearität zwischen den Prädiktoren und die relativ hohe Stabilität der Noten) zu dem Ausbleiben eines Effekts des Fähigkeitsselbstkonzepts in Beitrag III geführt hat. Zu Bedenken ist außerdem, dass die in Metaanalysen gefundenen durchschnittlichen Effekte auf Noten von Selbsteinschätzungen im

Allgemeinen (β = .08; Valentine et al., 2004) wie auch vom Fähigkeitsselbstkonzept im Speziellen (β = .08; Wu et al., 2021) gering ausfallen. Im Vergleich dazu sind die in den Beiträgen I und II gefunden Effekte des Fähigkeitsselbstkonzepts auf Noten größer (.19 ≤ β ≤ .32). Auch wenn eine simple Berechnung des Mittelwerts der in den drei Beiträgen gefundenen Effektstärken nicht zulässig ist, deuten die Befunde aller drei Beiträge zusammengenommen somit doch darauf hin, dass die gefundenen Ergebnisse um den metaanalytisch bestimmten Durchschnittswert streuen und mit diesem vergleichbar sind. Somit kann Hypothese 1 insgesamt unter Einschränkungen als gestützt angesehen werden.

3.3.2 Hypothesen 2a bis 2d – Mediation von Effekten des Fähigkeitsselbstkonzepts

Die Mediation von Effekten des Fähigkeitsselbstkonzepts auf Noten durch Erfolgserwartungen und subjektive Werte wurde in Beitrag III untersucht. Da in Beitrag III keine Effekte des Fähigkeitsselbstkonzepts auf Noten gefunden wurden, gab es somit auch keine Hinweise auf eine Mediation, sodass die entsprechenden Hypothesen zurückgewiesen werden. Die einzige Ausnahme stellte die Analyse von intrinsischen Werten als Mediator dar, in der das vollständige Modell Teil des Konfidenzsets war. In diesem Modell zeigte sich ein signifikanter indirekter und positiver Effekt des Fähigkeitsselbstkonzepts auf die Noten, welcher von den intrinsischen Werten mediiert wurde. Da sich aber gleichzeitig der direkte Effekt des Fähigkeitsselbstkonzepts auf die Noten als negativ erwies, war der Gesamteffekt nicht signifikant. Inhaltlich würde dies bedeuten, dass ein gewisser Aspekt des Fähigkeitsselbstkonzepts sich positiv auf intrinische Werte auswirkt, während ein anderer Aspekt einen direkten negativen Effekt auf die Noten ausübt. Wie in der Diskussion innerhalb von Beitrag III dargelegt, lässt sich dieser Befund möglicherweise mit der self-determination theory (Ryan & Deci, 2000 2019) erklären. Demnach ist es der intrinsischen Motivation zuträglich, wenn die grundlegenden Bedürfnisse einer Person nach sozialer Eingebundenheit, Autonomie und Kompetenzerleben erfüllt sind. Somit erklärt die self-determination theory, weshalb speziell die der intrinsischen Motivation konzeptionell sehr ähnlichen intrinsischen Werte, nicht aber die im Vergleich stärker der extrinsischen Motivation ähnelnden Wichtigkeitswerte und Nützlichkeitswerte von dem Fähigkeitsselbstkonzept beeinflusst wurden. Möglicherweise war es das mit einem hohen Fähigkeitsselbstkonzept einhergehende Erleben eigener Kompetenz, welches die Schüler*innen motivierte, sich gerne und aus Spaß und Interesse mit mathematischen Aufgaben zu beschäftigen, während dasselbe Kompetenzerleben weniger Einfluss darauf hat, für wie wichtig und nützlich sie die Aufgaben halten. Der über den beschriebenen positiven indirekten Effekt hinaus bestehende negative direkte Effekt des Fähigkeitsselbstkonzepts auf die Noten, könnte demgegenüber auf die Annahme, mit weniger Anstrengung hohe Leistungen zu erzielen, zurückzuführen sein. Schüler*innen, die davon ausgehen, aufgrund ihrer hohen mathematischen Kompetenz im Fach Mathematik auch ohne bedeutende Anstrengung gute Noten erzielen zu können, könnten ihre Lernanstrengungen reduzieren und somit schlechtere Noten erzielen. Nach dieser Interpretation hätte ein hohes Fähigkeitsselbstkonzept also einerseits einen positiven Einfluss auf akademische Leistung aufgrund der intrinsisch motivierenden Wirkung hohen Kompetenzerlebens (positiver indirekter Pfad in Beitrag III) und andererseits einen negativen Einfluss aufgrund der damit einhergehenden Gefahr, dass Schüler*innen davon ausgehen, weniger Lernen zu müssen (negativer direkter Pfad in Beitrag III). Allerdings muss bedacht werden, dass 1) das Ausbleiben eines positiven Gesamteffekts des Fähigkeitsselbstkonzepts wie oben beschrieben untypisch für vergleichbare Studien ist und 2), dass das vollständige Modell auch für die intrinsischen Werte zwar Teil des Konfidenzsets war, aber wesentlich weniger Unterstützung erfuhr als das Modell positiver Effekte der Kompetenz. Somit sind die in diesem Absatz besprochenen Erklärungsansätze lediglich als aus den Ergebnissen ableitbare Hypothesen und Fragestellungen zu verstehen.

3.3.3 Fragestellung 1 – SE Bias Effekte

In keinem der drei empirischen Beiträge wurden unter Verwendung moderner statistischer Verfahren (CRA; Humberg et al., 2018; RSA; Edwards & Parry1993; Edwards, 2002; Humberg et al., 2019a) SE Bias Effekte gefunden, weder lineare Effekte (alle Beiträge) noch nonlineare Effekte (Beiträge II und III), weder direkte Effekte (alle Beiträge) noch indirekte Effekte vermittelt über Erfolgserwartungen und subjektive Werte (Beitrag III) und weder Effekte in Mathematik (alle Beiträge) noch in Deutsch (Beitrag I). In Beitrag I wurden neben einem moderneren Verfahren (CRA) auch zwei ältere und in der Literatur häufiger verwendete Verfahren zur Untersuchung von SE Bias Effekten eingesetzt: Algebraische Differenzwerte und Residualwerte. Dies geschah, um die empirische Bedeutung des Einsatzes modernerer Verfahren exemplarisch zu untersuchen und zu veranschaulichen. Die Analyse der algebraischen Differenzwerte ergab in Übereinstimmung mit der CRA keine signifikanten SE Bias Effekte, wohingegen die Analyse der Residualwerte in dem Befund signifikanter positiver SE Bias Effekte resultierte. Dies ist wie folgt zu erklären. Wie in Abschnitt 1.2.3.1 erläutert, fallen algebraische Differenzwerte, welche als die Differenz aus der Selbsteinschätzung und der Kompetenz berechnet werden, umso größer aus, je größer die Selbsteinschätzung ist und umso kleiner, je größer die Kompetenz ist. Dies hat zur Folge, dass SE Bias Effekte unter Verwendung algebraischer Differenzwerte mit Effekten der Selbsteinschätzung positiv konfundiert sind: Je größer der (positive) Effekt der Selbsteinschätzung, desto größer der (positive) Effekt des SE Bias. Ebenso führt dieselbe Tatsache dazu, dass SE Bias Effekte negativ mit Effekten der Kompetenz konfundiert sind: Je größer der (positive) Effekt der Kompetenz, desto kleiner der (positive) Effekt des SE Bias (Edwards & Parry, 1993; Edwards, 2002; Humberg et al., 2018). In Beitrag I war in Mathematik wie auch in Deutsch sowohl der Effekt der Selbsteinschätzung als auch der Effekt der Kompetenz positiv. Somit führten die positiven Selbsteinschätzungseffekte dazu, dass die SE Bias Effekte, berechnet über algebraische Differenzwerte, größer wurden, während die positiven Effekte der Kompetenz dazu führten, dass die SE Bias Effekte kleiner wurden. Mit anderen Worten: Die ähnlich großen positiven Effekte der Selbsteinschätzung und der Kompetenz glichen sich in dem Effekt ihres Differenzwerts aus, sodass kein signifikanter SE Bias Effekt bestand. Demgegenüber sind Residualwerte statistisch nur (und zwar in noch größerem Maß als algebraische Differenzwerte) mit der Selbsteinschätzung konfundiert, aber nicht mit der Kompetenz (Humberg et al., 2019a). Da in Beitrag I positive und signifikante Selbsteinschätzungseffekte vorlagen, erklärt dies, dass unter Verwendung von Residualwerten auch SE Bias Effekte gefunden wurden. Da die Verwendung von Residualwerten in der neueren Forschung populärer ist als die Verwendung von algebraischen Differenzwerten (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côté et al., 2014; Leduc & Bouffard, 2017; Lee, 2021; 2022; Praetorius et al., 2016; sie auch Abschnitt 1.2.3.2) und diese als statistisch bessere Alternative zu algebraischen Differenzwerten empfohlen werden (Connell & Illardi, 1987; John & Robins, 1994; Paulhus & John, 1998; Robins & John, 1997) ist dieser Befund zudem von exemplarischer Bedeutung. Er zeigt, wie in einer empirischen Studie zu SE Bias Effekten, die Verwendung von Residualwerten im Gegensatz zu moderneren Verfahren zum Auffinden artifizieller SE Bias Effekte führen kann. Da in keinem der drei Beiträge unter Einsatz moderner Verfahren SE Bias Effekte nachgewiesen werden konnten, scheinen diese zumindest für die untersuchten Populationen nicht zu bestehen. Zwar bleibt zu bedenken, dass in Beitrag III die Effekte des Fähigkeitsselbstkonzepts unerwartet ausgeblieben sind, was die Frage aufwirft, inwieweit die Befunde aus dieser Studie allein aussagekräftig sind. Doch zeigt die Einheitlichkeit der Befundlage über alle drei Beiträge mit unterschiedlichen methodischen Verfahren, untersuchten Domänen und untersuchten Effekten hinweg, dass die empirische Evidenz eindeutig gegen SE Bias Effekte auf akademische Leistung spricht. Für die akademische Leistung von

Schüler*innen scheinen also sowohl die Selbsteinschätzung ihrer Kompetenzen als auch ihre tatsächlichen Kompetenzen relavant zu sein, nicht aber die Diskrepanz zwischen den beiden. Ob Schüler*innen sich selbst überschätzen, unterschätzen, oder akkurat einschätzen, hat über die absoluten Effekte von Selbsteinschätzung und Kompetenz hinaus somit keine nachweisbare Bedeutung für die Schulnoten.

3.3.4 Fragestellungen 2a bis 2d – Mediation von SE Bias Effekten

Da in Beitrag III keine SE Bias Effekte nachgewiesen wurden, lag auch keine Mediation von SE Bias Effekten durch Erfolgserwartung oder subjektive Werte vor. Zwar stellt sich, wie im vorigen Abschnitt diskutiert, die Frage nach der Repräsentativität der Ergebnisse aus Beitrag III, doch kann, da SE Bias Effekte unter Betrachtung aller drei Beiträge insgesamt zurückgewiesen werden können, auch die Annahme einer Mediation selbiger Effekte verworfen werden.

3.3.5 Weitere Befunde – Effekte der Kompetenz, der Erfolgserwartung und der subjektiven Werte

Die empirischen Beiträge beinhalten einige diskussionswürdige Befunde, welche nicht unmittelbar zur Bewertung und Beantwortung der aufgestellten Hypothesen und Fragestellungen beitragen. Diese Befunde sind in diesem Abschnitt kurz zusammengefasst.

Zunächst ist festzuhalten, dass sich mit einer Ausnahme in allen drei Beiträgen positive Effekte der Kompetenz auf die Noten gezeigt haben. Die einzige Ausnahme ist der nicht signifikante Effekt der Deutschkompetenz auf die Deutschnote in Beitrag I, welcher aber, wie bereits erläutert, vermutlich auch der im Vergleich zur Mathematikkompetenz schwierigeren Erfassung der Deutschkompetenz geschuldet ist. Der Befund signifikanter positiver Kompetenzeffekte ist nicht überraschend, da eine höhere Kompetenz in einer bestimmten Domäne sich unmittelbar auf die schulische Leistung in selbiger Domäne, etwa in Form besserer Leistungen in Klassenarbeiten und in der mündlichen Beteiligung, auswirken sollte. Erbrachte Leistung gehört zu den zentralen Kriterien, welche Lehrkräfte bei der Notengebung berücksichtigen (Brookhart et al., 2016) und in Nordrhein-Westfalen, wo die empirischen Arbeiten durchgeführt wurden, gesetzlich berücksichtigen müssen (§48 Abs. 1-2, Schulgesetz NRW - SchulG). Bemerkenswert ist, dass die objektiv erfasste Kompetenz auch unter Kontrolle zahlreicher Kovariaten (Fähigkeitsselbstkonzept, vorige Noten, sozioökonomischer Status, Geschlecht, Erfolgserwartung und subjektive Werte) einen nachweisbaren Einfluss auf die Noten ausübte. Zudem konnte in Beitrag III gezeigt werden, dass der positive Effekt der Kompetenz partiell durch die Erfolgserwartung nicht aber durch die subjektiven Werte mediiert wird. Da für den Ausgangswert der Erfolgserwartung und andere Kovariaten kontrolliert wurde, deutet dies auf einen tatsächlichen Mediationseffekt hin. Demnach würde eine höhere Kompetenz dazu führen, dass Schüler*innen eher an ihren eigenen Erfolg glauben, was sich positiv auf ihren tatsächlichen Erfolg im Sinne besserer Noten auswirkt. Allerdings war der indirekte Effekt vermittelt über die Erfolgserwartung kleiner als der direkte Effekt.

Zweitens zeigten sich in Beitrag III Effekte sowohl der Erfolgserwartung als auch der subjektiven Werte auf die Noten. Dieser Befund stimmt mit den Annahmen des Erwartungs-Wert-Modells überein (Eccles & Wigfield; 2020; 2023; Wigfield, 1994; Wigfield et al., 2020) und stellt eine Replikation früherer Befunde dar (z.B., Brown & Putwain, 2022; Froiland & Davison, 2016; Geng et al., 2022; Steinmayr et al., 2019; Weidinger et al., 2020). Bemerkenswert ist, dass sich die positiven Effekte von Erfolgserwartung und subjektiven Werten selbst unter Kontrolle der objektiv erfassten Kompetenz und des Fähigkeitsselbstkonzepts in derselben Domäne zeigten. Zwar haben auch andere Studien zur Bedeutung subjektiver Werte für akademische Leistung objektiv erfasste kognitive Fähigkeiten (Weidinger et al., 2020) oder objektiv erfasste Fähigkeiten und das Fähigkeitsselbstkonzept (Steinmayr et al., 2019) berücksichtigt. Allerdings wurden in diesen Studien Tests zur Erfassung der allgemeinen, numerischen, verbalen und/oder figuralen Intelligenz statt Schulleistungstests eingesetzt. Somit ist die Interpretation der Befunde dieser Studien eine etwas andere. Während frühere Studien gezeigt haben, dass subjektive Werte auch über allgemeine und inhaltsspezifische (z.B. numerische) kognitive Fähigkeiten hinaus akademische Leistung vorhersagen, zeigt die vorliegende Arbeit, dass dies auch unter Kontrolle der domänenspezifischen Kompetenz der Fall ist. Dies ist insbesondere deshalb der Fall, da in Beitrag III mit dem DEMAT 9 ein curricular valider Mathematikkompetenztest eingesetzt wurde. Zudem ist mir keine Studie bekannt, in der der Einfluss der Erfolgserwartung auf akademische Leistung über den Einfluss des Fähigkeitsselbstkonzepts hinaus untersucht wurde. Somit ist die vorliegende Arbeit die erste, in der ein entsprechender Effekt untersucht und nachgewiesen wurde.

3.3 Stärken und Limitationen der vorliegenden Arbeit

Der zentrale Vorteil der vorliegenden Arbeit ist, dass ein in der pädagogischen Psychologie seit längerem diskutiertes Thema erstmalig mit geeigneten statistischen Methoden untersucht wurde. Auf diese Weise konnten SE Bias Effekte auf akademische Leistung untersucht werden, ohne dass diese mit Effekten der Selbsteinschätzung und/oder Effekten der Kompetenz konfundiert wurden.

Eine weitere Stärke besteht in der engen inhaltlichen Passung zwischen den eingesetzten Verfahren zur Erfassung der Selbsteinschätzung und der Kompetenz in allen drei Beiträgen. Es ist eine wichtige Voraussetzung der SE Bias Forschung, dass die Selbsteinschätzung und das Außenkriterium inhaltlich möglichst eng verwandt sind (Edwards, 2002; Humberg et al., 2019b; Shanock et al., 2010). Ist dies nicht der Fall, könnte jede beobachtete Diskrepanz zwischen Selbsteinschätzung und Außenkriterium statt auf eine tatsächliche Selbstüber- oder unterschätzung auch darauf zurückzuführen sein, dass mit dem Test für das Außenkriterium eine Kompetenz erfasst wurde, die sich von der Kompetenz, welche die Proband*innen eingeschätzt haben, unterscheidet. Dies könnte zum Beispiel der Fall sein, wenn ein

Mathematikkompetenztest eine bestimmte Komponente der Mathematikkompetenz nicht berücksichtigt (beispielsweise das Lösen linearer Gleichungen), welche aber für die Schüler*innen aktuell von hoher Relevanz ist, etwa da sie das entsprechende Thema kürzlich im Unterricht behandelt haben. Stützen Schüler*innen sich bei ihrer Selbsteinschätzung dann unter anderem auf ihre wahrgenommene Kompetenz im Lösen linearer Gleichungen, ist die Selbsteinschätzung nicht mehr ohne weiteres mit dem Außenkriterium vergleichbar. Durch den Einsatz breit gefasster Kompetenztests, welche zahlreiche verschiedene Subkomponenten der Mathematikkompetenz abdecken (TIMSS, Baumert et al., 1998; DEMAT 9; Schmidt et al., 2013), wurde die Vergleichbarkeit mit dem Fähigkeitsselbstkonzept in Mathematik sichergestellt. In Deutsch wurden aus diesem Grund zwei Tests zur Erfassung der Kompetenz eingesetzt, ein Lesetest und ein Rechtschreibtest. Allerdings muss bedacht werden, dass die Deutschkompetenz nach dem Verständnis von Zehntklässler*innen vermutlich mehr beinhaltet als diese beiden Kompetenzen. Zwar sind Leseverständnis und Orthografie im Kernlehrplan für die Sekundarstufe I an Gymnasien in NRW als Kompetenzen vertreten, allerdings auch weitere Kompetenzen, wie das Verfassen von Texten, das Abgeben eigener Urteile und ein differenzierter Wortschatz (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, 2019, S. 16-17). Eine weitgehend vollständige Erfassung der Deutschkompetenz unter Berücksichtigung der genannten und weiterer Subkomponenten war aufgrund des Fehlens derart umfangreicher Tests und zeitlicher Beschränkungen bei der Untersuchungsdurchführung nicht möglich. Somit ist als Limitation zu nennen, dass in Beitrag I die Vergleichbarkeit zwischen Selbsteinschätzung und Außenkriterium in Deutsch vermutlich geringer ausfällt als in Mathematik. Aus diesem Grund sowie aufgrund statistischer Probleme bei der Erfassung der Deutschkompetenz (zum Beispiel große Standardfehler, siehe Beitrag I, sowie schlechtere Modelfits im Vergleich zu Mathematik) wurden die Untersuchungen in den Beiträgen II und III auf die Domäne Mathematik beschränkt.

Dass die Befunde aus Beitrag I trotz dieser Limitation in beiden Domänen ähnlich ausfielen und auch das Modell in Deutsch über gute Fit-Indizes verfügte, spricht allerdings dafür, dass die Ergebnisse in Deutsch in Beitrag I nicht in zu großem Ausmaß von dieser Einschränkung beeinträchtigt sind.

Zuletzt sei außerdem auf die Bedeutung der informationstheoretischen Modellselektion auf Basis von Akaike-Gewichten nach dem Schema von Humberg et al. (2019a) in den Beiträgen II und III verwiesen. Wie bereits erläutert, liegen in der Literatur zahlreiche verschiedene Annahmen über lineare und nonlineare SE Bias Effekte auf akademische Leistung vor, von denen einige ineinander genestete Zusammenhänge repräsentieren, andere hingegen nicht. Beispielsweise stellt die Annahme eines umgekehrt u-förmigen Zusammenhangs mathematisch einen Spezialfall der optimal margin Hypothese mit einer optimalen Marge von null dar und ist daher in dieser genestet (Humberg et al., 2019a). Zwar wurden nonlineare SE Bias Effekte auf akademische Leistung bereits in anderen Arbeiten empirisch untersucht (Lee, 2021; Lopez et al., 1998; Praetorius et al., 2016; Wright, 2000), allerdings nicht unmittelbar gegen konkurrierende Hypothesen getestet. Selbst wenn beispielweise ein signifikanter optimal margin Effekt gefunden würde, bedeutet dies nicht zwangsläufig, dass er einen besseren Erklärungswert für die zugrundeliegenden Zusammenhänge bietet als einfachere lineare SE Bias Effekte oder lineare Haupteffekte der Selbsteinschätzung und der Kompetenz. Humberg et al. (2019a, S. 840) gehen daher davon aus, dass das isolierte Testen einzelner miteinander konkurrierender Hypothesen einen der Gründe für widersprüchliche Befunde in der SE Bias Forschung darstellt. Liegen konkurrierende Hypothesen vor, ist es daher der bessere Ansatz, diese unmittelbar gegeneinander zu testen, statt sie, wie in den meisten vorigen Arbeiten, jeweils einzeln auf Signifikanz zu prüfen (Humberg et al., 2022). Informationstheoretische Modellvergleiche anhand von Akaike-Gewichten sind dazu gut geeignet, da sie erstens einen Vergleich sowohl genesteter wie nicht genesteter Modelle erlauben und zweitens, eine leicht zu interpretierende quantitative Abschätzung der Sicherheit bei der Modellselektion ermöglichen (die prozentuale Wahrscheinlichkeit eines Modells das beste Modell gegeben der Daten und konkurrierenden Modelle darzustellen). Die Verwendung dieses Ansatzes in den Beiträgen II und III stellt somit sicher, dass die gefundenen Effekte (beispielsweise positive Haupteffekte von Fähigkeitsselbstkonzept und Kompetenz in Beitrag II), nicht nur signifikant sind, sondern auch, dass sie im Vergleich zu allen anderen untersuchten Effekten die beste Erklärung des zugrundeliegenden Zusammenhangsmusters der Daten darstellen.

Neben der bereits genannten Limitation in Bezug auf die Erfassung der Deutschkompetenz in Studie I weist die vorliegende Arbeit noch andere nennenswerte Limitationen auf. Die nach meiner Ansicht bedeutendste Einschränkung ist die Tatsache, dass die Selbsteinschätzungen und die Außenkriterien in allen Beiträgen sehr ähnlich operationalisiert wurden. In allen Fällen wurde das Fähigkeitsselbstkonzept erfasst und in Bezug gesetzt zu einem möglichst umfassenden Kompetenztest in derselben Domäne. Wie bereits argumentiert, ist dieses Vorgehen gut geeignet, um SE Bias Effekte zu untersuchen. Allerdings stellt es nicht die einzige Möglichkeit dar. Ebenso wäre es beispielsweise denkbar, statt dem relativ allgemeinen Fähigkeitsselbstkonzept, die spezifische Überzeugung, bestimmte Aufgaben oder Aufgabengruppen lösen zu können, zu erfassen (aufgabenspezifische Selbstwirksamkeit), oder die Überzeugung, in bestimmten Domänen gute Leistungen erzielen zu können (domänenspezifische Selbstwirksamkeit; z.B. Jungert et al., 2014; Lu et al., 2023; Siefer et al., 2021). Diese Selbsteinschätzungen könnten dann mit Leistungen in den jeweiligen Aufgaben beziehungsweise Domänen als Außenkriterien in Bezug gesetzt werden. Eine solche Operationalisierung könnte besser als jene in der vorliegenden Arbeit geeignet sein, um umgekehrt u-förmige SE Bias Effekte, wie sie von Theorien des selbstregulierten Lernens impliziert werden, aufzudecken. In Studien zum Einfluss akkurater Selbsteinschätzungen auf selbstreguliertes Lernen wird meist erfasst, wie sicher sich die Proband*innen sind, bestimmte Aufgaben oder Aufgabengruppen lösen zu können, oder diese gelöst zu haben (z.B. Dunlosky & Rawson, 2012; Hadwin & Webster, 2013; Händel et al., 2020; Pijeira-Díaz et al., 2023; van Loon & Oeri, 2023). Es ist denkbar, dass der Bias in derartigen aufgabenspezifischen Selbsteinschätzungen während eines selbstregulierten Lernprozesses sich von dem Bias in relativ allgemeineren domänenspezifischen Kompetenzen, wie in der vorliegenden Arbeit erfasst, unterscheidet. So könnte beispielsweise eine Schülerin korrekt erkennen, dass sie eine bestimmte Mathematikaufgabe nicht lösen kann (aufgabenspezifischer SE Bias nahe null), gleichzeitig aber ihre eigene allgemeine Mathematikkompetenz überschätzen, etwa falls sie fälschlich davon ausgeht, dass ihre mangelnde Fähigkeit, die Aufgabe zu lösen, nicht auf eine geringe Mathematikkompetenz zurückzuführen ist, sondern auf andere Faktoren wie eine unfaire oder zu schwere Aufgabenstellung oder mangelnde Vorbereitung (positiver SE Bias für die Mathematikkompetenz im Allgemeinen). Empirisch zeigt sich in der Tat, dass diese beiden Formen des SE Bias zwar positiv zusammenhängen, aber trennbar sind (Lee, 2022). Somit kann aufgrund der Befunde der vorliegenden Arbeit nicht ausgeschlossen werden, dass ein SE Bias im Fähigkeitsselbstkonzept zwar keinen Einfluss auf akademische Leistung ausübt, ein Bias in der Fähigkeit, in konkreten Lernkontexten akkurat einzuschätzen, welche Aufgaben wie sicher gelöst werden können, hingegen schon. Denkbar wäre etwa, dass es am förderlichsten ist, eigene Stärken und Defizite in der Bearbeitung spezifischer Aufgaben zu erkennen (umgekehrt u-förmiger aufgabenspezifischer SE Bias Effekt wie in der Forschung zum selbstregulierten Lernen angenommen) und gleichzeitig über ein hohes Fähigkeitsselbstkonzept zu verfügen. So könnten Schüler*innen von beidem profitieren: Realistisch und akkurat gesteuerten Lernprozessen und dem motivierenden Einfluss eines hohen Fähigkeitsselbstkonzepts. Empirische Ergebnisse von

Lee (2022) weisen erstmals auf derartige differentielle Effekte hin. In dieser Studie zeigte sich, dass hochleistende Schüler*innen ihre allgemeinen mathematischen Kompetenzen überschätzten, aber ihre Leistung in spezifischen Mathematikaufgaben akkurat einschätzten. Zudem zeigten sich differentielle Effekte der beiden Formen des SE Bias auf Lernverhalten. Da diese Studie allerdings methodisch dem Zwei-Schritte-Ansatz folgte, sind die Ergebnisse mit Vorsicht zu interpretieren, da die gefundenen Zusammenhänge mit den beiden SE Bias Formen, insbesondere der positive Zusammenhang des SE Bias in allgemeinen mathematischen Fähigkeiten mit der Leistung, Artefakte von Selbsteinschätzungseffekten darstellen können. Dennoch zeigen die Befunde, dass SE Bias Effekte von der Operationalisierung der Selbsteinschätzung und des Außenkriteriums abhängen können. Zusammenfassend bedeutet dies, dass die Ergebnisse der vorliegenden Arbeit Rückschlüsse auf die Effekte eines SE Bias in domänenspezifischen Fähigkeitsselbstkonzepten erlauben, nicht aber auf die Bedeutung der akkuraten Einschätzung im Hinblick auf die eigene Fähigkeit, in konkreten Lernkontexten bestimmte Aufgaben zu lösen. Zukünftige Studien sollten sich daher der Frage widmen, ob sich differentielle Effekte derart unterschiedlicher Formen des SE Bias, wie sie bei Lee (2022) nachgewiesen wurden, auch dann zeigen, wenn SE Bias Effekte statistisch von Selbsteinschätzungseffekten getrennt werden.

Desweiteren stellt sich die Frage, inwieweit die Ergebnisse der vorliegenden Arbeit auf andere Populationen generalisierbar sind. Da in den empirischen Beiträgen Gymnasiast*innen der 9. bis 11. Klassenstufe in Deutschland untersucht wurden, sind die Stichproben innerhalb der Beiträge wie auch zwischen den Beiträgen sehr homogen. Somit kann nicht ausgeschlossen werden, dass SE Bias Effekte in anderen Populationen, zum Beispiel bei jüngeren oder älteren Schüler*innen oder Studierenden, an anderen Schulformen oder in anderen Nationen und Kulturen existieren. Allerdings zeigt sich zumindest für Selbsteinschätzungseffekte, dass diese weitgehend unabhängig von den genannten Faktoren sind. So zeigen sich etwa Effekte des

Fähigkeitsselbstkonzepts auf akademische Leistung in zahlreichen Ländern und Kulturen, darunter Australien (z.B. Seaton et al., 2014; 2015), Deutschland (z.B. Weidinger et al., 2018; Preckel et al., 2017), Taiwan (Lee, & Kung, 2018; Chen et al., 2013), Griechenland (Kalogiannis et al., 2011) und Hong Kong (Marsh et al., 2002) sowie an regulären Schulen und Klassen als auch an Schulen und Klassen für Hochbegabte (Seaton et al., 2015; Preckel et al., 2017). Zwar kann von einer Stabilität von Selbsteinschätzungseffekten zwischen verschiedenen Nationen und Fähigkeitsniveaus nicht unmittelbar auf eine Stabilität von SE Bias Effekten geschlossen werden, da aber die Selbsteinschätzung eine zentrale Komponente des SE Bias darstellt, erscheint es plausibel, dass SE Bias Effekte in Abhängigkeit der genannten Faktoren ebenfalls zumindest nicht stark variieren sollten. Empirisch lässt sich diese Frage auf Basis der aktuellen Literatur nicht beantworten, da nach meinem Wissen keine spezifischen Studien zum Vergleich von SE Bias Effekten in verschiedenen Populationen vorliegen und einzelne SE Bias Studien aufgrund ihrer heterogenen und zumeist problematischen Methodik nur eingeschränkt vergleichbar und interpretierbar sind. In Hinblick auf Altersunterschiede zeigt sich, dass die Effekte des Fähigkeitsselbstkonzepts auf akademische Leistung mit zunehmender Klassenstufe ansteigen. Im Grundschulalter ist die Befundlage heterogen, etwa ab einem Alter von 10 Jahren sind entsprechende Effekte aber gut nachweisbar (z.B. Ehm et al., 2019; Guay et al., 2003; Marsh et al., 2007; Niepel et al., 2014; Skaalvik & Valas, 1999; Viljaranta et al., 2014; Weidinger et al., 2018; Wu et al., 2021). Somit sind die untersuchten Stichproben vom Alter her geeignet, um Effekte eines Bias in Fähigkeitsselbstkonzepten zu untersuchen. Dennoch sollten die Befunde der vorliegenden Arbeit aufgrund der angesprochenen Stichprobenhomogenität nur unter Vorbehalt auf andere Populationen generalisiert werden. Ob dieselben Effekte in anderen Ländern, Schulformen und Klassenstufen tatsächlich bestehen beziehungsweise nicht bestehen, kann letztlich nur durch zukünftige empirische Arbeiten beantwortet werden.

In den empirischen Beiträgen wurde nicht für die Klassen- beziehungsweise Kurszugehörigkeit (im folgenden kurz Klassenzugehörigkeit) der Schüler*innen kontrolliert. Da somit die geclusterte Datenstruktur nicht berücksichtigt wurde, können Effekte zwischen Schüler*innen innerhalb von Klassen und Effekte zwischen Klassen nicht auseinandergehalten werden. So kann beispielsweise zumindest nicht rein statistisch bestimmt werden, ob ein höheres Fähigkeitsselbstkonzept bei einem Individuum zu einer besseren Note führt, oder ob ein höheres mittleres Fähigkeitsselbstkonzept einer Klasse zu im Mittel besseren Noten führt. Allerdings erlaubte es die Fragestellung nach SE Bias Effekten nicht, für die Klassenzugehörigkeit zu kontrollieren, da Unterschiede zwischen Klassen relevante Quellen eines SE Bias darstellen können. Beispielsweise führt ein höheres mittleres Leistungsniveau in einer Klasse zu einem geringeren Fähigkeitsselbstkonzept der einzelnen Schüler*innen (Fischteicheffekt; z.B. Fleischmann et al., 2022; Jansen et al., 2022). Somit kann die Zugehörigkeit zu einer besonders leistungsstarken beziehungsweise zu einer besonders leistungsschwachen Klasse dazu führen, dass Schüler*innen ihre Kompetenzen unter- beziehungsweise überschätzen. Das Fähigkeitsselbstkonzept und/oder die Noten innerhalb von Klassen zu standardisieren oder die Klassenzugehörigkeit anderweitig zu kontrollieren, würde diese realen Selbstüber- und unterschätzungen maskieren und die entsprechenden empirischen Effekte verzerren. Dies sei kurz an einem konkreten Beispiel erläutert. Angenommen eine Schülerin befindet sich in einer hoch leistungsstarken Klasse, in der die meisten Schüler*innen gute bis sehr gute Mathematiknoten erzielen. Weiterhin angenommen, diese Schülerin erzielt selbst ebenfalls gute Noten in Mathematik. So hätte sie aufgrund des Fischteicheffekts unter Umständen nur ein durchschnittliches (oder zumindest in Relation zu ihren guten Noten niedrigeres) mathematisches Fähigkeitsselbstkonzept, würde sich also unterschätzen. Würde man nun für die Klassenzugehörigkeit kontrollieren, würde diese Selbstunterschätzung maskiert, da die Schülerin durchaus über eine führ ihre Klasse durchschnittliche Mathematikkompetenz und ein für ihre Klasse durchschnittliches mathematisches Fähigkeitsselbstkonzept verfügen könnte (da auch die anderen Schüler*innen im Mittel gleichermaßen vom Fischteicheffekt betroffen sind). Mit anderen Worten, Selbstüber- und unterschätzungen kommen auch aufgrund der Klassenzugehörigkeit zustande und diese zu kontrollieren verzerrt somit den SE Bias. Dies würde sich auch auf die Bestimmung der SE Bias Effekte auswirken, da mögliche Assoziationen derartiger aufgrund der Klassenzugehörigkeit zustandegekommener Fehleinschätzungen mit anderen Variablen nicht mehr bestimmt werden können. Dass nicht für die Klassenzugehörigkeit kontrolliert wurde, stellt somit eine Einschränkung dar, die allerdings in diesem Forschungskontext nicht ohne weiteres behoben werden kann. In der Tat ist mir keine Studie zu SE Bias Effekten bekannt, in der in den individuellen Werten für die Klassenzugehörigkeit kontrolliert wurde (z.B. Bonneville-Roussy et al., 2017; Chung et al., 2016; Côté et al., 2014; Dupeyrat et al., 2011; Gonida & Leondari, 2011; Humberg et al., 2018; 2019a; Leduc & Bouffard 2017; Lee, 2021; 2022; Robins & Beer, 2001; Wright, 2000).

Obwohl das Fähigkeitsselbstkonzept und die Kompetenz in allen Beiträgen mit mehreren Items erfasst wurden, wurden sie lediglich in Beitrag I als latente Variablen modelliert. In den Beiträgen II und III wurden alle Variablen als manifeste Variablen behandelt. Die Modellierung latenter Variablen hat statistische Vorteile wie eine größere Robustheit gegenüber Verletzung statistischer Annahmen, etwa der Unabhängigkeit der Fehlerterme (Little, 2013; Seo et al., 2015). Die in Beitrag I verwendete condition-based regression analysis stellt in ihrer ursprünglichen Form ebenfalls ein manifestes Verfahren dar (Humberg et al., 2018), wurde in Beitrag I aber zur Modellierung latenter Variablen adaptiert. Hingegen konnte die in den Beiträgen II und III verwendete response surface analysis (RSA; Edwards, 2002; Edwards & Parry, 1993; Humberg et al., 2019a) nicht ohne weiteres als latentes Verfahren adaptiert werden. Dennoch erschien ihr

Einsatz aufgrund anderer Vorteile (Testung nonlinearer Effekte, direkter Vergleich konkurrierender Hypothesen) sinnvoll. Wie in Abschnitt 1.2.3.2 dargestellt, wird die RSA in anderen Disziplinen seit längerem und zunehmend auch in der pädagogischen und Persönlichkeitspsychologie eingesetzt und dabei auch genutzt, um latente Daten zu untersuchen, obwohl diese nicht als solche modelliert werden können (z.B. Bang & Park, 2015; Compagnoni & Losenno, 2020; Dust & Tims, 2020; Gilbreath et al., 2010; Jing et al., 2021; Meyer et al., 2010). Auch wenn die fehlende latente Modellierung einen Nachteil darstellt, ist die RSA (neben verwandten Verfahren wie spline regressions) aktuell das beste zur Verfügung stehende Verfahren zur Untersuchung von konkurrierenden Hypothesen über lineare und nonlineare SE Bias Effekte und (In-)kongruenzeffekten im Allgemeinen (Edwards & Parry, 2018; Humberg et al., 2019a; 2022)

Zuletzt muss darauf hingewiesen werden, dass alle Beiträge ein längsschnittliches und kein experimentelles Design aufweisen. Somit ist eine kausale Interpretation der Befunde als „Effekte“ von Fähigkeitsselbstkonzept, Kompetenz und SE Bias nur eingeschränkt möglich. Diese Limitation betrifft nicht allein die vorliegende Arbeit sondern den überwiegenden Großteil empirischer Studien zu SE Bias Effekten und Selbsteinschätzungseffekten. Zu den zentralen Maßnahmen, um Kausalschlüsse aus Längsschnittstudien ziehen zu können, gehört die theoretische, nicht statistische, Herleitung der Hypothesen und Forschungsfragen (Rohrer & Murayama, 2023) sowie die Kontrolle von Ausgangswerten und Störvariablen (McNamee, 2005). Beidem wurde in der vorliegenden Arbeit Rechnung getragen, da die überprüften SE Bias Effekte auf Basis theoretischer Annahmen in der Literatur abgeleitet wurden und in allen Beiträgen relevante potentielle Störvariablen, wie der Ausgangswert in den Noten, das Geschlecht und der sozioökonomische Status kontrolliert wurden. Dennoch kann nicht ausgeschlossen werden, dass weitere relevante nicht erfasste Störvariablen vorlagen, welche die kausale Interpretierbarkeit der Ergebnisse einschränken.

3.4 Implikationen für Forschung und Praxis

3.4.1 Forschungsimplikationen

Implikationen für weitere Forschungsarbeiten lassen sich aus den im vorigen Abschnitt diskutierten Stärken und Limitationen der vorliegenden Arbeit ableiten. So verdeutlichen die Ergebnisse insbesondere aus Beitrag I, dass SE Bias Effekte in zukünftigen Studien ausschließlich mittels des Ein-Schritt-Ansatzes untersucht werden sollten, da der Zwei-Schritte-Ansatz zum Auffinden artifizieller SE Bias Effekte führen kann. Darüberhinaus ergeben sich aus meiner Sicht vor allem zwei Ansätze für weiterführende Forschung.

Erstens sollte der Fokus, welcher in der vorliegenden Arbeit auf dem SE Bias relativ allgemeiner domänenspezifischer Kompetenzen lag, stärker auf spezifische Lernsituationen gerichtet werden. Beispielsweise könnte die Frage untersucht werden, welche Auswirkungen es auf das konkrete Lernverhalten und Lernergebnisse hat, wenn Schüler*innen in Lernsituationen über- oder unterschätzen, wie gut sie die jeweiligen Inhalte bereits gelernt haben. Vergleichbare Studien gibt es bereits (z.B. Cogliano et al., 2021; Dunlosky & Rawson, 2012; Hadwin & Webster, 2013; Lee, 2022; van Loon & Oeri, 2023) und der Konsens in der Forschung ist, dass akkurate Selbsteinschätzungen am förderlichsten für selbstreguliertes Lernen sind (siehe auch Hacker & Bol, 2019). Allerdings weisen diese Studien neben Vorteilen gegenüber der vorliegenden Arbeit auch bedeutsame Nachteile auf. Der Vorteil dieser Studien ist, dass die Akkuratesse beziehungsweise der Bias in den Selbsteinschätzungen meist auf einer sehr spezifischen Ebene erfasst wird. Beispielsweise lernten die Proband*innen bei Dunlosky und Rawson (2012) verschiedene Definitionen aus einem Psychologielehrbuch auswendig und sollten diese Definitionen während des Lernprozesses wiederholt wiedergeben. Dabei wurde abgefragt, wie gut sie mit der jeweiligen Antwort ihrer Ansicht nach in einem Test abschneiden würden. Die Akkuratesse dieser Einschätzung wurde dann eingesetzt um die spätere Leistung in einem tatsächlichen Test vorherzusagen. Somit erlaubt diese Methodik, im Gegensatz zu jener der vorliegenden Arbeit, Aussagen darüber, ob ein Bias in der Einschätzung des eigenen Lernerfolgs in konkreten Lernsituationen sich auf den tatsächlichen Lernerfolg auswirkt. Allerdings weist die Studie von Dunlsoky und Rawson (2012), wie die weitaus meisten zu diesem Thema (sie z.B. die oben genannten) das Problem auf, dass die Akkuratesse nach dem Zwei-Schritte-Ansatz bestimmt wurde. Somit sollte etwa unter Einsatz der RSA untersucht werden, ob entsprechende Effekte auch unabhängig von reinen Effekten der Selbsteinschätzung und der Kompetenz an sich bestehen. Eine solche Studie wäre in der Lage genau aufzuzeigen, bei welchen Kombinationen von selbsteingeschätztem und tatsächlichem Lernerfolg die besten späteren Testresultate erzielt werden. Auch Reanalysen bereits durchgeführter Studien sind zu dieser Fragestellung denkbar, da die notwendigen Daten in den entsprechenden Studien bereits vorliegen und lediglich unter Einsatz einer anderen statistischen Methode untersucht werden müssten. Interessant wäre zudem, in derselben Studie neben dem selbsteingeschätzten Lernerfolg auch, wie in der vorliegenden Arbeit, die selbsteingeschätzte domänenspezifische Kompetenz zu erfassen. Auf diese Weise könnte verglichen werden, ob es differentielle Effekte eines domänenspezifischen SE Bias und eines aufgaben- beziehungsweise lernkontextspezifischen SE Bias auf akademische Leistung gibt. Dass solche differentiellen Einflüsse bestehen könnten, wurde von Lee (2022) gezeigt, allerdings ebenfalls unter Einsatz des Zwei-Schritt-Ansatzes.

Zweitens sollten die untersuchten Selbsteinschätzungseffekte und SE Bias Effekte im domänenspezifischen Fähigkeitsselbstkonzept auch in anderen Populationen und für andere Domänen untersucht werden, um die Generalisierbarkeit der gefundenen Ergebnisse zu überprüfen. Mit Hinblick auf Beitrag III wär es von besonderem Interesse, eine vergleichbare

Studie mit kürzeren zeitlichen Intervallen und in anderen Klassenstufen (speziell vor der Wahl von Grund- und Leistungskursen) durchzuführen.

3.4.2 Implikationen für die pädagogische Praxis

Das primäre Ziel der vorliegenden Arbeit ist die Beantwortung einer theoretischen Fragestellung der pädagogischen Psychologie. Somit ist die vorliegende Arbeit eher grundlagenwissenschaftlich als praxisorientiert. Implikationen für die pädagogische Praxis sollten daher nur mit Vorsicht abgeleitet und eher als Hypothesen für zukünftige stärker praxisorientierte Forschung und weniger als ausgereifte Handlungsempfehlungen verstanden werden. Allgemein werden hohe Selbstkonzepte in der Psychologie als wünschenswert und im akademischen wie anderen Kontexten förderlich angesehen (siehe Trautwein & Möller, 2016). Die vorliegende Arbeit stützt diese Annahme, da sie zeigt, dass selbst im Falle einer Überschätzung eigener Kompetenzen keine negativen Konsequenzen eines zu hohen Fähigkeitsselbstkonzepts zu befürchten sind. Dies bestärkt auch die förderliche Wirkung von Selbstkonzeptinterventionen, welche in der pädagogischen Psychologie vielfach untersucht wurden. O'Mara et al. (2006) identifizierten in einer Metaanalyse 145 Studien zu insgesamt 200 Interventionen und deren Effekten auf das Selbstkonzept. Selbstkonzeptinterventionen stellen zudem einen interessanten Ansatz zu möglichen experimentellen Studien zu SE Bias Effekten dar. Wenn das Fähigkeitsselbstkonzept durch Selbstkonzeptinterventionen und die Kompetenz durch Kompetenztrainings gesteigert wird, können deren Effekte und somit auch die Effekte des SE Bias im Ein-Schritt-Ansatz experimentell überprüft werden. Dadurch könnte zudem der praktisch relevanten Frage nachgegangen werden, ob alle Schüler*innen in gleichem Maße von Selbstkonzeptinterventionen profitieren, oder ob beispielsweise Schüler*innen, die ihre eigenen Kompetenzen unterschätzen, stärker von einer Intervention profitieren als solche, die ihre Kompetenzen akkurat einschätzen oder überschätzen. In Anbetracht des Ergebnisses der vorliegenden Arbeit, dass nur das Fähigkeitsselbstkonzept, nicht aber der SE Bias, einen Einfluss auf akademische Leistung hat, kann die Hypothese abgeleitet werden, dass alle Schüler*innen gleichermaßen von der Intervention profitieren sollten, zumindest falls diese das Fähigkeitsselbstkonzept aller Schüler*innen in gleichem Maße zu steigern vermag.

3.5 Fazit

Über drei empirische Beiträge hinweg konnten die in der Literatur vertretenen Hypothesen zu SE Bias Effekten auf akademische Leistung zurückgewiesen werden. Schüler*innen profitieren bei ihren Noten sowohl von einem hohen Fähigkeitsselbstkonzept als auch von einer hohen Kompetenz. Diese Effekte sind linear und voneinander unabhängig, wobei der Effekt der Kompetenz partiell über die Erfolgserwartung, nicht aber über subjektive Werte vermittelt wurde. Die Diskrepanz zwischen Selbsteinschätzung und Kompetenz hat hingegen keinen nachweisbaren Einfluss. Diese Befunde gelten allerdings nur für SE Bias Effekte im Fähigkeitsselbstkonzept, also in der relativ allgemeinen Einschätzung der eigenen Kompetenz in einer bestimmten Domäne. Ob sie sich auch auf andere beispielsweise aufgabenspezifische SE Bias Effekte übertragen lassen, muss in weiteren Studien untersucht werden.

Literaturverzeichnis II

Albert, S. (1977). Temporal comparison theory. *Psychological Review, 84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Arens, A. K., & Niepel, C. (2023). Formation of academic self-concept and intrinsic value within and across three domains: Extending the reciprocal internal/external frame of reference model. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.1257

Asendorpf, J. B., & van Aken, M. A. G. (2003). Personality-relationship transaction in adolescence: Core versus surface personality characteristics. *Journal of Personality, 71* , 629–662. https://doi.org/10.1111/1467-6494.7104005

Bakadorova, O., & Raufelder, D. (2020). The relationship of school self-concept, goal orientations and achievement during adolescence. *Self and Identity, 19*(2), 235–249. https://doi.org/10.1080/15298868.2019.1581082

Bang, H., & Park, J. G. (2015). The double-edged sword of task conflict: Its impact on team performance. *Social Behavior and Personality: An International Journal, 43*(5), 715–728. https://doi.org/10.2224/sbp.2015.43.5.715

Baumeister, R. F. (1989). The optimal margin of illusion. *Journal of Social and Clinical Psychology, 8*(2), 176-189. https://doi.org/10.1521/jscp.1989.8.2.176

Baumert, J., Lehmann, R., Lehrke, M., Clausen, M., Hosenfeld, I., Neubrand, J. et al. (1998). *Testaufgaben Mathematik TIMSS 7./8. Klasse (Population 2).* Max-Planck-Institut für Bildungsforschung. Retrieved from https://pure.mpg.de/rest/items/item_2103201/component/file_2103200/content

Bonneville-Roussy, A., Bouffard, T., & Vezeau, C. (2017). Trajectories of self-evaluation bias in primary and secondary school: Parental antecedents and academic consequences. *Journal of School Psychology, 63*, 1–12. https://doi.org/10.1016/j.jsp.2017.02.002

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. doi:10.1016/j.lindif.2012.07.02

Chung, J., Schriber, R. A., & Robins, R. W. (2016). Positive illusions in the academic context: A longitudinal study of academic self-enhancement in college. *Personality and Social Psychology Bulletin, 42*(10), 1384-1401. https://doi.org/10.1177/0146167216662866

Compagnoni, M., & Losenno, K. M. (2020). “I’m the best! Or am I?”: Academic self-concepts and self-regulation in kindergarten. *Frontline Learning Research, 8*(2), 131–152. https://doi.org/10.14786/flr.v8i2.605

Connell, J. P., & Illardi, B. C. (1987). Self-system concomitants of discrepancies between children’s and teachers’ evaluations of academic competence. *Child Development, 58*, 1297–1307. https://doi.org/10.2307/1130622

Côté, S., Bouffard, T., & Vezeau, C. (2014). The mediating effect of self-evaluation bias of competence on the relationship between parental emotional support and children’s academic functioning. *British Journal of Educational Psychology, 84*(3), 415–434. https://doi.org10.1111/bjep.12045

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students’ learning and retention. *Learning and Instruction, 22*(4), 271–280. https//doi.org10.1016/j.learninstruc.2011.08.003

Dust, S. B., & Tims, M. (2020). Job crafting via decreasing hindrance demands: The motivating role of interdependence misfit and the facilitating role of autonomy. *Applied Psychology: An International Review, 69*(3), 881–912. https://doi.org/10.1111/apps.12212

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation.

Contemporary Educational Psychology, 61. https://doi.org/10.1016/j.cedpsych.2020.101859.

Eccles, J. S., & Wigfield, A. (2023). Expectancy-value theory to situated expectancy-value theory: Reflections on the legacy of 40+ years of working together. *Motivation Science*, *9*(1), 1–12. https://doi.org/10.1037/mot0000275

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. W. Schmitt (Hrsg.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (S. 350–400). Jossey-Bass.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36,* 1577–1613. http://dx.doi.org/10.2307/256822

Edwards, J. R., & Parry, M. E. (2018). On the Use of Spline Regression in the Study of Congruence in Organizational Research. *Organizational Research Methods, 21*, 68-110. https://do.org/10.1177/1094428117715067

Ehm, J.-H., Hasselhorn, M., & Schmiedek, F. (2019). Analyzing the developmental relation of academic self-concept and achievement in elementary school children: Alternative models point to different results. *Developmental Psychology, 55*(11), 2336–2351. https://doi.org/10.1037/dev0000796.supp (Supplemental)

Fleischmann, M., Hübner, N., Marsh, H. W., Guo, J., Trautwein, U., & Nagengast, B. (2022). Which class matters? Juxtaposing multiple class environments as frames-of-reference for academic self-concept formation. *Journal of Educational Psychology, 114*(1), 127–143. https://doi.org/10.1037/edu0000491.supp (Supplemental)

Gilbreath, B., Kim, T.-Y., & Nichols, B. (2011). Person-environment fit and its effects on university students: A response surface methodology study. *Research in Higher Education, 52*(1), 47–62. https://doi.org/10.1007/s11162-010-9182-3

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*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Hrsg.), *The Cambridge handbook of cognition and education*. (S. 647–677). Cambridge University Press. https://doi.org/10.1017/9781108235631.026

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction*, *24*, 37–47. https://doi.org/10.1016/j.learninstruc.2012.10.001

Händel, M., Harder, B., & Dresel, M. (2020). Enhanced monitoring accuracy and test performance: Incremental effects of judgment training over and above repeated testing. *Learning and Instruction, 65*. https://doi.org/10.1016/j.learninstruc.2019.101245

Helmke, A. (1998). Vom Optimisten zum Realisten? Zur Entwicklung des Fähigkeitsselbstkonzeptes vom Kindergarten bis zur 6. Klassenstufe. In F. E. Weinert (Hrsg.), *Entwicklung im Kindesalter* (pp. 115-132). Beltz.

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., Küfner, A. C. P., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2019a). Is accurate, positive, or inflated self-perception most advantageous for psychological adjustment? A competitive test of key hypotheses. *Journal of Personality and Social Psychology, 116*(5), 835–859. https://doi.org/10.1037/pspp0000204

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2018). Enhanced versus simply positive: A new condition-based regression analysis to disentangle effects of self-enhancement from effects of positivity of self-view. *Journal of Personality and Social Psychology, 114*(2), 303–322. https://doi.org/10.1037/pspp0000134

Humberg, S., Dufner, M., Schönbrodt, F. D., Geukes, K., Hutteman, R., van Zalk, M. H. W., Denissen, J. J. A., Nestler, S., & Back, M. D. (2022). The true role that suppressor effects play in condition-based regression analysis: None A reply to Fiedler (2021). *Journal of Personality and Social Psychology, 123*(4), 884–888. https://doi.org/10.1037/pspp0000428

Humberg, S., Nestler, S., & Back, M. D. (2019b). Response surface analysis in personality and social psychology: Checklist and clarifications for the case of congruence hypotheses. *Social Psychological and Personality Science, 10*(3), 409–419. https://doi.org/10.1177/1948550618757600

Jansen, M., Boda, Z., & Lorenz, G. (2022). Social comparison effects on academic self-concepts—Which peers matter most? *Developmental Psychology, 58*(8), 1541–1556. https://doi.org/10.1037/dev0001368.supp (Supplemental)

Jansen, M., Lüdtke, O., & Robitzsch, A. (2020). Disentangling different sources of stability and change in students' academic self-concepts: An integrative data analysis using the STARTS model. *Journal of Educational Psychology, 112*, 1614–1631. https://doi.org/10.1037/edu0000448

Jing, E. L., Lupton, N. C., & Ansari, M. A. (2021). A cultural value congruence approach to organizational embeddedness. *Journal of Personnel Psychology, 20*(4), 164–175. https://doi.org/10.1027/1866-5888/a000280

John, O. P., & Robins, R. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219. https://doi.org/10.1037//0022-3514.66.1.206

Jungert, T., Hesser, H., & Träff, U. (2014). Contrasting two models of academic self-efficacy—Domain-specific versus cross-domain—In children receiving and not receiving special instruction in mathematics. *Scandinavian Journal of Psychology, 55*(5), 440–447. https://doi.org/10.1111/sjop.12139

Kalogiannis, P., Papaioannou, A., Sagovich, A., & Abatzoglou, G. (2011). Reciprocal effects between self-concept and school performance, preparation for school, and life satisfaction: A longitudinal study. *Hellenic Journal of Psychology, 8*(1), 96–122.

Kersting, M., & Althoff, K. (2004). *RT Rechtschreibungstest.* Hogrefe.

Leduc, C., & Bouffard, T. (2017). The impact of biased self-evaluations of school and social competence on academic and social functioning. *Learning and Individual Differences, 55*, 193–201. https://doi.org/10.1016/j.lindif.2017.04.006

Lee, E. J. (2021). Biased self-estimation of maths competence and subsequent motivation and achievement: Differential effects for high- and low-achieving students. *Educational Psychology, 41*(4), 446–466. https://doi.org/10.1080/01443410.2020.1821869

Lee, E. J. (2022). Are overconfidence and the accurate calibration of performance mutually incompatible? *Japanese Psychological Research*. https://doi.org/10.1111/jpr.12409

Lee, C.-Y., & Kung, H.-Y. (2018). Math self-concept and mathematics achievement: Examining gender variation and reciprocal relations among junior high school students in Taiwan. *Eurasia Journal of Mathematics, Science & Technology Education, 14*(4), 1239–1252. https://doi.org/10.29333/ejmste/82535

Little, T. D. (2013). *Longitudinal structural equation modeling: Methodology in the social sciences*. Guildford press.

Lu, H., Chen, X., & Qi, C. (2023). Which is more predictive: Domain- or task-specific self-efficacy in teaching and outcomes? *British Journal of Educational Psychology, 93*(1), 283–298. https://doi.org/10.1111/bjep.12554

Marsh, H. W. (2022). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology.* https://doi.org/10.1037/edu0000750.supp (Supplemental)

Marsh, H. W. (2023). Extending the reciprocal effects model of math self-concept and achievement: Long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology*, *115*(2), 193–211. https://doi.org/10.1037/edu0000750.supp

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology, 96*(3), 518-522. https://doi.org/10.1037/0022-0663.96.3.518

Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, O., & Brettschneider, W.-D. (2007). Longitudinal study of preadolescent sport self-concept and performance: Reciprocal effects and causal ordering. *Child Development, 78*, 1640–1656. https://doi.org/10.1111/j.1467-8624.2007.01094.x

Marsh, H. W., Hau, K.-T., & Kong, C.-K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal, 39*(3), 727–763. doi:10.3102/00028312039003727

Marsh, H. W., & Martin, A. W. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59-77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Parker, P. D., Guo, J., Basarkod, G., Niepel, C., & Van Zanden, B. (2021). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology, 121*(1), 168–183. https://doi.org/10.1037/pspp0000306.supp (Supplemental)

Marsh, H. W., Pekrun, R., & Lüdtke, O. (2022). Directional ordering of self-concept, school grades, and standardized tests over five years: New tripartite models juxtaposing within- and between-person perspectives. *Educational Psychology Review.* https://doi.org/10.1007/s10648-022-09662-9

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., Xu, K. M., Parker, P. D., Hau, K.-T., Pekrun, R., Elliot, A., Guo, J., Dicke, T., & Basarkod, G. (2021). Moderation of the big-fish-little-pond effect: Juxtaposition of evolutionary (Darwinian-economic) and achievement motivation theory predictions based on a Delphi approach. *Educational Psychology Review, 33*(4), 1353–1378. https://doi.org/10.1007/s10648-020-09583-5

McNamee, R., (2005). Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine, 62*, 500-506. https://doi.org/10.1136/oem.2002.001115.

Meyer, J. P., Hecht, T. D., Gill, H., & Toplonytsky, L. (2010). Person–organization (culture) fit and employee commitment under conditions of organizational change: A longitudinal study. *Journal of Vocational Behavior, 76*(3), 458–473. https://doi.org/10.1016/j.jvb.2010.01.001

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2019). *Kernlehrplan für die Sekundarstufe I Gymnasium in Nordrhein-Westfalen. Deutsch.* Abgerufen von https://www.schulentwicklung.nrw.de/lehrplaene/lehrplannavigator-s-i/gymnasium-aufsteigend-ab-2019-20/index.html

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional comparisons: Theoretical assumptions and empirical results. In J. D. Wright (Hrsg.), *International encyclopedia of the social and behavioral sciences 2nd ed*. (S. 430–436). Elsevier.

Möller, J., & Köller, O. (1998). Dimensionale und soziale Vergleiche nach schulischen Leistungen. *Zeitschrift Für Entwicklungspsychologie und Pädagogische Psychologie, 30*(3), 118–127.

Möller, J., Zitzmann, S., Helm, F., Machts, N., & Wolff, F. (2020). A meta-analysis of relations between achievement and self-concept. *Review of Educational Research, 90*(3), 376–419. https://doi.org/10.3102/0034654320919354

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: An International Journal, 20*(4), 849–873. https://doi.org/10.1007/s11218-017-9395-9

Niepel, C., Brunner, M., & Preckel, F. (2014b). 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. https://doi.org/10.1037/a0036307

Parker, P., Dicke, T., Guo, J., Basarkod, G., & Marsh, H. (2021). Ability stratification predicts the size of the big-fish-little-pond effect. *Educational Researcher, 50*(6), 334–344. https://doi.org/10.3102/0013189X20986176

Paulhus, D. L. & John, O. P. (1998). Egoistic and moralistic bias in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025–1060. https://doi.org/10.1111/1467-6494.00041

Pijeira-Díaz, H. J., van de Pol, J., Channa, F., & de Bruin, A. (2023). Scaffolding self-regulated learning from causal-relations texts: Diagramming and self-assessment to improve metacomprehension accuracy? *Metacognition and Learning*. https://doi.org/10.1007/s11409-023-09343-0

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? Zusammenhänge zwischen optimistischen, realistischen und pessimistischen Selbstkonzepten und der Leistungsentwicklung von Grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 48*(1), 14–26. https://doi.org/10.1026/0049-8637/a000140

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A test of the reciprocal-effects model of academic achievement and academic self-concept in regular classes and special classes for the gifted. *Gifted Child Quarterly, 61*(2), 103–116. doi:10.1177/0016986216687824

Robins, R. W. & John, O. P. (1997). The quest for self-insight: Theory and research on the accuracy of self-perception. In H. Hogan, J. Johnson & S. Briggs (Hrsg.), *Handbook of personality psychology* (S. 649–679). Academic Press.

Rohrer, J. M., & Murayama, K. (2023). These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis. *Advances in Methods and Practices in Psychological Science, 6*, 1-14. https://doi.org/10.1177/25152459221140842

Ross, M., & Wilson, A. E. (2003). Autobiographical memory and conceptions of self: Getting better all the time. *Current Directions in Psychological Science, 12* , 66–69. https://doi.org/10.1111/1467-8721.01228

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., & Deci, E. L. (2019). Brick by brick: The origins, development, and future of self-determination theory. In A. J. Elliot (Hrsg.), *Advances in motivation science, Volume 6* (S. 111–156). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.01.001

Schmidt, S., Ennemoser, M., & Krajewski, K. (2013). *Deutscher Mathematiktest für neunte Klassen (DEMAT 9)*. Hogrefe.

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits- und -verständnistest für die Klassen 6–12.* Hogrefe.

Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., & Yeung, A. S. (2015). The reciprocal effects model revisited: Extending its reach to gifted students attending academically selective schools. *Gifted Child Quarterly, 59*(3), 143–156. https://doi.org/10.1177/0016986215583870

Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2014). The reciprocal relations between self-concept, motivation and achievement: Juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology, 34*(1), 49–72. doi:10.1080/01443410.2013.825232

Seo, H., Little, T. D., Shogren, K. A., & Lang, K. M. (2016). On the benefits of latent variable modeling for norming scales: The case of the Supports Intensity Scale – Children's Version. *International Journal of Behavioral Development, 40*(4), 373-384. https://doi.org/10.1177/0165025415591230

Sewasew, D., & Koester, L. S. (2019). The developmental dynamics of students' reading self-concept and reading competence: Examining reciprocal relations and ethnic-background patterns. *Learning and Individual Differences, 73*, 102–111. https://doi.org/10.1016/j.lindif.2019.05.010

Sewasew, D., & Schroeders, U. (2019). The developmental interplay of academic self-concept and achievement within and across domains among primary school students. *Contemporary Educational Psychology, 58*, 204–212. https://doi.org/10.1016/j.cedpsych.2019.03.009

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25*(4), 543–554. https://doi.org/10.1007/s10869-010-9183-4

Siefer, K., Leuders, T., & Obersteiner, A. (2021). Which task characteristics do students rely on when they evaluate their abilities to solve linear function tasks?—A task-specific assessment of self-efficacy. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.596901

Skaalvik, E. M., & Valas, 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. https://doi.org/10.1080/00220979909598349

Stocker, J., Abu-Hilal, M., Hermena, E., AlJassmi, M., & Barbato, M. (2021). Internal/external frame of reference model and dimensional comparison theory: A novel exploration of their applicability among arab high school students. *Educational Psychology*. https://doi.org/10.1080/01443410.2021.1887455

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21–27. https://doi.org/10.1037/0033-2909.116.1.21

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Hrsg.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (S. 187-214). Springer International Publishing.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133. https://doi.org/10.1207/s15326985ep3902_3

van der Westhuizen, L., Arens, A. K., Greiff, S., Fischbach, A., & Niepel, C. (2022). The generalized internal/external frame of reference model with academic self-concepts, interests, and anxieties in students from different language backgrounds. *Contemporary Educational Psychology, 68*, 1–13. https://doi.org/10.1016/j.cedpsych.2021.102037

van Loon, M. H., & Oeri, N. S. (2023). Examining on-task regulation in school children: Interrelations between monitoring, regulation, and task performance. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000781

Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J.-E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian Journal of Educational Research, 58*, 734–756. https://doi.org/10.1080/00313831.2014.904419

Wan, S., Lauermann, F., Bailey, D. H., & Eccles, J. S. (2021). When do students begin to think that one has to be either a “math person” or a “language person”? A meta-analytic review. *Psychological Bulletin, 147*(9), 867–889. https://doi.org/10.1037/bul0000340.supp (Supplemental)

Wan, S., Lauermann, F., Bailey, D. H., & Eccles, J. S. (2023). Developmental changes in students’ use of dimensional comparisons to form ability self-concepts in math and verbal domains. *Child Development, 94*(1), 272–287. https://doi.org/10.1111/cdev.13856

Wang, Z. (2020). When large-scale assessments meet data science: The big-fish-little-pond effect in fourth- and eighth-grade mathematics across nations. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.579545

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2018). Changes in the relation between competence beliefs and achievement in math across elementary school years. *Child Development, 89*(2), 138–156. https://doi.org/10.1111/cdev.12806

Werts, C. E., & Watley, D. J. (1969). A student’s dilemma: Big fish-little pond or little fish-big pond. *Journal of Counseling Psychology, 16*(1), 14–19. https://doi.org/10.1037/h0026689

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49–78. https://doi.org/10.1007/BF02209024

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students’ subjective task values and motivation: A look back and a look forward. In A. J. Elliot (Hrsg.), *Advances in* *motivation science, Voume 7* (S. 161–198). Elsevier Academic Press. https://doi.org/10.1016/bs.adms.2019.05.002

Wolff, F. (2022). Dimensional comparisons: Much is known about the effects, less about the processes. *Social and Personality Psychology Compass*, *16*(6). https://doi.org/10.1111/spc3.12664

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology, 110*(7), 1005–1025. https://doi.org/10.1037/edu0000248

Wolff, F., Lüdtke, O., Helm, F., & Möller, J. (2021b). Integrating the big-fish-little-pond effect, the basking-in-reflected-glory effect, and the internal/external frame of reference model predicting students' individual and collective academic self-concepts. *Contemporary Educational Psychology, 65*. https://doi.org/10.1016/j.cedpsych.2021.101952

Wolff, F., & Möller, J. (2021). Dimensional comparison theory: Minimal intervention affects strength of dimensional comparison effects. *Journal of Experimental Education, 89*(4), 625–642. https://doi.org/10.1080/00220973.2020.1843128

Wolff, F., & Möller, J. (2022). An individual participant data meta-analysis of the joint effects of social, dimensional, and temporal comparisons on students' academic self-concepts. *Educational Psychology Review, 34*(4), 2569–2608. https://doi.org/10.1007/s10648-022-09686-1

Wolff, F., Nagy, G., Retelsdorf, J., Helm, F., Köller, O., & Möller, J. (2019). The 2I/E model: Integrating temporal comparisons into the internal/external frame of reference model. *Journal of Educational Psychology, 111*(7), 1131–1161. https://doi.org/10.1037/edu0000319.supp (Supplemental)

Wolff, F., Sticca, F., Niepel, C., Götz, T., Van Damme, J., & Möller, J. (2021a). The reciprocal 2I/E model: An investigation of mutual relations between achievement and self-concept levels and changes in the math and verbal domain across three countries. *Journal of Educational Psychology*, *113*(8), 1529–1549. https://doi.org/10.1037/edu0000632

Wolff, F., Wigfield, A., Möller, J., Dicke, A.-L., & Eccles, J. S. (2020). Social, dimensional, and temporal comparisons by students and parents: An investigation of the 2I/E model at the transition from elementary to junior high school. *Journal of Educational Psychology, 112*(8), 1644–1660. https://doi.org/10.1037/edu0000440

Wolff, F., Zitzmann, S., & Möller, J. (2021c). Moderators of dimensional comparison effects: A comprehensive replication study putting prior findings on five moderators to the test and going beyond. *Journal of Educational Psychology, 113*(3), 621–640. https://doi.org/10.1037/edu0000505.supp (Supplemental

Wu, H., Guo, Y., Yang, Y., Zhao, L., & Guo, C. (2021). A meta-analysis of the longitudinal relationship between academic self-concept and academic achievement. *Educational Psychology Review.* https://doi.org/10.1007/s10648-021-09600-1

Zhang, J., Chiu, M. M., & Lei, H. (2023). Achievement, self-concept and anxiety in mathematics and English: A three-wave cross-lagged panel study. *British Journal of Educational Psychology*, *93*(1), 56–72. https://doi.org/10.1111/bjep.12539

Zell, E., & Lesick, T. L. (2021). Taking social comparison to the extremes: The huge-fish-tiny-pond effect in self-evaluations. *Social Psychological and Personality Science, 12*(6), 1030–1038. https://doi.org/10.1177/1948550620956535

4 Anhang

4.1 Eigenanteile des Doktoranden bei den Beiträgen der Dissertation

4.1.1 Veröffentlichte Beiträge

Beitrag I:

Paschke, P., Weidinger, A. F., & Steinmayr, R. (2020). Separating the effects of self-evaluation bias and self-view on grades [Trennung der Effekte des Selbsteinschätzungsbias und der Selbsteinschätzung auf Schulnoten]. *Learning and Individual Differences, 83-84*. https://doi.org/10.1016/j.lindif.2020.101940

Formulierung der Fragestellung: Die Fragestellung wurde von Patrick Paschke entworfen.
Konzeption des Beitrags: Der Beitrag wurde hauptverantwortlich von Patrick Paschke konzeptualisiert, wobei die Ko-Autorinnen Beratung leisteten.
Statistische Auswertungen: Die statistischen Auswertungen wurden von Patrick Paschke durchgeführt.
Schriftliche Abfassung des Beitrags: Patrick Paschke verfasste den Text des Beitrags hauptverantwortlich. Beide Ko-Autorinnen gaben Rückmeldungen zu den Textentwürfen, wonach Patrick Paschke den Text überarbeitete und finalisierte.

Beitrag II:

Paschke, P, Weidinger, A. F., & Steinmayr, R. (2023). Linear and nonlinear relationships between self-evaluation and self-evaluation bias with grades [Lineare und nicht lineare Zusammenhänge von Selbsteinschätzung und Selbsteinschätzungsbias mit Schulnoten]. *Learning and Individual Differences, 102*. https://doi.org/10.1016/j.lindif.2023.102266

Formulierung der Fragestellung: Die Fragestellung wurde von Patrick Paschke entworfen.

Konzeption des Beitrags: Der Beitrag wurde hauptverantwortlich von Patrick Paschke konzeptualisiert, wobei die Ko-Autorinnen Beratung leisteten.

Statistische Auswertungen: Die statistischen Auswertungen wurden von Patrick Paschke durchgeführt.

Schriftliche Abfassung des Beitrags: Patrick Paschke verfasste den Text des Beitrags hauptverantwortlich. Beide Ko-Autorinnen gaben Rückmeldungen zu den Textentwürfen, wonach Patrick Paschke den Text überarbeitete und finalisierte.

4.1.2 Weiterführende Analyse

Beitrag III:

Paschke, P. & Steinmayr, R. (2023). *The Effects of Self-Evaluation, Competence, and their Discrepancy on Academic Achievement in Math Mediated by Expectancy of Success and Subjective Task Values [Die Effekte von Selbsteinschätzung, Kompetenz und ihrer Diskrepanz auf akademische Leistung in Mathematik mediiert durch Erfolgserwartung und subjektive Werte].* [Manuscript in preperation for publication]. Institut für Psychologie, TU Dortmund University.

Formulierung der Fragestellung: Die Fragestellung wurde von Patrick Paschke entworfen.

Konzeption des Beitrags: Der Beitrag wurde hauptverantwortlich von Patrick Paschke konzeptualisiert, wobei Ricarda Steinmayr Beratung leistete.

Statistische Auswertungen: Die statistischen Auswertungen wurden von Patrick Paschke durchgeführt.

Schriftliche Abfassung des Beitrags: Patrick Paschke verfasste den Text des Beitrags hauptverantwortlich. Ricarda Steinmayr gab Rückmeldung zu den Textentwürfen, wonach Patrick Paschke den Text überarbeitete und finalisierte.

______________________ am 01.07.2023	______________________ am 29.06.2023
Patrick Paschke	Ricarda Steinmayr
______________________ am 01.07.2023	
Anne F. Weidinger	

4.2 Eidesstattliche Erklärung

Hiermit versichere ich **schriftlich** und **eidesstattlich** gemäß § 11 Abs. 2 PromO v. 08.02.2011/08.05.2013:

1. Die von mir vorgelegte Dissertation ist selbstständig verfasst und alle in Anspruch genommenen Quellen und Hilfen sind in der Dissertation vermerkt worden.
2. Die von mir eingereichte Dissertation ist weder in der gegenwärtigen noch in einer anderen Fassung an der Technischen Universität Dortmund oder an einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegt worden.

________________________________ ________________________________

Ort, Datum Unterschrift

3. Weiterhin erkläre ich **schriftlich** und **eidesstattlich**, dass mir der „Ratgeber zur Verhinderung von Plagiaten“ und die „Regeln guter wissenschaftlicher Praxis der Technischen Universität Dortmund“ bekannt sind und von mir in der vorgelegten Dissertation befolgt worden sind.

________________________________ ________________________________

Ort, Datum Unterschrift