Perceiving Achievement in Schools: How do Self-appraisals, Peer Appraisals and Achievement Relate to Each Other?

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Thomas Lösch, M. Sc.

aus Darmstadt

Tübingen

2016

Tag der mündlichen Prüfung:	13.01.2017
Dekan:	Professor Dr. rer. soc. Josef Schmid
1. Gutachter:	Professor Dr. Augustin Kelava
2. Gutachter:	Professor Dr. Benjamin Nagengast

Acknowledgements

This dissertation was financed by the LEAD Graduate School & Research Network at the University of Tübingen and I acknowledge this support with deep gratitude. This work has greatly benefited from the guidance, support, and inspiration of several people.

I would especially like to thank my three supervisors, Prof. Dr. Augustin Kelava, Prof. Dr. Benjamin Nagengast, and Prof. Dr. Ulrich Trautwein, for their guidance, feedback, for providing a dynamic and highly inspiring research environment, and for many opportunities to develop and learn, academically and personally. To Benjamin Nagengast, for continuously stimulating me to develop clear ideas. To Ulrich Trautwein, for the clarity and precision of his arguments and writing. And to Augustin Kelava, for all the things that I learned from him through the years.

Additionally, I would like to thank Prof. Dr. Oliver Lüdtke, who got into the role of being my unofficial, fourth supervisor. I had a great time in your lab in Kiel and I am deeply grateful for your advices, for your ideas, and your constructive management of our projects.

Moreover, I want to thank Dr. Isabelle Häfner and Dr. Marion Spengler. You were available, when there was a dire need for someone to be available. Thank you so much for your help! In a similar vein, I want to thank Dr. Norman Rose for all the times I could ask you for advice (or for a chance to grant me access to some more working memory…)!

I am highly grateful to be able to work in the LEAD Graduate school and for the continuous support of the LEAD staff. I want to thank especially LEAD's shepherd, Mareike Bierlich, who constantly keeps the Graduate School in the most functional state possible. Similarly, I want to thank Sophie Freitag, and all the other administrative staff as well as the junior research group leaders, Dr. Aiste Jusyte, Dr. Nicole Tieben, and Dr. Thomas Dresler, who keep everything running. Thank you for being here and for having lunch together every once in a little while!

The time in LEAD was so much richer thanks to inspiring conversations. I want to thank all the other doctoral candidates in the LEAD Graduate School, be it from Europastraße or Gartenstraße, old ones, new ones, and finished ones, office roommates or office neighbors for the time together. I am so happy to have the chance to work with you and get to know all of you. I learned so much, from different disciplines, different countries, and different ideas. Every one of

you is very excellent! Representing all the others, I want to thank especially Sina, Christina, Leona, Michèle, Fabian, Michael, Eike, and Maria. Thanks to all of you!

Very special thanks go to Dr. Tanja Gerlach. Thank you, for being there in the decisive moments of my career. You provided me with the best advices and helped me through some tough decisions. Without you, I would probably not have started this thesis in the first place and in the end, I did research in fields you have introduced me to – way before I started this dissertation.

I am very grateful for my friends, especially Jann, Daniel and Patrick. Thank you so much for your patience, and the energizing time we had together!

In einem Maß, das in einen Absatz nicht hinein passen will, bedanke ich mich bei allen meinen Eltern und ganz besonders meiner Mutter Lydia. Ihr habt mich in der Dissertation und schon lange davor mit ganzem Herzen und voller Kraft unterstützt. Ich wusste immer, dass ich mich an euch wenden kann und dass ihr hinter mir steht.

Finally, I want to thank you, Juliane. In so many ways, this dissertation would have been less complete without you. Thank you for going through the dissertation together, for providing your comments, your support, for all the inspiration, and for being a team. Thank you for being there when I came in!

To all of you: Thank you!

Abstract

Self-appraisals of achievement (short: self-appraisals) are guiding students' behavior in school and they determine academic success. In addition, *peer appraisals* about a student were shown to have an effect on that student's future development, including academic achievement. In turn, academic achievement is undoubtedly a precursor of both self-appraisals and peer appraisals. However, beyond the notion that these three constructs are associated, major aspects of their interrelations are not well understood. Since self-appraisals and peer appraisals provide a potential lever to positively affect students' academic achievement, it is vital to better understand the strength of their connection, their functional relationship, and their relatedness in terms of shared correlates. Therefore, the present thesis aims at investigating the links between self-appraisals, peer appraisals, and achievement with schools as context.

Study 1 targeted the link between self-appraisals and achievement by investigating the association between postdictions of and performance in the final school exam in Germany (the Abitur). Using the domains of math and English, the average correlation between both postdictions and exam grades, their average deviation (i.e., postdiction bias), as well as individual differences in postdiction bias were explored in a sample of 2,164 high school students. On average, students' postdictions were highly associated with achievement, indicating that students could accurately appraise their performance. Individual differences in postdiction bias were predicted by cognitive abilities, internal and contextual references, as well as evaluative tendencies.

Study 2 targeted the triangular connection of self-appraisals, peer appraisals, and achievement. An interpersonal approach to academic self-concept in math was proposed that disentangled variance of self-concept ratings into several components: a "trait" of perceived achievement that was shared with peers (target effect), a "method" of how a student rated peers' achievement (perceiver effect), and a student's self-appraisal that was not shared with others (self-enhancement). The interpersonal approach was investigated using a sample of 1,549 school students in 87 classes of two age cohorts, who rated their own and their peers' competence. Results supported the validity of the target effect as representing a shared competence appraisal. In turn, this shared appraisal predicted achievement gains. Conversely, self-enhancement had a comparatively small effect on improvements in achievement.

Study 3 mainly targeted the link between achievement and peer appraisals by investigating whether the internal/external frame of reference (I/E) model would be equally valid for peer ratings of competence as it is valid for academic self-concept. The I/E model posits that students' academic self-concept in one domain (e.g., math) is positively associated with achievement in the same domain, but negatively associated with achievement in a comparison domain (e.g., English). The I/E model was compared between peer ratings of competence and academic self-concept in math and English, using a sample of 850 seventh grade students in 47 classes, who rated every classmate's competence in both domains within a round-robin design. Results supported the I/E model for academic self-concept but not for peer ratings.

In a general discussion, findings of the three studies are summarized against the background of the current state of research that is relevant for interpersonal achievement perception, including research in education, motivation, cognition, and personality. Thereby, it is emphasized how this thesis may be used to inform future research on an interpersonal conception of achievement appraisals and how the results may inform practice in schools to foster the optimal level of self-appraisals.

Zusammenfassung

Leistungs-Selbsteinschätzungen (kurz: Selbsteinschätzungen) beeinflussen sowohl das schulische Verhalten als auch den schulischen Erfolg von SchülerInnen. Zusätzlich wirkt sich die Leistungs-Einschätzung von MitschülerInnen über einzelne SchülerInnen (kurz: Peer-Einschätzungen) auf dessen zukünftige Entwicklung der schulischen Leistung aus. Umgekehrt beeinflussen auch schulische Leistungen sowohl Selbsteinschätzungen als auch Peer-Einschätzungen. Trotzdem ist über die Auffassung hinaus, dass diese drei Konstrukte miteinander in Verbindung stehen, nicht viel über die genauen Zusammenhänge bekannt. Da jedoch Selbsteinschätzungen und Peer-Einschätzungen eine Möglichkeit darstellen, den schulischen Erfolg von SchülerInnen positiv zu beeinflussen, ist es wichtig zu wissen, warum und in welchem Ausmaß die Konstrukte miteinander und mit anderen Konstrukten zusammenhängen. Deshalb untersucht die vorliegende Dissertation die Zusammenhänge zwischen Selbsteinschätzungen, Peer-Einschätzungen und Leistung im Schulkontext.

Studie 1 beschäftigte sich mit dem Zusammenhang zwischen Selbsteinschätzugen und Leistung. Dazu wurde die Beziehung zwischen der Abiturleistung und Selbsteinschätzungen der Abiturleistung (im folgenden *Postdiktionen*)untersucht. In einer Stichprobe von 2164 Gymnasiasten wurden die durchschnittlichen Korrelationen zwischen Postdiktionen und Abiturnoten in Mathematik und Deutsch, deren durchschnittliche Abweichung (Postdiktions-Bias) und individuelle Unterschiede in im Postdiktions-Bias analysiert. Im Durchschnitt stimmten die Postdiktionen der SchülerInnen stark mit der Leistung überein. Dies zeigt, dass SchülerInnen ihre Leistung akkurat einschätzen konnten. Individuelle Unterschiede des Postdiktions-Bias wurden durch kognitive Fähigkeiten, innere und äußere Bezugsrahmen und evaluative Tendenzen vorhergesagt.

Studie 2 beschäftigte sich mit dem Zusammenhang zwischen den drei Konstrukten Selbsteinschätzungen, Peer-Einschätzungen und Leistung. Dazu wurde ein interpersoneller Ansatz für das Fähigkeitsselbstkonzept verwendet. In diesem Ansatz wird die Varianz von Einschätzungen des Fähigkeitsselbstkonzept in folgende Komponenten zerlegt: einen "Trait" von wahrgenommener Leistung, die mit Peers geteilt wurde (Target Effekt), eine "Methode“, wie SchülerInnen ihre Peers einschätzten (Beobachter Effekt) und die Selbsteinschätzung von SchülerInnen, die nicht mit anderen geteilt wurde (Selbstüberschätzung). Der Ansatz wurde

anhand einer Stichprobe von 1549 SchülerInnen in 87 Klassen aus zwei Jahrgängen untersucht. Die SchülerInnen mussten ihre eigene Leistung und die Leistung ihrer Peers einschätzen. Ergebnisse bestätigten die Validität des Target Effekts als geteilte Einschätzung einer Leistung. Diese geteilte Einschätzung sagte Leistungszuwächse vorher. Im Gegensatz dazu hatte Selbstüberschätzung nur einen schwachen Effekt auf Leistungszuwächse.

Studie 3 beschäftigte sich hauptsächlich mit dem Zusammenhang zwischen Leistung und Peer-Einschätzungen. Dazu wurde untersucht, ob das internal/external frame of reference (I/E) Modell im gleichen Ausmaß für Peer-Einschätzungen wie für das Fähigkeitsselbstkonzept gelten würde. Das I/E Modell postuliert, dass das Fähigkeitsselbstkonzept von SchülerInnen in einem Fach (zum Beispiel Mathematik) positiv mit der Leistung im selben Fach, aber negativ mit der Leistung in einem Vergleichs-Fach (zum Beispiel Englisch) zusammenhängt. Das I/E Modell für Peer-Einschätzungen in Mathe und Englisch wurde mit dem I/E Modell für das Fähigkeitsselbstkonzept in Mathe und Englisch verglichen. Dazu wurden 850 Siebtklässler aus 47 Klassen befragt, die die Leistung aller MitschülerInnen in einem Round-Robin-Design einschätzen mussten. Ergebnisse bestätigten das I/E Modell für das schulische Selbstkonzept aber nicht für Peer-Einschätzungen.

In einer allgemeinen Diskussion werden die drei Studien vor dem Hintergrund aktueller Forschung aus den Bereichen Bildung, Motivation, Kognition und Persönlichkeit zu interpersoneller Wahrnehmung von Leistung zusammengefasst. Dabei wird dargestellt, wie die vorliegende Dissertation zukünftige Forschung zu interpersoneller Wahrnehmung von Leistung anregen kann. Zusätzlich wird ausgeführt, wie die Ergebnisse der Studien verwendet werden können, um ein optimales Niveau von Selbsteinschätzungen in der Schule zu fördern.

Content

1

Introduction and Theoretical Framework

1. Introduction and Theoretical Framework

Hat der alte Hexenmeister	That old sorcerer has vanished
Sich doch einmal wegbegeben!	And for once has gone away!
Und nun sollen seine Geister	Spirits called by him, now banished,
Auch nach meinem Willen leben.	My commands shall soon obey.
Seine Wort und Werke	Every step and saying
Merkt ich und den Brauch,	That he used, I know,
Und mit Geistesstärke	And with sprites obeying
Tu ich Wunder auch.	My arts I will show.

With this verse, Goethe starts his famous poem *The sorcerer's apprentice*, which is not only a classic of German literature but also a classic of German school lessons. The poem tells the story of the apprentice, who is left alone by his master. Finally freed from the restrictions his master imposed on him, he is eager to practice the magic he has learned. Believing in his magic abilities, he starts out by cleaning the house through magical force. Yet, shortly after he has started, the apprentice suddenly recognizes that he has forgotten how to end the spell. The magic gets out of hand, and the apprentice is unable to stop what he has started. What began as a bucket of water to clean the house ends as a tidal wave that washes over the apprentice. The poem concludes with the arrival of the old sorcerer, who is finally able to end the uncontrolled chaos. In short, the poem tells a story about a student who overestimates his abilities and who has to face dire consequences that arose from this misguided appraisal.

As demonstrated by this poem, the topic of individuals who fail to correctly appraise their own abilities is not a new one. Nevertheless, the link between appraisals of abilities, actual abilities, and the consequences of misperceiving ones abilities is not entirely understood. Whereas Goethe's poem suggests that one should better be aware of ones capabilities in line with the classic phrase "know thyself!" (Wilson, 2009), research in psychology draws a picture that is not conclusive yet. Taylor and Brown (1988) proposed that an overestimation of one's virtues is a normal behavior that is linked to a positive individual development, for instance, in terms of achievement. In a reply, Colvin and Block (1994) disagreed with this idea and argued for the contrary: Accurate appraisals of the self would be a sign of mental health and would lead to positive outcomes (Colvin, Block, & Funder, 1995).

Because the debate also revolved around achievement, it has implications for the school context. In educational psychology, self-appraisals of achievement (including academic self-

concept, judgments of learning and more) are a central element of several theories targeting motivation (Marsh & Shavelson, 1985; Wigfield & Eccles, 2000) and cognition (Bjork, Dunlosky, & Kornell, 2013; Nelson & Narens, 1990) in the context of schools. These theories align in their statement that self-appraisals are determinants of academic behavior. Yet the theories diverge in the proposed link of how self-appraisals translate into academic success. Research focusing on motivation in educational contexts proposes that self-appraisals should be optimistic, because more positive self-appraisals would lead to more effort and a higher persistence in academic tasks (Wigfield & Eccles, 2000). This idea is outlined by the notion that "educational policy statements throughout the world list self-concept enhancement as a central goal of education" (Marsh & Hau, 2003, p. 364). However, research focusing on cognition in educational contexts proposes that self-appraisals should be realistic representations of achievement, because accurate self-appraisals would be a necessary condition for controlling and regulating one's learning effectively (Dunlosky & Lipko, 2007). This idea is summarized with the statement that "greater accuracy in a person's judgments of performance (i.e., being well calibrated) creates greater potential for self-regulation" (Bol & Hacker, 2012, p. 1). Because these positions have not been brought together, it is yet unclear whether self-appraisals might be accurate representations of achievement and whether an inaccurate representation might be maladaptive or even beneficial in the specific contexts of schools.

But specifically in the school context, more factors affect students' self-appraisals than just their achievement: Students are situated in the social context of their class and their classmates. Individuals care about their social context and what others think about them (Leary & Baumeister, 2000). Therefore, an important, yet overlooked aspect of self-appraisals in schools might be peers' appraisals. It was theoretically outlined (Wigfield & Eccles, 2000) and empirically found (Gest, Domitrovich, & Welsh, 2005) that peer appraisals of achievement predict students' self-appraisals. Moreover, what classmates thought about individual students' achievement predicted future achievement of that student as well (Gest, Rulison, Davidson, & Welsh, 2008). But contrasting self-appraisals with peer appraisals, it is not understood which of them is more indicative of a positive development. For instance, it is unknown whether students develop positively, even when their self-appraisal diverges from their peers' appraisal.

In this thesis, I investigated the triangular connection of self-appraisals, achievement, and peer appraisals. The overarching goal of the thesis was to describe and identify associations

between all three constructs and investigate links of how they affect each other within the specific context of schools. Therein, emphasis was placed upon comparing approaches of educational psychology that rather focus on motivation with approaches that rather focus on cognition. To achieve this goal, this thesis comprises three studies. The first study investigated the link between self-appraisals and objective achievement, their average association and predictors of this association. Thereby, *postdictions* were used, because they are a specific self-appraisal that has mainly been used in cognitive research and was now investigated in a design typical for motivation research. The second study investigated the interrelations of the whole triangle and advocated for an interpersonal approach to self-appraisals. This approach was able to disentangle common and shared variance between self-appraisals and peer appraisals and so showed their unique and shared effect on achievement. Furthermore, this approach allowed to test whether overestimating one's achievement was associated with maladaptive or beneficial outcomes. Finally, the third study targeted the link between achievement and peer appraisals from a multidimensional perspective. Therein, it extended the interpersonal approach presented in the second study, because the study investigated whether comparison process between multiple domains formed peer appraisals in a similar way as self-appraisals.

The thesis begins with a presentation of the theoretical background that aims at an integrative view on the three major constructs of the thesis, which are *self-appraisals, achievement, and peer-appraisals.* First, I will start by shortly describing students' *self* as main actor of this thesis. Subsequently, the three main constructs and their link to each other will be introduced. First, I will present a classification model of self-appraisals that integrates disconnected lines of research. Second, I will describe the connection between self-appraisals and *achievement*, where a motivational perspective and a cognitive perspective ought to be be contrasted. Third, I will show how self-appraisals and peer appraisals are interrelated and I will introduce a model that connects all three constructs of this thesis. The theoretical background concludes with an outline of the three empirical studies of this thesis. The following chapters will describe the three studies of this thesis. Study 1 investigated the link between postdictions and exam grades in the Abitur. Study 2 introduced an interpersonal approach to investigate the interrelatedness of academic self-concept, peer ratings of competence, and academic achievement. Study 3 tested, whether peer ratings of competence were similarly affected by dimensional comparisons as academic self-concept. The thesis ends with a general discussion

that summarizes results to the three major constructs of this thesis. The thesis will conclude with practical recommendations about how teachers should handle students' appraisals and directions for further research that might build upon results of this thesis and its triangular conception of self-appraisals, peer appraisals, and achievement.

1.1 The Self and its Function

The self is one of the most fundamental constructs in psychology and is investigated (at least) in educational, social, and personality psychology. In one of the early works of psychology, James (1890) was one of the first to describe the self and its structure as constructs of psychological science. He distinguished between two aspects of the self: the *me* and the *ego*. The *me* describes self-appraisals and attributes belonging to an individual: "a man's Self is the sum total of all that he *can* call his" (James, 1890, p. 291, highlighted as in the original). This description refers to material (e.g., the body, possessions) and immaterial (e.g., reputation, abilities) things alike. Thus, the *me* could be seen as the object of thought. Conversely, the *ego* describes the self as actor or subject. That is, the *ego* describes the mental instance of individuals that experiences thoughts, in sum, individuals' consciousness.

In a more recent view on the self, Leary and Tangney (2003) reviewed multiple approaches on the term *self* and proposed an integrative definition. They define the core of the self as "the human capacity for reflexive thinking – the ability to take onself as the object of one's attention and thought" (Leary & Tangney, 2003, p. 6). Moreover, it was added that, the self is "a set of systems that permit human beings to reflect on themselves and to respond to those self-reflections cognitively, emotionally, and behaviorally" (Leary, 2004, p. 2). That is, the self is involved in behavior and perception where an individual is both the subject and the object of the action. Therein, the major function of this reflective process may be to "enable people to make predictions about their worlds, guide behavior, and maintain a sense of continuity, place, and coherence" that finally results in "guiding behavior" of individuals (Kwang & Swann, 2010, p. 264).

In order to guide students' academic behavior, a major function of the self in the school context is thus to reflect about ones achievement and learning. This general idea is found in theories on academic motivation (Wigfield & Eccles, 2000) and learning cognition (Nelson & Narens, 1990). The common idea of these positions is that the self's reflections about the own achievement guide the choices of an individual – be it what material to learn or what advanced course to take. Thus, one approach in educational psychology is to describe and explain how the self reflects on its achievement and how these reflections form, and how they result in behavior.

Therein, peers may play a major role in formation of the self. Already James (1890) proposed that the *social self* is an important part of the self, that is, the self as seen by others and connected with others. Similarly, Leary and Baumeister (2000) proposed that the positivity of self-reflections was a function of the positivity of others views about the self. And finally, this thought was also outlined in educational psychology, where others' appraisals were investigated as a determinant of students academic self (Gest et al., 2005; Wigfield & Eccles, 2000).

That is, the self and peers are related and they share a view on achievement. However, different research traditions focused on different aspects of the self and its associations with other variables. The following chapters will describe the state of research regarding achievement and its appraisal by the self and by peers in the context of schools.

1.2 Self-appraisals

There is a vast number of constructs that describe the self's reflection about its ability, competence or performance. Following, I will first outline their similarity and present reasons to compare them systematically. Afterward, I will present a classification model that was developed to suit these needs. I will then use the model to classify academic self-concept, judgments of learning, and some related constructs. I will conclude the section by presenting frame of reference effect, which are one major determinant of self-appraisals.

1.2.1 Reasons to Distinguish Self-appraisals

As an umbrella term, with *self-appraisals of achievement* (short: self-appraisals), I refer to ratings, judgments, or evaluations that use the self as rater or perceiver and that have the self's ability, competence, or performance in an academic context as the target. They are rather cognitive evaluations and may be contrasted to rather affective evaluations like values or interests (Eccles & Wigfield, 2002). Summarized by the term self-appraisals are several constructs, including the *academic self-concept* (Marsh & Shavelson, 1985), *academic self-efficacy* (Pajares, 1996), *success expectancies* (Eccles & Wigfield, 1995), *judgments of learning* (JOL; Nelson & Narens, 1990), *postdictions* (Hacker, Bol, Horgan, & Rakow, 2000), and many similar constructs, sometimes just labeled *self-evaluations* or *self-assessments* (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Zell & Krizan, 2014). The common core of all these constructs is that they describe an individual's perceived achievement (Bong & Skaalvik, 2003).

Although different self-appraisals all share a "common theme" (Zell & Krizan, 2014) in describing the self as a rater of its own achievement, these constructs are rooted in distinct theories like self-regulation or motivation. Accordingly, they have different theoretical underpinnings and different processes are proposed explaining their formation. More than ten years ago, Wigfield and Eccles (2000, p. 72) observed these differences and stated that "an important task remaining for future research is to examine more closely how similar and different these various measures are" and furthermore, Eccles and Wigfield (2002, p.127) forecasted that "the further integration of work on cognition, motivation, and self-regulation will remain an important topic for motivation researchers during the next decade". But despite these

calls for integrative work, not much was done to connect these different research traditions. Nevertheless, there are at least three reasons to distinguish between different self-appraisals.

First, similar constructs not being sufficiently differentiated and defined might lead to fallacies in their understanding (Block, 1995; Kelly, 1927; Marsh, Craven, John, Debus, & Hinkley, 2003) like it was found for academic self-concept by Shavelson, Hubner, and Stanton (1976). On the one hand, self-appraisals might be termed differently, but in fact may have an identical meaning. On the other hand, self-appraisals might be termed the same, but in fact may have a different meaning.

Second, different conceptualizations of self-appraisals might lead to different correlates. For example, Bong and Skaalvik (2003) compared in their seminal work two of the most relevant self-appraisals in motivational research: academic self-concept and academic self-efficacy. They characterized academic self-concept as past-oriented, domain-specific self-perception of competence. Conversely, self-efficacy should represent a future-oriented, domain- and context-specific self-perception of confidence. These characterizations illustrated similarities like their domain-specificity and multidimensionality. Furthermore, both constructs shared some predictors like achievement. Yet, Bong and Skaalvik (2003) noticed important differences as well. For instance, the noted differences in the frame of reference they are based on: Whereas self-efficacy is mainly a appraisal in reference to a certain goal of an academic task (like solving an equation), academic self-concept is an appraisal in reference to a given norm (like a social norm). Recently, several investigations further clarified empirically distinct features. Self-efficacy compared to academic self-concept seemed to be much less influenced by social references (comparing one's achievement to achievement of others) and dimensional references (comparing one's achievement to one's achievement in a different domain; Jansen, Scherer, & Schroeders, 2014; Marsh et al., 2015; Marsh, Trautwein, Lüdtke, & Köller, 2008; Möller, Pohlmann, Köller, & Marsh, 2009; Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014). Regarding outcomes, it appears that self-efficacy better predicts aspects of achievement and academic self-concept better predicts aspects of academic choices (Jansen, Scherer, et al., 2014; Parker et al., 2014). For instance, in the study by Jansen et al. (2014), self-efficacy in science had a stronger association to science achievement, but self-concept in science had a stronger association to career aspirations in science. Similarly, in Parker et al. (2014) self-efficacy in math predicted university entry (regardless of university major) and better predicted achievement in

the final school year, whereas self-concept in math better predicted the selection of science or math courses in university. Although, both constructs are similar in the core of their definition, they can be distinguished on a theoretical and an empirical level.

That is, theoretical differences and the subsequent operationalization of different constructs might lead to the empirically observed features of the self-appraisals. Hence, it might prove useful to understand how differences in the conceptualizations of self-appraisals translate into empirical differences between constructs. For instance, it is not entirely clear which difference rendered self-efficacy to be less affected by reference effects than self-concept.

Third, and relatedly, the operationalization self-appraisals may be the reason for differing results in research traditions. Educational research in motivation (Wigfield & Eccles, 2000) and cognition (Bjork et al., 2013) find different links between self-appraisals and other constructs like achievement. One reason for these differences might lie in the self-appraisals used. That is, differences between research traditions might be caused by differences in self-appraisals.

1.2.2 Introducing a Classification Model of Self-appraisals

Following, I present a model that has the goal to relate differences in conceptualizations of self-appraisals to differences in how self-appraisals are associated with other constructs. The comparison by Bong and Skaalvik (2003) is highly useful, yet it focused only on two specific self-appraisals and disregarded many others. To fill this gap, Jansen & Lösch (in prep.) proposed a classification model that describes self-appraisals according to their central features (see Figure 1.1). The model focuses on the operational definition of self-appraisals, that is, on the mode and principles of how they are typically measured. The model classifies self-appraisals on three dimensions adapted from Bong and Skaalvik (2003): *frame of reference*, *specificity level*, and *time orientation*.

In the subsequent chapters, this model will be used to guide the description of self-appraisals in different research traditions. The model helps to show that there is a strong common idea behind self-appraisal. In addition, it allows describing and defining constructs in comparison to each other. A proposition by the classification model is that two self-appraisals that are similar to each other on the three dimensions will also be similar with regard to their features (e.g., their correlates). Yet, this proposition still needs empirical confirmation.

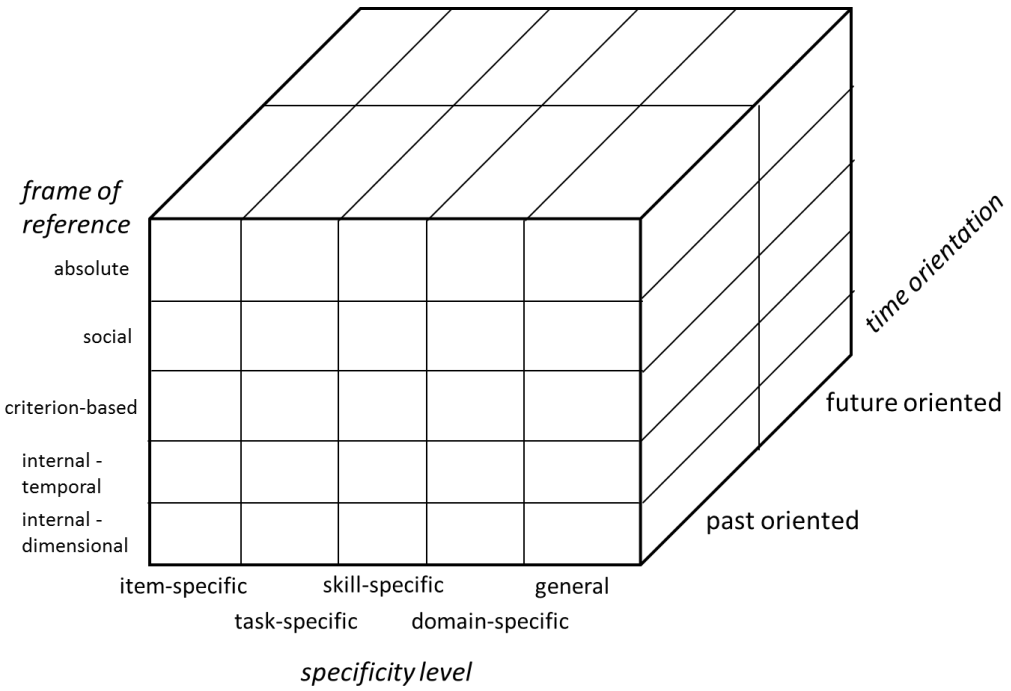

Figure 1.1. A classification model for self-appraisals of achievement as described in Jansen and Lösch (in prep.).

The *frame of reference* refers to comparisons evoked within items to measure a specific self-appraisal. *Absolute* means that no specific comparison is triggered, which is often the case for academic self-concept. Academic self-concept is often measured with items like "I am good in math" (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005), whereby the word "good" is not further referenced. An absolute standard in the operational definition of a self-appraisal does not preclude that reference standards are evoked. In fact, academic self-concept is well known to be affected by social and dimensional references (Marsh, Seaton, et al., 2008; Möller & Marsh, 2013). Yet these references are not (always) explicitly triggered within the items. A *social* frame of reference would mean that within the item a comparison with some other individuals, for instance, with the class of students or all students of a university is triggered. One example would be an appraisal like "If you were to order all the students in your math class from the worst to the best in math, where would you put yourself?" (Eccles & Wigfield, 1995). In this item, an explicit comparison between students and their classmates is evoked. A *criterion-based* reference means that the self-appraisal is compared to a measurable behavior like achieving a

certain grade, solving a number of items correctly, or being able to perform a given task. Often, self-efficacy and JOLs are criterion based. Hacker et al. (2000), for instance, assessed JOLs by asking students how much percent of items they will solve correctly in an upcoming test. Finally, *internal* frames of reference trigger students to compare different aspects of their achievement. Therein, *temporal* comparisons would mean that students compare their current achievement to their achievement at another point in time. An example item could be "Compared to one year ago, I am good in math". In contrast, *dimensional* references would describe a comparison between different achievement domains or tasks. An example item could be "Some kids are better in one subject than in another. For example, you might be better in math than in reading. Compared to most of your other school subjects, how good are you in math?" (Wigfield & Eccles, 2000).

The *specificity level* refers to the type of achievement that is estimated and roughly ranges from a very specific to a very general achievement. The most specific level is *item*, which means that students are asked if they got a specific item in a given test correct (Schraw, Kuch, Gutierrez, & Richmond, 2014). A *task* would be more general and refers to a number of items or a number of steps that need to be solved. Examples would include the performance in a specific test in terms the percentage of correctly solved items (Hacker et al., 2000) or the performance in solving a specific complex problem (Baars, Vink, van Gog, de Bruin, & Paas, 2014). With *skill* we refer to a class of tasks that could range from "predict how changes in an environment will affect the survival of certain species" in science (Jansen, Scherer, et al., 2014) to more general "reading" or "writing" in a language (Arens & Jansen, 2016). Contrary to task, a skill does not need to refer to a specific situation, but to the skill in general or to a hypothetical situation where that skill would be applied. A *domain specific* operationalization refers to a broader domain where multiple skills are included. Most often, this description refers to academic domains like the school subjects "math" or "English" (Marsh & Shavelson, 1985). Finally, *general* refers to the broadest operationalization and might include appraisals of the academic achievement in general like "general academic self-concept" (Brunner et al., 2010) or beliefs about the own intelligence (Denissen, Schönbrodt, van Zalk, Meeus, & van Aken, 2011; Dufner et al., 2012).

Time orientation is the last dimension of the model and refers to the time for which the achievement is rated. Thus, a *past-oriented* self-appraisal describes achievement that lies in the past. For instance, postdications are past-oriented as they describe the belief of how good an

individual performed in a given test situation in the past (Hacker et al., 2000). On the contrary, *future-orientation* refers to achievement that has not yet happened. For instance, JOLs about an expected test performance can be categorized as future-oriented (Baars et al., 2014).

Following the model will be used in presenting academic self-concept and judgments of learning in more detail. In the following sections, both self-appraisals will first be described and then classified.

1.2.3 Academic Self-concept – the Marsh/Shavelson Model

Academic self-concept is one of the most prominent constructs in educational psychology (Marsh & Hau, 2003). Most modern research regarding academic self-concept is grounded in the reviews by Shavelson et al. (1976) as well as Marsh and Shavelson (1985). They define self-concept as "a person's perceptions of him- or herself" (Marsh & Shavelson, 1985, p. 107) and further outlined that self-concept is multidimensional and hierarchical. That is, self-concept applies to a wide range of life domains, including social life, sports, physical appearance, or academic competence. Hierarchical means, that both there is both a more general and a more specific level of self-concept. The more general level was originally described to be a summative evaluation of the whole self. A more specific level can be described with regard to *academic* self-concept, which refers to the self-perception of a person in the academic domain. General academic self-concept corresponds to general academic ability (Brunner et al., 2010; Marsh & Shavelson, 1985). On a more specific level, factors of academic self-concept correspond to ability and achievement in specific domains. Most self-concept research targets these domain-specific factors of academic self-concept, for instance, self-concept in math, English, or the sciences (Marsh & Craven, 2006).

Although the Marsh/Shavelson model of academic self-concept is well established, organization and structure of academic self-concept is subject to ongoing research. For instance, Brunner et al. (2010) argued for a bifactor model instead of a hierarchical model of academic self-concept. That is, general academic self-concept should be represented by a bifactor instead of a second-order factor. In another line of reasoning, Jansen, Schroeders, and Lüdtke (2014) demonstrated the validity of three separate science domains (i.e., biology, physics, chemistry) instead of one global science domain. Furthermore, their research indicated that the organization of academic self-concept in the sciences seemed to rely on the organization of school subjects

(Jansen, Schroeders, Lüdtke, & Pant, 2014): Students that were taught science subjects in an integrated, domain-general format in contrast to separate subjects for physics, chemistry, and biology had less correlated self-concepts in these domains. Recently, Arens and Jansen (2016) proposed that language self-concept might have sub-facets referring to reading, writing, listening, and speaking.

Notwithstanding this ongoing research, academic self-concept can be described using the dimensions of the classification model by Jansen & Lösch (in prep.) because it is measured in a relatively consistent way in a wide range of studies (but see below). The most prominent measure is the *self-description questionnaire* (e.g., Marsh, 1989), which also has a German version (Schwanzer et al., 2005). It is often measured in large-scale studies like PISA and other nationally representative samples (e.g., Marsh, Hau, Artelt, Baumert, & Peschar, 2006). Most often, academic self-concept is assessed with items like "I usually do well in mathematics" (Marsh et al., 2012). In this instance "well" is not further referenced and accordingly, academic self-concept is best described as measured with an absolute frame of reference. Yet, sometimes, comparisons are directly triggered with items like "Mathematics is one of my best subjects" (Marsh et al., 2006) or "Mathematics is harder for me than for many of my classmates" (Marsh et al., 2012). The specificity of academic self-concept can most often be said on the domain level, including school subjects like math or English, (Marsh, 1993). But in line with the hierarchical organization, academic self-concept is sometimes also measured on a general specificity level (Brunner et al., 2010; Brunner, Lüdtke, & Trautwein, 2008) or on a skill level, like reading or writing (Arens & Jansen, 2016). Finally, the time orientation is past-oriented, as academic self-concept refers to past achievement and experiences in that domain (Bong & Skaalvik, 2003).

Highly related to academic self-concept are success expectancies and ability beliefs stemming from expectancy-value theory (Eccles & Wigfield, 1995). It was stated that both constructs are "conceptually distinct" but "empirically these constructs are highly related" (Wigfield & Eccles, 2000). The major distinction between both constructs is probably the time orientation, with expectancies being future oriented. Yet in line with their empirical similarity, academic self-concept is often used interchangeably to expectancies (Guo et al., 2016; Guo, Marsh, Morin, Parker, & Kaur, 2015; Guo, Marsh, Parker, Morin, & Yeung, 2015; Guo, Parker, Marsh, & Morin, 2015; Nagengast et al., 2011; Trautwein et al., 2012).

1.2.4 Judgments of Learning and Postdictions

The theoretical foundation of judgments of learning (JOLs) and similar constructs like postdictions is given by the metamemory framework (Nelson & Narens, 1990). The metamemory framework describes self-regulated learning as a process consisting of an object level and a meta-level. The object level refers to the actual learning behavior or behavior in an achievement situation. The meta-level is described as instance that reflects about the object level. The meta-level has two important functions: monitoring the object level and regulating it according to the monitoring. According to the metamemory framework, self-appraisals including JOLs and postdictions can best be referred to as judgments of the meta-level about actions on the object level (Nelson & Narens, 1990). These monitoring judgments are described as pivotal aspects of students' self-regulated learning, because only accurate JOLs should be able to guide learning effectively (Bjork et al., 2013).

Whereas this description gives a theoretical guideline to interpret JOLs, the operationalization of JOLs is highly varying, as they are usually adapted to the specific task at hand. Thus, it is difficult to clearly categorize them within the classification model. Some examples may illustrate this variation. Hacker et al. (2000) asked university students before and after exams to estimate the percentage of items they got correct. In a more specific approach, Schraw et al. (2014) let participants work on several tests including several multiple-choice items. After each item, participants were asked whether they believed their answer was correct. Thiede, Anderson, and Therriault (2003) investigated students' metacomprehension, that is, they let students read several texts. After each text, students should indicate how well they believed that they understood the text on a likert scale. This judgment was later compared with students' performance in questions about the texts. As a last example, Baars et al. (2014) studied calibration in solving complex problems. Before and after working on complex problems, they asked students to judge how many steps of solving a problem they would be able to make.

Whereas these examples show some variation, they also highlight similarities regarding the classification model. Predominantly, JOLs are criterion referenced, that is, they refer to an objectively determined performance. Yet, regarding metacomprehension, also self-appraisals with an absolute reference might be asked for (Thiede et al., 2003). JOLs are often highly specific, referring to specific items (Schraw et al., 2014) or tasks in the form of whole tests or exams (Hacker et al., 2000). The time orientation is varying, with some studies investigating

JOLs that are future oriented, some investigating postdictions that are past-oriented, and some investigating both. Therein, both future- and past-oriented judgments are interpreted as results of the same ability to correctly monitor one's achievement and calibrate one's self-appraisals (Baars et al., 2014).

1.2.5 Reference Effects of Self-appraisals

A defining feature of self-appraisals is that they are necessarily made in comparison with certain references, as every evaluation can only be made in relation to a standard (Kahneman & Miller, 1986). As outlined by the classification model, self-appraisals may trigger references, but reference effects might arise without being triggered explicitly (Jansen & Lösch, in prep.). There is a vast literature on frame of reference effects for academic self-concept with a special focus on social and dimensional frames of reference (Möller & Marsh, 2013). But also self-appraisals like JOLs are made with regard to certain references often labeled as *cues* (Koriat, 1997) that are sometimes similar to social or dimensional reference effects. Following, I will describe social and dimensional reference effects with regard to several self-appraisals.

A social frame of reference means that students compare their own achievement with achievement of others to form their self-concept. This effect is represented by the big-fish-little-pod (BFLPE) effect (Marsh, Seaton, et al., 2008): Controlling for individual achievement, students in a class with higher average achievement have a lower academic self-concept than students in a class with lower average achievement. Overall, the BFLPE was repeatedly and consistently shown across cultures and domains (Marsh, 1987; Marsh & Hau, 2003; Nagengast & Marsh, 2012; Seaton, Marsh, & Craven, 2010).

Yet, studies that investigated the BFLPE for self-appraisals like self-efficacy found much weaker or no effects (Jansen, Scherer, et al., 2014; Marsh, Trautwein, et al., 2008; Parker et al., 2014). Zhao and Linderholm (2011) experimentally tested the effect of a social frame of reference on JOLs and postdictions. They provided individuals with the information that peers experienced a test either as difficult or as easy. Thereby, the information that peers found a test easy elicited more positive JOLs and postdictions than the information that peer found the test difficult. That is, the social frame of reference had the opposite effect to the BFLPE. That is, whereas there is a clear effect of a social frame of reference on academic self-concept, the case for other self-appraisals is less clear.

A similar case can be made for dimensional frames of reference, which are often tested within the internal/external frame of reference model (I/E model) for academic self-concept (Marsh, 1986). The I/E model describes the association between achievement and academic self-concept in a verbal and mathematical domain. This association is positive within domains but negative across domains. That is, verbal self-concept is positively associated with verbal achievement. But controlling for this association, verbal self-concept is negatively associated with math achievement (and analogously for math self-concept). This pattern leads to the paradoxical result that achievement is usually positively correlated across domains, but self-concept is not (Möller et al., 2009). Overall, the I/E model provides evidence that individuals perceive their ability in one domain in reference to their ability in a different domain. This was found across a broad range of studies that employed various methodologies including experimental and observational studies and so it led to the recent proposition of *dimensional comparison theory* (DCT; Möller & Marsh, 2013) that proposes dimensional comparisons as a pervasive feature of human appraisals.

DCT is explained in terms of its domain-specificity. According to the multidimensional model of self-concept, some domains are more similar to each other (e.g., math and physics) than some other domains (e.g., math and English). The dissimilarity of domains, in turn, results in dimensional comparison effects between achievement in these domains (Helm, Mueller-Kalthoff, Nagy, & Möller, 2016; Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2014, 2015). Thereby, dimensional comparison effects were not only found for academic self-concept, but also for a range of other domain-specific outcomes like academic emotions or interest (Arens, Möller, & Watermann, 2016; Goetz, Frenzel, Hall, & Pekrun, 2008; Schurtz, Pfost, Nagengast, & Artelt, 2014).

However, the I/E model was only found to a weaker extent for other self-appraisals. That is, studies that investigated the I/E model for self-efficacy only found weak or no effects (Jansen, Scherer, et al., 2014; Marsh et al., 2015; Möller et al., 2009; Parker et al., 2014). Thus, differences in the operationalization between self-concept and self-efficacy seem to moderate the I/E model.

Overall, frames of reference play a major role in the formation of academic self-concept, but their role in the formation of other self-appraisals is less clear. That is, reference effects might be moderated by dimensions of the classification model (Jansen & Lösch, in prep.). This

moderating effect was shown between academic self-concept and self-efficacy: For self-efficacy, reference effects were consistently weaker. However, to further corroborate this notion it is necessary to test reference effects on a broader range of self-appraisals.

1.3 Self-appraisals and Achievement

The defining core of self-appraisals is their reference to achievement, thus a logical question is in how far they are actually associated to achievement. In this section, the link between self-appraisals and achievement is described. I will start by describing calibration as a measure of the association between achievement and self-appraisals. Thereafter, I will describe the cross-sectional correlation between both constructs, the development of their association, and finally, their effect on each other. Throughout this chapter, several research traditions are compared. These traditions differ in the self-appraisals they refer to. That is, the different perspectives on calibration are paralleled by differences of self-appraisals in the classification model.

1.3.1 Operationalization of Calibration

The correspondence or fit between a self-appraisal and actual achievement is termed *calibration* (Alexander, 2013; de Bruin & van Gog, 2012; Keren, 1991).Calibration can be operationalized in several ways, but these measures are usually correlated with each other (Gutierrez, Schraw, Kuch, & Richmond, 2016; Maki, Shields, Wheeler, & Zacchilli, 2005; Schraw et al., 2014). A major distinction is between relative or absolute measures of calibration: relative measures assess calibration as some sort of correlation whereas absolute measures assess calibration as a difference score (Maki et al., 2005). Consequently, relative measures inform whether the rank order of achievement fits the rank order of self-appraisals, that is, it informs about the accuracy of an appraisal in relation to other appraisals. Contrary, absolute measures rather inform whether an individual was inaccurate and about the direction of that inaccuracy, that is, without the relation to other appraisals. Therein, *overestimation* can be defined as a positive absolute calibration and *underestimation* as a negative absolute calibration (Moore & Healy, 2008). Both relative and absolute calibration can be investigated as individual differences, yet to investigate individual differences in relative accuracy several achievement measures and several appraisals per individual are necessary (e.g., appraisals and test questions about multiple texts; Dunlosky & Lipko, 2007).

A number of constructs are associated with the concept of calibration. For instance, *self-enhancement* is often used interchangeably with overestimation. Yet self-enhancement has a broader meaning (Sedikides & Gregg, 2008). Self-enhancement can be defined as the motive to increase the positivity or decrease the negativity of one's self-concept (Leary, 2007). That is, self-enhancement is interpreted as a motivation to see or present oneself as positive, whereas overestimation is rather seen as the act of having an overly positive self-appraisal. Additionally, self-enhancement also describes more behaviors than just being not calibrated (Alicke & Sedikides, 2009) like the tendency of the majority of people to believe that they have more positive attributes than the average person (the *better-than-average-effect*; Guenther & Alicke, 2010).

1.3.2 Cross-sectional Association of Self-appraisals and Achievement

A first question regarding calibration is whether individuals' self-appraisals correspond to their actual achievement or ability on average and without considering temporal developements. Zell and Krizan (2014, p. 111) investigated this topic with the question "Have people insight into their abilities?". To answer this question is especially important against the background of some authors consenting that "self-assessments of skill and character are often flawed in substantive and systematic ways" (Dunning, Heath, & Suls, 2004, p. 69) or that "people's judgment accuracy is often poor" (Dunlosky & Lipko, 2007).

In their study, Zell and Krizan (2014) aggregated several meta-analyses on the correlation between self-appraisals and measures of objective achievement. That is, they focused on relative accuracy of self-appraisals. The authors collected 22 meta-analyses that investigated correlations between various measures of self-appraisals and various measures of objective achievement. The meta-analyses were set in a number of domains including education, sports, or vocational achievement. Overall, the study found a moderate effect of $M = .29$, $SD = 0.11$ with a range of .09 to .63. This result suggested that individuals are somewhat realistic, but they are far from perfect judges of their own capabilities.

However, the range of effect sizes indicated potential moderators. Indeed, Zell and Krizan (2014) found support for some moderators regarding operationalization of self-appraisals and measurement of achievement. For instance, they could replicate that more specific self-appraisals resulted in higher correlations with achievement and that past-oriented self-appraisals

resulted in higher correlations (Mabe & West, 1982; Valentine, DuBois, & Cooper, 2004). That is, there is evidence that at least two of the dimensions of the classification model (Jansen & Lösch, in prep.) are associated with the accuracy of the self-appraisals. Moreover, tasks that were more objective or tasks that were more familiar to individuals elicited higher accuracy of self-appraisals.

1.3.3 Development of Self-appraisals

The association between self-appraisals and achievement was also studied over time. Regarding this association, it is well known that self-appraisals (especially academic self-concept) decline through time in school. This trend was found in elementary school (Eccles, Wigfield, Harold, & Blumenfeld, 1993; Spinath & Steinmayr, 2008; Weidinger, Spinath, & Steinmayr, 2015; Wigfield et al., 1997) and in middle school (Wigfield, Eccles, Mac Iver, Reuman, & et al, 1991), often using cross-sequential designs (Marsh, 1989; Spinath & Spinath, 2005; Watt, 2004) spanning up to twelve years (Fredricks & Eccles, 2002; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002) and moreover, in Germany, Australia, and the United States (Nagy et al., 2010). There were some differences between domains in the strength of the decline, yet in all domains a decline was present.

Simultaneously to the development of self-appraisals, calibration seems to develop as well. In their seminal work, Stipek and Mac Iver (1989) describe that children shift from overconfident self-appraisals to accurate self-appraisals throughout their time in school. In a sense, the decrease in self-appraisals described above could at the same time be seen as an increase in calibration, with self-appraisals aligning with actual achievement. In line with this proposition, correlations between self-appraisals and achievement increased through years in school (Denissen, Zarrett, & Eccles, 2007). Further support for an increase in accuracy comes from studies that investigated peer appraisals of achievement. It was shown that peers increased in accuracy with age, too (Malloy, Yarlas, Montvilo, & Sugarman, 1996)

1.3.4 Longitudinal Interplay of Self-appraisals and Achievement

Achievement and self-appraisals do not develop independent of each other, but rather they affect each other over time. Yet, different educational traditions propose and show different

links between achievement and self-appraisals. That is, self-appraisals guide behavior on several levels of granularity - so far educational theories agree. However, research on different self-appraisals proposes different optima of the level of self-appraisals to achieve academic success. Educational psychology focusing on motivation highlights the effect of positive self-appraisals (Marsh & Hau, 2003). In contrast, educational psychology focusing on cognition highlights the positive effect of calibrated self-appraisals (Bol & Hacker, 2012; Dunlosky & Lipko, 2007). These positions will be contrasted in the following sections.

1.3.4.1 The Motivational Perspective

With the *motivational perspective*, mainly research on academic self-concept, self-efficacy and success expectancies is described. Especially academic self-concept is seen as one of the most important constructs in educational psychology due to its high predictive validity for a wide range of outcomes (Marsh & Hau, 2003). The association between academic self-concept and achievement is reciprocal: the effect of academic self-concept on subsequent achievement (controlling for prior achievement) and the reciprocal effect of achievement on subsequent academic self-concept (controlling for prior academic self-concept) is well documented in a number of reviews and meta-analyses (Huang, 2011; Marsh & Craven, 2006; Marsh & Martin, 2011; Valentine et al., 2004). But furthermore, academic self-concept predicted academic effort and engagement (Guo et al., 2016; Marsh et al., 2016; Trautwein, Lüdtke, Schnyder, & Niggli, 2006), career aspirations (Guo, Marsh, Parker, et al., 2015; Jansen, Scherer, et al., 2014; Nagengast et al., 2011; Nagengast & Marsh, 2012), and educational attainment (Guay, Larose, & Boivin, 2004; Guo, Marsh, Morin, et al., 2015; Marsh & O'Mara, 2008). Finally, academic self-concept was identified as a crucial predictor of academic choices, including course choice in schools or choice of university majors often in line with expectancy-value theory (Guo, Parker, et al., 2015; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Parker et al., 2012, 2014; Simpkins, Davis-Kean, & Eccles, 2006; Wang, 2012; Watt et al., 2012). Results for academic self-efficacy are similar to results for academic self-concept. Despite the conceptual differences of these two constructs, the general pattern is highly similar: the higher students' levels of self-efficacy, the more positive outcomes should be expected (Pajares, 1996; Usher & Pajares, 2008).

1.3.4.2 The Cognitive Perspective

The *cognitive perspective* mainly focuses on JOLs and related self-beliefs as well as on their calibration. Regarding JOLs as outcome, it is usually found that higher achieving students give higher JOLs or postdictions (e.g., Baars et al., 2014; Hacker et al., 2000). But what is more relevant in the cognitive perspective is the association between achievement and calibration. In a famous article, Kruger and Dunning (1999) proclaimed that students with lower achievement suffer a "double course of incompetence" meaning that they are "unskilled and unaware of it". That is, low achieving students overestimated their achievement and accordingly high achieving students underestimated their achievement. Their result was often replicated with different self-appraisals (Chiu & Klassen, 2009, 2010; de Bruin, Kok, Lobbestael, & de Grip, 2016; Hacker et al., 2000; Maki et al., 2005; Nietfeld & Schraw, 2002). However, Krueger and Mueller (2002) pointed out that the so called "unskilled and unaware effect" might be an statistical artifact: Self-appraisals would necessarily have a lower variance than achievement and accordingly they would regress to the mean (Fiedler & Unkelbach, 2014). One solution to control this regression effect would be a different operationalization of calibration (John & Robins, 1994), yet this was barely done in educational psychology.

JOLs were also found to predict behavior, but a different optimum is proposed than in the motivational perspective. Based on the notions of the metamemory framework (Nelson & Narens, 1990), metacognitive judgments guide learners choice of study tasks and study time (Metcalfe, 2009; Son & Metcalfe, 2000). Based on this premise, an accurate calibration is seen as essential for successful learning, for instance, to choose the right things to learn, to invest sufficient time to study, or to terminate study after a concept is understood (Alexander, 2013; Bol & Hacker, 2012; de Bruin & van Gog, 2012). In line with this theoretical expectation, there are a number of studies that demonstrated a successful increase in students' learning outcomes by interventions that targeted students' calibration (Dunlosky & Lipko, 2007). Several forms of training were able to improve calibration, for instance, providing learners with guidelines on how to reflect about learning material (Bol, Hacker, Walck, & Nunnery, 2012), showing learners video model examples where models displayed successful calibration (Kostons, van Gog, & Paas, 2012), or presenting material on overconfidence biases before making JOLs (Roelle, Schmidt, Buchau, & Berthold, 2016). Another successful strategy to improve calibration is to use generation strategies alongside learning, which require learners to produce summaries, keywords, or diagrams, that contain main ideas of a learning material (Bugg & McDaniel, 2012;

Fukaya, 2013; Redford, Thiede, Wiley, & Griffin, 2012; Thiede et al., 2003; van Loon, de Bruin, van Gog, van Merriënboer, & Dunlosky, 2014). Finally, calibration was improved by providing standards of how a task should have been solved correctly (Baars et al., 2014; Dunlosky & Rawson, 2012).

Beside these studies that used interventions to increase calibration and subsequently achievement, there were studies that demonstrated the link from calibration to achievement via feedback. Two studies (Försterling & Morgenstern, 2002; Kim, Chiu, & Zou, 2010) provided participants either with correct or incorrect feedback about their performance in a test. Correct feedback compared to incorrect feedback led to better calibration and to better subsequent achievement in a following test. Overall, studies in the tradition of the cognitive approach show that calibration in the learning process shapes control of study behavior and subsequently leads to better learning outcomes.

1.3.4.3 Comparison and Integration of Perspectives

These two traditions are highly similar in that they acknowledge the importance of self-appraisals as a guidance for subsequent behavior. But at the same time they differ in a number of characteristics including central elements of their design and their theoretical underpinnings.

Theoretically, the two traditions differ in their take on self-appraisals. On the one hand, self-appraisals are seen as the product of a cognitive estimation of achievement (Nelson & Narens, 1990). That is, they are interpreted as the result of a cognitive process, which tries to be as close as possible to the truth. Consequently, accuracy of self-appraisals gives an indicator of the ability to correctly perceive the own achievement. On the other hand, self-appraisals are conceptualized as a motivational characteristic that predicts behavior and educational choices (Eccles & Wigfield, 2002). Especially seen in the framework of expectancy-value theory the expectation to be able to achieve something will be followed by the impetus to do this action. In this regard, accuracy of a self-appraisal is not an integral part of the construct as it is in the cognitive perspective. Rather, self-appraisals are seen as a facilitator of a positive academic development. They are seen as important constructs in their own right, much more than from a cognitive perspective.

Studies on motivation often use large-scale data spanning several years, they are mostly observational, externally valid, and they mostly focus on individual differences in self-appraisals

or individual differences in development of self-appraisals. Studies on cognition often use smaller samples; they are conducted in the lab or in a controlled environment, rely on (experimental) intervention studies, and they use mixtures of within- and between-person designs. As a result of these differing designs, it is unclear whether results from one tradition generalize to the other.

So far, there have been few integrative approaches that combine paradigms or constructs from both approaches and moreover, they are not entirely coherent. On the one hand, Chiu and Klassen (2009, 2010) studied calibration of math and reading self-concept using data from PISA 2000 (OECD, 2002). They found that calibration of reading and math self-concept was positively associated to achievement and, furthermore, that especially low-achieving students tended to overestimate their achievement and high-achieving students tended to underestimate their achievement. Contrary to these results, no negative link from overestimations to other academic outcomes could be found in a set of other studies (Bouffard & Narciss, 2011; Bouffard, Vezeau, Roy, & Lengelé, 2011; Côté, Bouffard, & Vezeau, 2014; Dupeyrat, Escribe, Huet, & Régner, 2011; Gonida & Leondari, 2011). Rather, it appeared that underestimations might have harmful consequences for students' motivation, whereas both overconfidence and accurate self-appraisals might be equally adaptive. Finally, in a recent German study, a small positive effect of overestimations on subsequent achievement over and above prior achievement could be found (Praetorius, Kastens, Hartig, & Lipowsky, 2016).

Yet, the studies above have some limitations. The studies by Chiu and Klassen (2009, 2010) were cross-sectional and did not show longitudinal effects of calibration on achievement. Other studies did not investigate achievement as outcome (Bouffard et al., 2011; Gonida & Leondari, 2011). And finally, it might be problematic to use achievement in a double function, namely, to compute a measure of calibration and at the same time use it as outcome of calibration (Chiu & Klassen, 2009, 2010; Côté et al., 2014; Dupeyrat et al., 2011; Praetorius et al., 2016). This might be circular, especially with cross-sectional data. Overall, there is no clear picture regarding the effect of overestimations on students' achievement outside of small-scale learning.

1.4 Self-appraisals and Peer Appraisals

Closely related to self-appraisals of achievement are peer appraisals of achievement. Self- and other appraisals may be associated in multiple ways and models of person perception show the interrelatedness of the self and others (Kenny, 1994; Kenny & West, 2010; Kwan et al., 2004; McAbee & Connelly, 2016; Vazire, 2010).

Others' appraisals influence self-appraisals. For instance, expectancy-value theory stresses the influence of significant others on individuals self-appraisals (Wigfield & Eccles, 2000). Accordingly, others' appraisals may be a precursor of self-appraisals. Furthermore, others' appraisals have an effect on achievement mediated by self-appraisals.

Especially this latter result may point to another relation: Self-perception and perception by others may be two methods to measure the same trait. That is, both perception processes might work similar and correlates of appraisals from different perspectives might be similar (Kenny & West, 2010; Kwan et al., 2004). Consequently, self-appraisals and others' appraisals might be two ways to measure the same construct like in a multi-trait multi-method framework (Campbell & Fiske, 1959; McAbee & Connelly, 2016).

In the following section, I will subsequently describe both notions in more detail. Thereby, the conceptualization of self-appraisals in a multi-trait multi-method framework allows to derive a model that integrates self-appraisals, peer appraisals, achievement, and their connections in one representation.

1.4.1 Effects of Others' Appraisals of Achievement

There are several lines of research showing that others' appraisals affect self-appraisals. According to expectancy-value theory (Wigfield & Eccles, 2000), one source of self-appraisals are the beliefs and the feedback of significant others. So far, research has mainly focused on appraisals by parents and teachers, but peers were shown to be important as well.

The effect of parents' and teachers' appraisals on students' self-appraisals and achievement is well documented. Several meta-analyses found consistent positive effects of parents' appraisals on student achievement (Castro et al., 2015; Fan & Chen, 2001; Jeynes, 2007). Similarly, several studies found an effect of parents' appraisals on (the development of)

students' self-appraisals (Bleeker & Jacobs, 2004; Fredricks & Eccles, 2002; Frome & Eccles, 1998; Simpkins, Fredricks, & Eccles, 2012). Moreover, students' self-appraisal mediated the effect of parents' appraisals and achievement (Neuenschwander, Vida, Garrett, & Eccles, 2007). Similarly, teachers' appraisals of students' achievement (the *Pygmalion effect;* Rosenthal, 1994) had a positive effect on students' future achievement as documented in several meta-analyses (Jussim & Harber, 2005; Tenenbaum & Ruck, 2007). Importantly, Friedrich, Flunger, Nagengast, Jonkmann, and Trautwein (2015) found that the effect of teacher appraisals on achievement was mediated by students academic self-concept.

A similar pattern can be found for peers' appraisals of a student's achievement. Peers' appraisals of a target student's achievement are conceptualized as *peer academic reputation* (PAR; Gest, Domitrovich, & Welsh, 2005). PAR is usually measured as an aggregate of appraisals by classmates about a target student's achievement. Therein, PAR was often measured with regard to achievement across subjects in contrast to academic self-concept, which is often measured with regard to specific subjects.

There is only a relatively small number of studies investigating PAR, yet they draw a coherent picture. In a sample of Grade 3 students, Gest et al. (2005, 2008) found that PAR was associated with academic achievement, effort, and general academic self-concept. Importantly, they found that PAR predicted school grades over and above earlier grades over the course of two years (and similarly, that grades predicted PAR over and above earlier PAR). In another sample of at-risk students who scored below median achievement in a reading test, two studies investigated further effects of PAR. Hughes et al. (2009) found that PAR predicted teacher rated effort, teacher rated achievement in math and reading, and performance in a reading test (but not in a math test) one year later. For these effects, the authors controlled for prior values of all measure including prior achievement. Chen et al. (2010) added to these results by finding that PAR predicted achievement in a standardized test three years later, controlling for prior achievement as well as IQ and economic status. Therein, the effect of PAR on achievement was mediated by general academic self-concept and engagement. Finally, Denissen et al. (2011) found in university students that PAR predicted and was predicted by grades in exams in a cross-lagged model over the course of eight months. Moreover, PAR predicted student dropout, whereas self-rated ability did not.

Overall, there is evidence that others' appraisals of a target student's achievement predict that student's future achievement and self-appraisals. That is, there appears to be a common core of appraisals by different perspectives (self, parents, teachers, and peers) that share a positive effect on future achievement.

1.4.2 Appraisals by the Self and Others in a Multi-rater Framework

Models of person perception explicitly target overlap in constructs from different perspectives and these models may be applied to appraisals of achievement. Following I present two approaches that have different origins, but result in similar conceptualizations of self-appraisals: a factor analytic approach and a componential approach. I will describe these models, how accuracy can be investigated within these models, and finally, how it is possible to integrate objective achievement within these models.

1.4.2.1 A Factor Analytic Representation of Appraisals

In personality psychology, person perception and the accuracy of person perception are important fields of research mainly because there is no easy way to assess personality objectively. Although it is assumed that there is a "true" personality, it is hard to measure (Funder, 1995). Accordingly, personality measures need to be validated against each other (McAbee & Connelly, 2016). Moreover, to come closer to a valid measurement of personality, multiple measures of personality are employed that are combined in a multi-trait multi-method (MTMM) design (Campbell & Fiske, 1959). In an MTMM design, each measure is assumed to consist of differences due to the measured trait and differences due to the method that is used to measure the trait (Eid, Geiser, & Koch, 2016). In this regard, different raters of the same trait can be conceptualized as multiple methods to measure one target trait. Following this conception, the self and peers might be seen as two different raters of the same trait. That is, a personality trait can be defined as a latent factor underlying multiple measures of the same trait while controlling for variance due to the measurement method. Following this definition, a given appraisal can be described as consisting of a meaningful "substance" of the trait and the "style" of the rater that is unrelated to the actual trait (Leising, Scherbaum, Locke, & Zimmermann, 2015; McCrae & Costa, 1983).

The *trait-identity-reputation (TRI) model* (see Figure 1.2; McAbee & Connelly, 2016) further developed the conceptualization of appraisals in a MTMM design. The major contribution of the TRI model is that it offers interpretations for method (i.e., rater) factors identified with multiple-rater data. The *trait* represents common variance that is shared by all raters and might thus be the most valid representation of the target trait. The *identity* is the variance in self-appraisals that is not explained by the trait and represents an individual's idiosyncratic self-perception. Thus, the identity factor can be seen as representation of self-enhancement (Leary, 2007; McAbee & Connelly, 2016; Sedikides & Gregg, 2008). Finally, *reputation* is variance in others' appraisals that is not explained by the trait and may represent the impression that a target elicits in others. The TRI model acknowledges that both the self and others can have insight into an underlying trait but at the same time that raters with different perspectives may have unique insights (Vazire, 2010). Accordingly, variance in appraisals is attributed to at least two sources: the trait and a perceiver-specific style of appraisal. Moreover, the TRI model offers a way to investigate the unique effect of each of those factors on a criterion by regressing criteria on the factor scores (McAbee & Connelly, 2016).

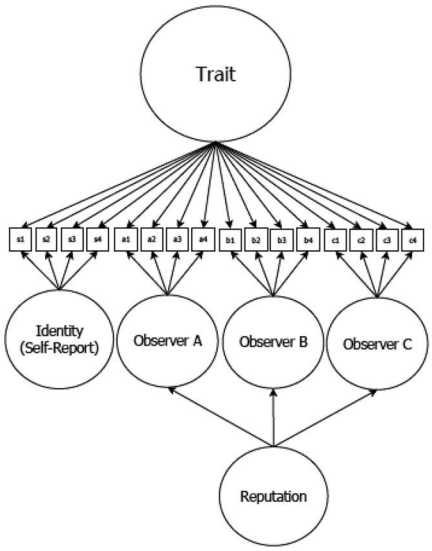

Figure 1.2. The trait-identity-reputation (TRI) model according to McAbee and Connelly (2016).

1.4.2.2 A Componential Representation of Appraisals

A componential view on appraisals - that was originally introduced by Cronbach (1955) - leads to a similar decomposition of interpersonal appraisals. One of the most famous

componential models is the *social relations model* (SRM; Kenny, 1994; Kenny, Kashy, & Cook, 2006). The SRM decomposes variance of interpersonal appraisals into three components: *perceiver effect*, *target effect* and *relationship effect*. That is, a positive appraisal may result because of either one of those components. The prototypical design to investigate the SRM is a round-robin assessment where each group member rates every other group member on one or more traits (see Figure 1.3 for an illustration). The *perceiver effect* describes a tendency to rate peers in general rather mildly or harshly. The perceiver effect may be seen as a mixture of a methodological response style and a positive perception of others in general (Kenny, 1994). This individual rating tendencies seems to be relatively stable (Srivastava, Guglielmo, & Beer, 2010; Wood, Harms, & Vazire, 2010). That is, a perceiver that tends to rate targets very positive will probably do so across traits and across points in time. The *target effect* describes a tendency to be rated as positive or negative on a given trait. That is, the target effect may be seen as an estimate for a trait as perceived by others (Kenny, 1994; Kenny et al., 2006). Finally, the *relationship effect* describes the unique tendency of one perceiver to rate one specific target especially positive or negative. That is, the relationship effect describes the idiosyncratic view within one specific dyad. In designs where only one item is used to measure appraisals, the relationship effect is confounded with measurement error.

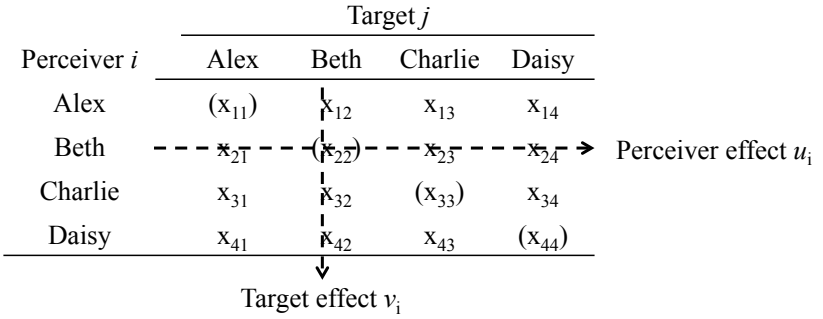

Figure 1.3. Illustration of an univariate round robin design with four participants (Alex, Beth, Charlie, and Daisy). The appraisal of member *i* about member *j* is represented by x_{ij}. The self-appraisal of member *i* is represented by (x_{ii}). The arrows indicate effects that are estimated in the social relations model (Kenny, 1994): the perceiver or row effects u_i (i.e., how student *i* tends to rate others) and the target or column effects v_j (i.e., how others rate student *i*).

Within the framework of the SRM, similar ideas are put forward as in the TRI model. In the SRM, self-appraisals are conceptualized as a peer appraisal for which the self is both the

perceiver and the target (Kenny, 1994). This idea is backed up in the meta-analysis by Kenny and West (2010) who concluded that "there are several pieces of evidence that self-perception and the perception of others are not very different" (p. 208). That is, similar processes and mechanisms might be assumed to contribute to self- and other appraisals. With this argumentation, Kenny (1994) proposed that self-appraisals can also be assumed to be a function of target and perceiver effects, but these effects might not be weighted like they are in appraisals about others. Note that the idea formulated in the SRM of the target effect underlying both self- and other appraisals is almost identical to the idea formulated in the TRI model of a common trait factor underlying self- and other appraisals. Similarly, based on the SRM the interpersonal approach to self-enhancement was proposed (Kwan et al., 2004). In the interpersonal approach, self-enhancement is conceptualized as the part of a self-appraisal that is not explained by target and perceiver effect. This operationalization of self-enhancement is almost identical to the identity factor in the TRI model.

There are minor differences between the TRI model and the SRM. First, the TRI model so far focuses on one target that is rated by several perceivers, whereas the prototypical SRM design is a round-robin design where every participant is both target and perceiver of appraisals. In line with this focus, the TRI model uses a factor analytic operationalization, whereas the SRM requires more specialized models (Lüdtke, Robitzsch, Kenny, & Trautwein, 2013; Nestler, 2015; Schönbrodt, Back, & Schmukle, 2012). As a second consequence, the TRI model does not explicitly model the perceiver effect. That is, the identity factor in the TRI model is confounded with an individual's tendency to rate others in specific ways.

1.4.2.3 Accuracy of Appraisals

In general, accuracy refers to the degree that a appraisal about a person's trait refers to that person's true level of a trait (Funder, 1995) or, put otherwise, "is Z's view of H correct?" (Kenny, 1994, p. 6), where Z is the perceiver and H the target of a appraisal. Yet, as personality can not easily be measured objectively, a general idea is to validate personality measures from different sources (e.g. different observers, observations, self-appraisals) against each other (Campbell & Fiske, 1959; McAbee & Connelly, 2016).

Funder's *realistic accuracy model* (Funder, 1995, 2012) follows this validation approach across measures and outlines three ways to estimate the accuracy of appraisals. These are *self-*

other agreement, *other-other agreement*, and *behavioral prediction* (Funder, 2012). These three measures describe different correlations: self-other agreement is the correlation between self-appraisals and appraisals by others, other-other agreement is the correlation between appraisals by different others, and behavioral prediction is the correlation between appraisals and more objective covariates.

The TRI model (McAbee & Connelly, 2016) accuracy can be rephrased in a highly similar way as outlined by Funder (2012). A measure of agreement between the self and multiple others is the explained variance by the trait factor. A measure of agreement among perceivers is the variance that is explained by the reputation factor. And behavioral prediction can be investigated as the proportion of variance that is explained in a criterion by the identity, reputation, and trait factor (McAbee & Connelly, 2016). In sum, both the realistic accuracy model and the TRI model allow similar inspections of accuracy, but the TRI model allows for a more detailed investigation about the unique contributions in accuracy by the multiple perspectives.

1.4.2.4 Applying Models of Person Perception to Achievement Perception

In principle, models of person perception like the TRI model and the SRM can be applied to appraisal of achievement. But the important difference is that achievement can be measured more objectively. That is, achievement tests are widely accepted as valid and reliable measure of students' abilities and skills in a given domain (e.g., OECD, 2012). Similarly, school grades may represent a valuable measure of students' achievement. Thus, it needs to be outlined how this measure may be integrated in models of person perception.

The adaptation of TRI model and SRM depend on the conceptualization of the latent trait that is measured with appraisals of achievement. One could think of two alternatives. The first one would be that achievement tests, self-appraisals and peers' appraisals are all different methods to measure achievement. In this conceptualization, the achievement measure would be defined as equal to the latent trait (see Figure 1.4a), because the achievement measure (e.g., a standardized test) is widely accepted as "true measure" of students' achievement. The second approach would be to conceptualize self-appraisals and peer appraisals as different methods to measure *perceived* achievement. This conceptualization would assume that multiple measures of perceived achievement (e.g., self appraisals, peer appraisals) share characteristics that they do

not share with objective achievement. In this approach, achievement would be regarded as covariate in the model (see Figure 1.4b).

Depending on the question at hand, either approach may be more useful. In most cases, conceptualizing achievement as equal to the latent trait of might be more useful. The first approach directly displays an estimate of accuracy as the path coefficient from the achievement to the appraisals (as outlined above). That is, calibration is one part of the model and can be compared to the accuracy of peers. But although the first approach is easily implemented in a factor analytic conceptualization, it is not as straightforward to handle in the SRM framework. Additionally, in the first approach, the correlation between achievement and self-enhancement is per definition zero, as self-enhancement is defined as residual of the self-appraisal regressed on achievement (with the same being true for reputation). Moreover, the second approach shows the association between a shared perception and objective achievement directly, which might inform about how accurate achievement appraisals may be. Overall, the second approach may prove more useful to clarify the interrelation between self-enhancement and achievement, and moreover, to identify the association between objective achievement on the one hand and the appraisal of achievement on the other hand.

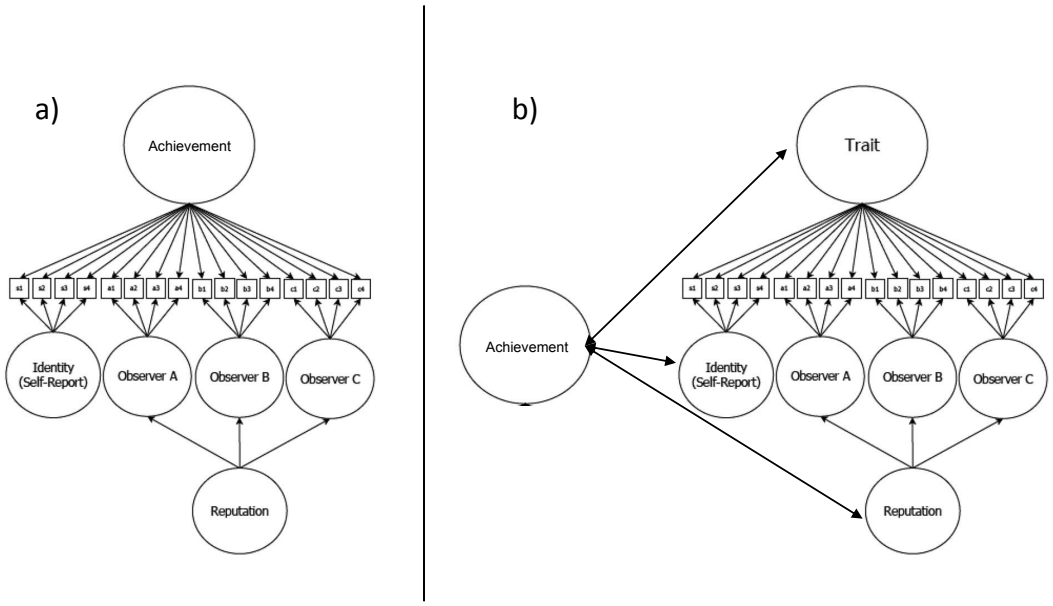

Figure 1.4. The trait-identity-reputation (TRI) model according to McAbee and Connelly (2016) in its adapted form to include a measure of objective achievement.

However, the adaptation of the TRI model to appraisal of achievement is built on the notion that self-appraisals and peer appraisals have overlapping variance. Both constructs were reciprocally related (Gest et al., 2005, 2008) but the amount of shared variance was not investigated so far. Furthermore, self-appraisals and peer appraisals should have similar correlates. Especially, both should be associated with achievement, which is supported in several studies (Chen et al., 2010; Denissen et al., 2011; Gest et al., 2008; Hughes et al., 2009). But moreover, to be seen as measures of the same constructs, they should also be similar processes involved in their formation. There are reasons to believe this (Kenny & West, 2010), but so far peer appraisals were studied in only few studies. So, little is known about predictors of PAR and whether they are similar to predictors of self-appraisals. Overall, to judge whether the TRI model of achievement appraisals operationalization is valid, these notions need to be clarified.

1.5 Outline of Studies

Students' self-appraisals guide behavior in different ways through small-scale learning and large-scale decisions. Accordingly, they are a focus of many educational theories (Marsh & Shavelson, 1985; Nelson & Narens, 1990; Wigfield & Eccles, 2000). On the one hand, they are highly associated with achievement, for instance, in terms of reciprocal associations (Marsh & Craven, 2006). On the other hand, self-appraisals are associated with peer appraisals (Gest et al., 2005). But also, peer appraisals are reciprocally associated with achievement (Gest et al., 2008). Overall, these three constructs are connected in a triangular shape. Yet several of these connections are not clear.

First, research traditions differ in self-appraisals they use. These differences can be described using the three dimensions *frame of reference*, *specificity*, and *time orientation* introduced in the classification model (Jansen & Lösch, in prep.). Potentially, differences of self-beliefs in these dimensions lead to differing correlations with achievement and to differing reference effects in their formation. Yet evidence for this proposal is limited, as studies that used different self-beliefs also differed markedly in several design features. In order to have a more integrated understanding of self-appraisals, they need to be compared more systematically.

Second, calibration describes the connection between self-appraisals and achievement and is an important predictor of future achievement (de Bruin & van Gog, 2012). However, calibration was mainly studied in controlled, small-scale settings. Moreover, studies that investigated calibration in more externally valid settings had several limitations (Chiu & Klassen, 2010). Accordingly, it is unclear whether calibration research has implications for the actual school context. That is, it is unclear how calibrated students are and what effect calibration might have.

Third, self-appraisals and peer appraisals of achievement may be described as different methods to measure the same construct. Both are reciprocally associated with achievement (Denissen et al., 2011; Gest et al., 2008; Marsh & Craven, 2006). So, based on models of person perception they might be regarded as measures of the same perceived achievement (Kenny, 1994; Kwan et al., 2004; McAbee & Connelly, 2016). Yet, these models have not been applied to appraisals of achievement. So, their validity needs to be examined empirically in order to support this connection. One would suppose that similar processes contribute to their formation.

Especially reference effects are a highly important predictor of self-appraisals that might also contribute to the formation of peer appraisals. Yet, no study has investigated this idea so far.

The present thesis addresses these limitations of prior research in three studies. All three studies target aspects of the triangular conception of self-appraisals, peer appraisals, and achievement. Thereby, the studies integrate psychological theories and models from the fields of education, motivation, cognition, and personality.

Study 1 targeted the connection between self-appraisals and achievement by investigating calibration in the school context. Therein, postdictions were used as self-appraisal, which are a usual measure in calibration research (Baars et al., 2014; Hacker et al., 2000). In this way, we tested whether results from calibration research generalize to the actual school context. Moreover, using postdictions allowed to test expectations from the classification model (Jansen & Lösch, in prep.). As postdiction differ in *specificity* and *frame of reference* from academic self-concept, we could test whether these differences resulted in different features. Overall, the study investigated the calibration of students' postdictions in a high-stakes exam, predictors of this calibration, as well as the domain-specificity of calibration. These topics were addressed in a large sample of 2,164 high school students that postdicted their grades in Abitur exams in English and math.

Study 2 targeted the whole triangle by investigating the overlap of self-appraisals, peer appraisals, as well as their shared association with achievement. By using a round-robin design of appraisals, the interpersonal approach (Kwan et al., 2004) was used to identify overlapping variance and to test whether this overlap would predict future achievement. Therein, this study may show the usefulness of the SRM for educational psychology. Furthermore, we could address the proposition of calibration research that an overestimation of achievement would result in lower future achievement. We analyzed self-appraisals and peer appraisals in math of 1,549 Grade 5 and Grade 8 students and tested how they were associated with current and future achievement.

Study 3 focused on achievement in English and math as predictor of self-appraisals and peer appraisals in English and math. Therein, two ideas were brought together. On the one hand, DCT would predict that comparisons between achievement in two domains predict peer appraisals in these domains. Similarly, according to the TRI model, similar processes should contribute to formation of self-appraisals and peer appraisals. Thus, we tested this notion and

investigated whether not only self-appraisals but also peer appraisals would be formed according to the I/E model. In a sample of 850 Grade 7 students, self-appraisals and peer appraisals were investigated within a bivariate SRM to test effects of the I/E model on peer appraisals.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. http://doi.org/10.1016/j.learninstruc.2012.10.003

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48. http://doi.org/10.1080/10463280802613866

Arens, A. K., & Jansen, M. (2016). Self-concepts in reading, writing, listening, and speaking: A multidimensional and hierarchical structure and its generalizability across native and foreign languages. *Journal of Educational Psychology*, *108*(5), 646–664. http://doi.org/10.1037/edu0000081

Arens, A. K., Möller, J., & Watermann, R. (2016). Extending the internal/external frame of reference model to social studies: Self-concept and achievement in history and politics. *Learning and Individual Differences*, *51*, 91–99. http://doi.org/10.1016/j.lindif.2016.08.044

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, *33*, 92–107. http://doi.org/10.1016/j.learninstruc.2014.04.004

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, *64*(1), 417–444. http://doi.org/10.1146/annurev-psych-113011-143823

Blanton, H., Buunk, B. P., Gibbons, F. X., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*(3), 420–430. http://doi.org/10.1037/0022-3514.76.3.420

Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in Math and Science: Do Mothers' Beliefs Matter 12 Years Later? *Journal of Educational Psychology*, *96*(1), 97–109. http://doi.org/10.1037/0022-0663.96.1.97

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215. http://doi.org/10.1037/0033-2909.117.2.187

Bol, L., & Hacker, D. J. (2012). Calibration research: where do we go from here? *Frontiers in Psychology*, *3*(July), 229. http://doi.org/10.3389/fpsyg.2012.00229

Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. a. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*, *37*(4), 280–287. http://doi.org/10.1016/j.cedpsych.2012.02.004

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

Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research*, *50*(4), 205–208. http://doi.org/10.1016/j.ijer.2011.08.001

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*(4), 221–229. http://doi.org/10.1016/j.ijer.2011.08.003

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. http://doi.org/10.1037/a0019644

Brunner, M., Lüdtke, O., & Trautwein, U. (2008). The Internal/External Frame of Reference Model Revisited: Incorporating General Cognitive Ability and General Academic Self-Concept. *Multivariate Behavioral Research*, *43*(1), 137–172. http://doi.org/10.1080/00273170701836737

Bugg, J. M., & McDaniel, M. a. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104*(4), 922–931. http://doi.org/10.1037/a0028661

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. http://doi.org/10.1037/h0046016

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-

analysis. *Educational Research Review*, *14*, 33–46. http://doi.org/10.1016/j.edurev.2015.01.002

Chen, Q., Hughes, J. N., Liew, J., & Kwok, O.-M. (2010). Joint contributions of peer acceptance and peer academic reputation to achievement in academically at-risk children: Mediating processes. *Journal of Applied Developmental Psychology*, *31*(6), 448–459. http://doi.org/10.1016/j.appdev.2010.09.001

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. http://doi.org/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. http://doi.org/10.1016/j.learninstruc.2008.11.002

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin*, *116*(1), 3–20. http://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*(6), 1152–1162. http://doi.org/10.1037/0022-3514.68.6.1152

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*, *8*, 415–434. http://doi.org/10.1111/bjep.12045

Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, *52*(3), 177–193. http://doi.org/10.1037/h0044919

de Bruin, A. B. H., Kok, E. M., Lobbestael, J., & de Grip, A. (2016). The impact of an online tool for monitoring and regulating learning at university: overconfidence, learning strategy, and personality. *Metacognition and Learning*. http://doi.org/10.1007/s11409-016-9159-5

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. http://doi.org/10.1016/j.learninstruc.2012.01.003

Denissen, J. J. A., Schönbrodt, F. D., van Zalk, M., Meeus, W. H. J., & van Aken, M. A. G. (2011). Antecedents and consequences of peer-rated intelligence. *European Journal of Personality*, *25*(2), 108–119. http://doi.org/10.1002/per.799

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I Like to Do It, I'm Able, and I Know I Am: Longitudinal Couplings Between Domain-Specific Achievement, Self-Concept, and Interest. *Child Development*, *78*(2), 430–447. http://doi.org/10.1111/j.1467-8624.2007.01007.x

Dufner, M., Denissen, J. J. A., 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*(3), 537–572. JOUR. http://doi.org/10.1111/j.1467-6494.2011.00742.x

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*(4), 228–232. http://doi.org/10.1111/j.1467-8721.2007.00509.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. http://doi.org/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. http://doi.org/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. http://doi.org/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. http://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. http://doi.org/10.1177/0146167295213003

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(1), 109–132. http://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., 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. http://doi.org/10.2307/1131221

Eid, M., Geiser, C., & Koch, T. (2016). Measuring Method Effects: From Traditional to Design-Oriented Approaches. *Current Directions in Psychological Science*, *25*(4), 275–280. http://doi.org/10.1177/0963721416649624

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, *13*(1), 1–22. http://doi.org/10.1023/A:1009048817385

Fiedler, K., & Unkelbach, C. (2014). Regressive Judgment: Implications of a Universal Property of the Empirical World. *Current Directions in Psychological Science*, *23*(5), 361–367. http://doi.org/10.1177/0963721414546330

Försterling, F., & Morgenstern, M. (2002). Accuracy of self-assessment and task performance: Does it pay to know the truth? *Journal of Educational Psychology*, *94*(3), 576–585. http://doi.org/10.1037//0022-0663.94.3.576

Fredricks, J. a, & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 519–533. http://doi.org/10.1037/0012-1649.38.4.519

Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology*, *41*, 1–12. http://doi.org/10.1016/j.cedpsych.2014.10.006

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology*, *74*(2), 435–452. http://doi.org/10.1037/0022-3514.74.2.435

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*(1), 1–18. http://doi.org/10.1007/s11409-012-9093-0

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. http://doi.org/10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological*

Science, *21*(3), 177–182. http://doi.org/10.1177/0963721412445309

Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology*, *97*(3), 337–346. http://doi.org/10.1037/0022-0663.97.3.337

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology*, *44*(3), 625–636. http://doi.org/10.1037/0012-1649.44.3.625

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33*(1), 9–33. http://doi.org/10.1016/j.cedpsych.2006.12.002

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. http://doi.org/10.1016/j.ijer.2011.08.002

Guay, F., Larose, S., & Boivin, M. (2004). Academic Self-concept and Educational Attainment Level: A Ten-year Longitudinal Study. *Self and Identity*, *3*(1), 53–68. http://doi.org/10.1080/13576500342000040

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better-than-average effect. *Journal of Personality and Social Psychology*, *99*(5), 755–770. http://doi.org/10.1037/a0020959

Guo, J., Marsh, H. W., Morin, A. J. S., Parker, P., & Kaur, G. (2015). Directionality of the Associations of High School Expectancy-Value, Aspirations and Attainment over Eight Years: A Multiwave, Longitudinal study. *American Educational Research Journal*, *52*(2), 371–402. http://doi.org/10.3102/0002831214565786

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). 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://doi.org/10.1016/j.lindif.2015.01.008

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., … Trautwein, U. (2016). Probing the Unique Contributions of Self-Concept, Task Values, and Their Interactions Using Multiple Value Facets and Multiple Academic Outcomes. *AERA Open*,

2(1), 1–20. http://doi.org/10.1177/2332858415626884

Guo, J., Parker, P. D., Marsh, H. W., & Morin, A. J. S. (2015). Achievement, motivation, and educational choices: A longitudinal study of expectancy and value using a multiplicative perspective. *Developmental Psychology*, *51*(8), 1163–1176. http://doi.org/10.1037/a0039440

Gutierrez, A. P., Schraw, G., Kuch, F., & Richmond, A. S. (2016). A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors. *Learning and Instruction*, *44*, 1–10. http://doi.org/10.1016/j.learninstruc.2016.02.006

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. a. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. http://doi.org/10.1037//0022-0663.92.1.160

Helm, F., Mueller-Kalthoff, H., Nagy, N., & Möller, J. (2016). Dimensional Comparison Theory: Perceived Subject Similarity Impacts on Students Self-Concepts. *AERA Open*, *2*(2), 1–9. http://doi.org/10.1177/2332858416650624

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. http://doi.org/10.1016/j.jsp.2011.07.001

Hughes, J. N., Dyer, N., Luo, W., & Kwok, O.-M. (2009). Effects of peer academic reputation on achievement in academically at-risk elementary students. *Journal of Applied Developmental Psychology*, *30*(2), 182–194. http://doi.org/10.1016/j.appdev.2008.12.008

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. http://doi.org/10.1111/1467-8624.00421

James, W. (1890). The consciousness of self. In *The principles of psychology, Vol I.* (pp. 291–401). New York, NY, US: Henry Holt and Co, Inc. http://doi.org/10.1037/10538-010

Jansen, M., & Lösch, T. (in preparation). Integrating perspectives on the self-appraisal of academic abilities: A classification model.

Jansen, M., Scherer, R., & Schroeders, U. (2014). Students' self-concept and self-efficacy in the sciences: differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24. http://doi.org/10.1016/j.cedpsych.2014.11.002

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, *30*, 11–21. http://doi.org/10.1016/j.lindif.2013.12.003

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and Assimilation Effects of Dimensional Comparisons in Five Subjects: An Extension of the I/E Model. *Journal of Educational Psychology*, Advance online publication. http://doi.org/10.1037/edu0000021

Jansen, M., Schroeders, U., Lüdtke, O., & Pant, H. A. (2014). Interdisziplinäre beschulung und die struktur des akademischen selbstkonzepts in den naturwissenschaftlichen fächern. *Zeitschrift Fur Padagogische Psychologie*, *28*(1–2), 43–49. http://doi.org/10.1024/1010-0652/a000120

Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education*, *42*(1), 82–110. http://doi.org/10.1177/0042085906293818

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*(1), 206–219. http://doi.org/10.1037/0022-3514.66.1.206

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. http://doi.org/10.1207/s15327957pspr0902_3

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153. http://doi.org/10.1037//0033-295X.93.2.136

Kelly, T. L. (1927). *Interpretation of educational measurements*. Oxford, England: World Book Co.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self-and other perception: A meta-analysis. *Personality and Social Psychology Review*, *14*(2), 196–213. http://doi.org/10.1177/1088868309353414

Keren, G. (1991). Calibration and probability judgements: Conceptual and methodological issues. *Acta Psychologica*, *77*(3), 217–273. http://doi.org/10.1016/0001-6918(91)90036-Y

Kim, Y.-H., Chiu, C.-Y., & 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. http://doi.org/10.1037/a0020555

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. http://doi.org/10.1037//0096-3445.126.4.349

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, *22*(2), 121–132. http://doi.org/10.1016/j.learninstruc.2011.08.004

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality & Social Psychology*, *82*(2), 180–188. http://doi.org/10.1037//0022-3514.82.2.180

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. http://doi.org/10.1037/0022-3514.77.6.1121

Kuyper, H., & Dijkstra, P. (2009). Better-than-average effects in secondary education: a 3-year follow-up. *Educational Research and Evaluation*, *15*(November 2014), 167–184. http://doi.org/10.1080/13803610902804416

Kuyper, H., Dijkstra, P., Buunk, A. P., & van der Werf, M. P. C. (2011). Social comparisons in the classroom: An investigation of the better than average effect among secondary school children. *Journal of School Psychology*, *49*(1), 25–53. http://doi.org/10.1016/j.jsp.2010.10.002

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*(1), 94–110. http://doi.org/10.1037/0033-295X.111.1.94

Kwang, T., & Swann, W. B. (2010). Do People Embrace Praise Even When They Feel Unworthy? A Review of Critical Tests of Self-Enhancement Versus Self-Verification.

Personality and Social Psychology Review, *14*(3), 263–280. http://doi.org/10.1177/1088868310365876

Leary, M. R. (2004). Editorial: What Is the Self? A Plea for Clarity. *Self and Identity*, *3*(1), 1–3. http://doi.org/10.1080/13576500342000004

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology*, *58*(1), 317–344. http://doi.org/10.1146/annurev.psych.58.110405.085658

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In *Advances in Experimental Social Psychology* (Vol. 32, pp. 1–62). http://doi.org/10.1016/S0065-2601(00)80003-9

Leary, M. R., & Tangney, J. P. (2003). The self as a Organizing Construct in the Behavioral and Social Sciences. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of Self and Identity* (pp. 3–14). New York: Guilford.

Leising, D., Scherbaum, S., Locke, K. D., & Zimmermann, J. (2015). A model of "substance" and "evaluation" in person judgments. *Journal of Research in Personality*, *57*, 61–71. http://doi.org/10.1016/j.jrp.2015.04.002

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. http://doi.org/10.1037/a0029252

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*(3), 280–296. http://doi.org/10.1037/0021-9010.67.3.280

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*(4), 723–731. http://doi.org/10.1037/0022-0663.97.4.7238

Malloy, T. E., Yarlas, A., Montvilo, R. K., & Sugarman, D. B. (1996). Agreement and accuracy in children's interpersonal perception: A social relations analysis. *Journal of Personality and Social Psychology*, *71*(4), 692–702. http://doi.org/10.1037/0022-3514.71.4.692

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. http://doi.org/10.3102/00028312023001129

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of*

Educational Psychology, *79*(3), 280–295. http://doi.org/10.1037//0022-0663.79.3.280

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, *81*(3), 417–430. http://doi.org/10.1037/0022-0663.81.3.417

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*(4), 841–860. http://doi.org/10.3102/00028312030004841

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., … Parker, P. (2012). Factorial, Convergent, and Discriminant Validity of TIMSS Math and Science Motivation Measures: A Comparison of Arab and Anglo-Saxon Countries. *Journal of Educational Psychology*, *105*(1), 108–128. http://doi.org/10.1037/a0029907

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. http://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Craven, R. G., John, W., Debus, R. L., & Hinkley, J. W. (2003). Evaluation of the big-two-factor theory of academic motivation orientations: An evaluation of jingle-jangle fallacies. *Multivariate Behavioral Research*, *38*(September 2013), 37–41. http://doi.org/10.1207/S15327906MBR3802

Marsh, H. W., & Hau, K.-T. (2003). Big-Fish--Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. http://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's Brief Self-Report Measure of Educational Psychology's Most Useful Affective Constructs: Cross-Cultural, Psychometric Comparisons Across 25 Countries. *International Journal of Testing*, *6*(4), 311–360. http://doi.org/10.1207/s15327574ijt0604_1

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology*, *39*(4), 326–341. http://doi.org/10.1016/j.cedpsych.2014.08.003

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

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. http://doi.org/10.1348/000709910X503501

Marsh, H. W., & O'Mara, A. (2008). Reciprocal Effects Between Academic Self-Concept, Self-Esteem, Achievement, and Attainment Over Seven Adolescent Years: Unidimensional and Multidimensional Perspectives of Self-Concept. *Personality and Social Psychology Bulletin*, *34*(4), 542–552. http://doi.org/10.1177/0146167207312313

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: Developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology*, *52*(8), 1273–1290. http://doi.org/10.1037/dev0000146

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. http://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. http://doi.org/10.1207/s15326985ep2003_1

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. http://doi.org/10.1037/0022-0663.100.3.510

McAbee, S. T., & Connelly, B. S. (2016). A Multi-Rater Framework for Studying Personality: The Trait-Reputation-Identity Model. *Psychological Review*. http://doi.org/10.1037/rev0000035

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, *51*(6), 882–888. http://doi.org/10.1037/0022-006X.51.6.882

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in*

Psychological Science, *18*(3), 159–163. http://doi.org/10.1111/j.1467-8721.2009.01628.x

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. http://doi.org/10.1037/a0032459

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. http://doi.org/10.3102/0034654309337522

Moore, D. a, & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. http://doi.org/10.1037/0033-295X.115.2.502

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053. http://doi.org/10.1037/a0027697

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*(8), 1058–1066. http://doi.org/10.1177/0956797611415540

Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, *12*(4), 323–345. http://doi.org/10.1080/13803610600765687

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence*, *20*(2), 482–506. http://doi.org/10.1111/j.1532-7795.2010.00644.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. *Psychology of Learning and Motivation*, *26*, 125–173. http://doi.org/10.1016/S0079-7421(08)60053-5

Nestler, S. (2015). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*. http://doi.org/10.1007/s11336-015-9474-9

Neuenschwander, M. P., Vida, M., Garrett, J. L., & Eccles, J. S. (2007). Parents' expectations

and students' achievement in two western nations. *International Journal of Behavioral Development*, *31*(5), 474–482. http://doi.org/10.1177/0165025407080589

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. http://doi.org/10.1080/00220670209596583

OECD. (2002). *PISA 2000 Technical Report*. Paris. Retrieved from http://www.oecd.org/pisa/pisaproducts/

OECD. (2012). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*. *PISA*. OECD Publishing. http://doi.org/10.1787/9789264201118-en

Pajares, F. (1996). Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research*, *66*(4), 543–578. http://doi.org/10.3102/00346543066004543

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29–48. http://doi.org/10.1080/01443410.2013.797339

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: a multicontext study. *Developmental Psychology*, *48*(6), 1629–42. http://doi.org/10.1037/a0029167

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *48*(1), 14–26. http://doi.org/10.1026/0049-8637/a000140

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction*, *22*(4), 262–270. http://doi.org/10.1016/j.learninstruc.2011.10.007

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2016). Effects of Informing Learners About the Dangers of Making Overconfident Judgments of Learning. *Journal of Educational Psychology*. http://doi.org/10.1037/edu0000132

Rosenthal, R. (1994). Interpersonal Expectancy Effects: A 30-Year Perspective. *Current*

Directions in Psychological Science, *3*(6), 176–179. http://doi.org/10.1111/1467-8721.ep10770698

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: an R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–70. http://doi.org/10.3758/s13428-011-0150-4

Schraw, G., Kuch, F., Gutierrez, A. P., & Richmond, A. S. (2014). Exploring a three-level model of calibration accuracy. *Journal of Educational Psychology*, *106*(4), 1192–1202. http://doi.org/10.1037/a0036653

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. http://doi.org/10.1016/j.learninstruc.2014.08.001

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. http://doi.org/10.1026/0012-1924.51.4.183

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond Effect: Generalizability and Moderation--Two Sides of the Same Coin. *American Educational Research Journal*, *47*(2), 390–433. http://doi.org/10.3102/0002831209350493

Sedikides, C., & Gregg, A. P. (2008). Self-Enhancement: Food for Thought. *Perspectives on Psychological Science*, *3*(2), 102–116. http://doi.org/10.1111/j.1745-6916.2008.00068.x

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research*, *46*(3), 407. http://doi.org/10.2307/1170010

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology*, *42*(1), 70–83. http://doi.org/10.1037/0012-1649.42.1.70

Simpkins, S. D., Fredricks, J. A., & Eccles, J. S. (2012). Charting the Eccles' expectancy-value model from mothers' beliefs in childhood to youths' activities in adolescence. *Developmental Psychology*, *48*(4), 1019–1032. http://doi.org/10.1037/a0027468

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204–221.

http://doi.org/10.1037/0278-7393.26.1.204

Spinath, B., & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction*, *15*(2), 87–102. http://doi.org/10.1016/j.learninstruc.2005.04.008

Spinath, B., & Steinmayr, R. (2008). Longitudinal Analysis of Intrinsic Motivation and Competence Beliefs: Is There a Relation Over Time? *Child Development*, *79*(5), 1555–1569. http://doi.org/10.1111/j.1467-8624.2008.01205.x

Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, *98*(3), 520–534. http://doi.org/10.1037/a0017057

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*(3), 521. http://doi.org/10.2307/1130719

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

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, *99*(2), 253–273. http://doi.org/10.1037/0022-0663.99.2.253

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. http://doi.org/10.1037/0022-0663.95.1.66

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, *98*(2), 438–456. http://doi.org/10.1037/0022-0663.98.2.438

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*(3), 763–777. http://doi.org/10.1037/a0027470

Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, *78*(4), 751–796. http://doi.org/10.3102/0034654308321456

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. http://doi.org/10.1207/s15326985ep3902_3

van Loon, M. H., de Bruin, A. B. H., van Gog, T., van Merriënboer, J. J. G., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica*, *151*, 143–154. http://doi.org/10.1016/j.actpsy.2014.06.007

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. JOUR. Retrieved from 10.1037/a0017908

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. http://doi.org/10.1037/a0027247

Watt, H. M. G. (2004). Development of Adolescents' Self-Perceptions, Values, and Task Perceptions According to Gender and Domain in 7th- through 11th-Grade Australian Students. *Child Development*, *75*(5), 1556–1574. http://doi.org/10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., Shapka, J. D., Morris, Z. a, Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*(6), 1594–1611. http://doi.org/10.1037/a0027838

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2015). Zur Bedeutung von Grundschulnoten für die Veränderung von Intrinsischer Motivation und Fähigkeitsselbstkonzept in Deutsch. *Zeitschrift Für Pädagogische Psychologie*, *29*(3–4), 193–204. http://doi.org/10.1024/1010-0652/a000160

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

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & et al. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-

esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552–565. http://doi.org/10.1037/0012-1649.27.4.552

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 the elementary school years: A 3-year study. *Journal of Educational Psychology*, *89*(3), 451–469. http://doi.org/10.1037/0022-0663.89.3.451

Wilson, T. D. (2009). Know Thyself. *Perspectives on Psychological Science*, *4*(4), 384–389. http://doi.org/10.1111/j.1745-6924.2009.01143.x

Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social Psychology*, *99*(1), 174–190. http://doi.org/10.1037/a0019390

Zell, E., & Krizan, Z. (2014). Do people have insight Into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. http://doi.org/10.1177/1745691613518075

Zhao, Q., & Linderholm, T. (2011). Anchoring effects on prospective and retrospective metacomprehension judgments as a function of peer performance information. *Metacognition and Learning*, *6*(1), 25–43. http://doi.org/10.1007/s11409-010-9065-1

2

"I thought I had an A": The Incidence and Correlates of Students' Postdiction Bias on a High-Stakes Exam

Lösch, T., Kelava, A., Nagengast, B., & Trautwein, U. (submitted). "I thought I had an A": The incidence and correlates of postdiction bias in a high stakes exam. *Learning and Individual Differences*.

Abstract

Students have been reported to differ on the trait to overestimate their performance on tests, yielding a positive *postdiction bias* (i.e., postdiction of test performance is higher than their actual score) in some students. There are several propositions about characteristics that predict postdiction bias, but they have not been brought together. Moreover, it is largely unknown whether inaccuracies in postdiction are domain-specific (overestimation in only some domains) or domain-general. This study examined the extent to which postdiction scores were biased in a sample of 2,164 high school students using the most important achievement test in German schools (*Abitur* exam grades). Students' postdictions were unbiased on average, but there were a considerable number of students who overestimated or underestimated their performance. Postdiction bias was domain-general and was related to students' cognitive abilities, internal and contextual references, and evaluative dispositions. These results were in line with a proposed model of postdiction formation.

Keywords: postdictions, postdiction bias, academic self-concept, narcissism

2. "I thought I had an A": The Incidence and Correlates of Students' Postdiction Bias on a High-Stakes Exam

2.1 Introduction

Students' self-appraisals, that is, their judgments of competence or achievement in a specific domain, test, or task, are key constructs for predicting academic achievement (Huang, 2011; Marsh & Martin, 2011; Valentine, DuBois, & Cooper, 2004). Yet, not only the absolute level of self-appraisals, but also their bias has been shown to be related to achievement (e.g., Chiu & Klassen, 2009, 2010). As documented by Kruger and Dunning (1999), an overestimation of achievement (i.e., a positive bias) was predominantly associated with lower performance.

In light of the supposed importance of accurate self-appraisals it may cause some concern that students' self-appraisals have been reported to be inaccurate, flawed, or biased (Dunning, Heath, & Suls, 2004). Moreover, bias in self-appraisals was found to be a stable individual difference (Pallier et al., 2002; Stankov, Lee, & Paek, 2009), indicating that there is a number of students who are systematically hampered in judging their achievement. But although several theories exist that describe which students have biased self-appraisals, these accounts have not been brought together.

This study addressed students' bias and aimed at identifying characteristics that were associated with this potentially worrisome trait. We investigated individual differences in students' ability to judge their achievement using their bias in *postdictions*. *Postdictions* are students' judgments of test performance following a test or exam, but before students know their result (Marsh, Trautwein, Lüdtke, & Köller, 2008). We investigated (1) students average bias, (2) predictors of postdiction bias, and (3) the domain-generality or domain-specificity of postdiction bias. To ensure external validity, we used a large sample and the most important test in German schools (the *Abitur*, in English and math) as performance criterion.

2.1.1 Differentiation of Academic Self-Concept, Postdictions, and Postdiction Bias

Postdictions conceptually overlap with the prominent construct of the academic self-concept in that both describe self-appraisals of ability or achievement. Empirically, they have been shown to be positively correlated at about $.50 < r < .80$ (Ackerman & Wolman, 2007; Marsh, Trautwein, et al., 2008). However, the two constructs differ in several important aspects that warrant further differentiation.

Academic self-concept (Eccles & Wigfield, 1995; Marsh & Shavelson, 1985) describes a person's perceived competence in a domain. It is past-oriented, domain-specific, and relatively stable (Bong & Skaalvik, 2003). It is based on social (Festinger, 1954), temporal (Albert, 1977), and dimensional (Möller & Marsh, 2013) references.

By contrast, the term *postdiction* refers to a retrospective self-evaluative judgment of how well a student performed on a specific test or exam (Hacker, Bol, Horgan, & Rakow, 2000; Marsh, Trautwein, et al., 2008). An example of a postdiction would be an answer to the question "What grade will you get on this math test that you just/recently completed?" In contrast to self-concept ratings, postdictions specify students' perceived performance in relation to a specific criterion and are, thus, goal-referenced and specific to a domain and task.

Postdictions can be studied from the perspective of their *calibration*, that is, their correspondence with a performance criterion (Alexander, 2013; de Bruin & van Gog, 2012). Calibration is argued to be pivotal for a successful regulation of learning tasks (Kostons, van Gog, & Paas, 2012; Metcalfe, 2009; Thiede, Anderson, & Therriault, 2003). However, several scholars pointed out that calibration should also be regarded as a matter of individual differences (Chiang, Therriault, & Franks, 2010; Jackson & Kleitman, 2014; Maki, Shields, Wheeler, & Zacchilli, 2005; Schraw & Nietfeld, 1998; Stankov et al., 2009). For instance, Pallier et al. (2002, p. 293) concluded that "there can now be little reason to question empirical evidence that human beings maintain a consistent, and at times substantial, relationship in the expression of confidence in the accuracy of their responses across a wide array of capabilities". That is, some individuals have a general tendency to overestimate their performance, some have a general tendency to underestimate their performance, and some tend to be accurate.

To operationalize and quantify calibration of postdictions, we use *postdiction bias* as indicator, which is the difference between postdictions and the performance criterion (Winne &

Jamieson-Noel, 2002). It refers to the absolute (i.e. signed) deviation of postdictions from achievement (cf. Chiu & Klassen, 2009, 2010). Postdiction bias has two directions: We refer to *overestimation* as a postdiction that is more positive than the actual performance (i.e., postdiction bias > 0). We refer to *underestimation* as a postdiction that is more negative than the actual performance (i.e., postdiction bias < 0). Postdiction bias is a useful measure of individual differences in the ability to calibrate oneself (Maki et al., 2005).

2.1.2 Average Postdiction Bias

There is some consensus that individuals and especially students are rather inaccurate in their self-appraisals on average (Dunlosky & Lipko, 2007; Dunning et al., 2004). Zell and Krizan (2014) found an average correlation of .29 by aggregating meta-analyses on the relation between self-appraisals and objective performance. However, several characteristics of self-appraisals moderated this correlation. In the case of postdictions, the s*elf-evaluation specificity*, *self-evaluation timing*, *task objectivity*, and *task familiarity* might yield a higher accuracy. Postdictions are highly *specific* (i.e. they refer to a specific test), they are assessed *after the performance*, the test performance might be *objective*, and exams in schools might be *tasks* that are *familiar* to students. Under these conditions, students may be more accurate, that is, their bias should be closer to 0. Thus, in a first step, mean level bias of postdictions should be identified.

Two measures inform about mean level calibration. First, the correlation between postdictions and the performance criterion serves as a measure of relative calibration in the whole sample (e.g., Glenberg & Epstein, 1985). That is, a higher correlation indicates that students are more calibrated in terms of their rank order. Second, the average postdiction bias informs about the absolute calibration of postdictions, for instance, whether students have a tendency to overestimate their performance (Chiu & Klassen, 2009, 2010). Both measures should be investigated, as they tap different aspects of calibration (Maki et al., 2005).

2.1.3 Predictors of Postdiction Bias

Postdiction bias might be regarded a highly important individual differences: Accurate postdictions are required for a successful study behavior, because postdictions were related to subsequent learning (Baars, Vink, van Gog, de Bruin, & Paas, 2014; Finn & Metcalfe, 2007;

Griffin, Jee, & Wiley, 2009). Thus, to identify students who suffer from the inability to judge their achievement, predictors of postdiction bias need to be identified. To identify predictors of postdiction bias, we propose a model of postdiction formation depicted in Figure 2.1. The model is built on earlier notions about formation of self-appraisals (Griffin et al., 2009; Koriat, 1997; Zhao & Linderholm, 2008), but with a focus that highlights individual differences

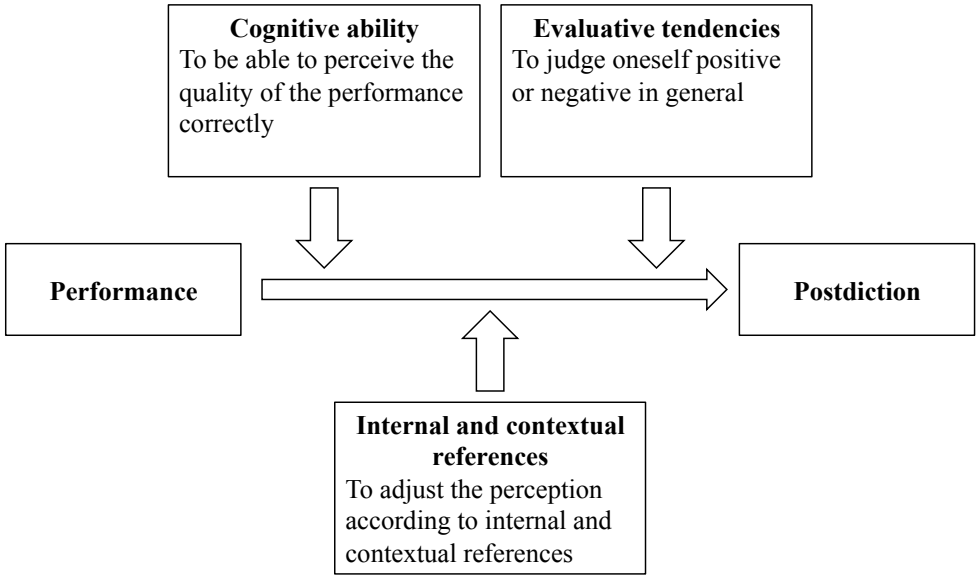

Figure 2.1. Model of postdiction formation.

According to the model, postdictions are formed in a perception and interpretation process whereby the single steps can be understood as lenses that may warp the original performance perception. Three processes are distinguished. First, students' perception is filtered by their cognitive ability to judge a performance accurately. Some students may lack the ability to identify the solution for a task as correct or incorrect (Kruger & Dunning, 1999). Second, the perception of performance may be adjusted based on internal and contextual references. For instance, if students know that they usually struggle with a specific type of task, they may lower their postdiction according to their prior experience (Zhao & Linderholm, 2008). And third, students' tendencies to perceive themselves as very positive or negative in general may alter their original performance interpretation (de Bruin, Kok, Lobbestael, & de Grip, 2016). According to the model, a performance perception is warped by these three lenses and is then manifested as a

postdiction. Based on this model, we derived six hypotheses about which characteristics should predict postdiction bias.

The first hypothesis refers to the operationalization of the cognitive ability lens and says *achievement in the same domain leads to underestimation (i.e., reduces overestimation)*. It has been shown repeatedly that achievement is negatively related to postdiction bias in the same domain; that is, low-performing students overestimated their performance more (Chiu & Klassen, 2009, 2010; Hacker et al., 2000; Kruger & Dunning, 1999; Maki et al., 2005; Nietfeld & Schraw, 2002). This may indicate that the ability to perform well in a test is also necessary to give an accurate postdiction.

The second, third, and fourth hypotheses specify internal and contextual references. Second, *achievement in a different domain leads to overestimation.* This effect is explained by an internal comparison (Möller & Marsh, 2013) that could also apply to postdictions. The effect of achievement in a very different domain (e.g., English as opposed to math) controlling for achievement in the same domain was investigated in self-concept research and termed the *internal/external frame of reference model* (I/E model; Möller & Marsh, 2013). That is, in a model where English self-concept is predicted by English and math achievement, the effect of English achievement was positive and the effect of math achievement was negative (Möller, Pohlmann, Köller, & Marsh, 2009). Because the I/E model was found for academic self-concept, it might also be valid for postdiction bias.

Third, *the domain-specific self-concept leads to overestimation in the same domain*. A high academic self-concept may evoke a false confidence: if students belief that they are generally good in math, they may suppose that the last exam (which they give a postdiction on) was no exception to that rule (Zhao & Linderholm, 2008). That is, the academic self-concept may work as a reference of prior experiences that biases postdictions. Furthermore, the feeling of being familiar in a domain (which is conceptually similar to the academic self-concept) led to more overestimations during a learning task (Shanks & Serra, 2014).

Fourth, *social references affect postdiction bias*, that is, when giving a postdiction, students could use their peers as reference. One of the most prominent social references is a school's average level of achievement. In case of the academic self-concept, a school's average level of achievement has consistently been found to be a negative predictor controlling for individual achievement differences (known as the *Big-Fish-Little-Pond-Effect*; Marsh, Seaton, et

al., 2008). Yet, with regard to other outcomes, there sometimes was a positive effect (a *behavioral assimilation effect* for achievement; Marsh et al., 2010; Stäbler, Dumont, Becker, & Baumert, 2016) or no effect (Marsh, Trautwein, et al., 2008). In an experimental study, a positive effect was found, that is, students, who were informed that peers performed very good in a task, overestimated themselves more (Zhao & Linderholm, 2011). Altogether, we suspect an effect of a school's average level of achievement, but have no hypothesis on its direction.

The fifth and sixth hypotheses refer to evaluative tendencies to see oneself overly positive or negative. Fifth, *narcissism leads to overestimation*. In terms of personality measures, the big five had moderate to low correlations to postdiction bias (Pallier et al., 2002). Contrary, *narcissism* describes the stable tendency to maintain a grandiose self (Back et al., 2013). In line with this definition, narcissism was related to overestimations of performance in university exams (de Bruin et al., 2016).

Sixth, *males show more overestimation than females*. In some studies, it was found that males overestimated themselves more (Lundeberg, Fox, & Puncochar, 1994; Stankov et al., 2009). Additionally, gender differences in the academic self-concept were more pronounced than gender differences in achievement, indicating overestimation by males (e.g., Pant et al., 2012). However, gender differences in postdictions were sometimes weak (Lundeberg, Fox, Brown, & Elbedour, 2000) and sometimes not present at all (e.g., Baars et al., 2014). Furthermore, gender differences in postdiction bias might depend on the specific domain (e.g., males overestimating performance in math and females overestimating performance in English; Marsh, 1993).

2.1.4 Postdiction Bias as a Domain-Specific or Domain-General Characteristic

A question that permeates the former hypotheses is whether postdiction bias is a domain-general or domain-specific construct. Postdictions are typically assessed at the domain-level, similar to the academic self-concept (e.g., in math or language; see also Bong & Skaalvik, 2003). However, there is evidence that there is only one person characteristic that describes the ability to be calibrated, that is, postdiction bias across domains might measure the same construct (Pallier et al., 2002; Schraw & Nietfeld, 1998).

To determine whether postdiction bias is domain-general or domain-specific, we propose three criteria. First, to be considered a domain-general characteristic, postdiction bias should be

positively correlated across different domains (cf. Goetz et al., 2007). Second, the I/E model (see hypothesis 3 of Section 1.3) should only be valid for domain-specific constructs (Schurtz, Pfost, Nagengast, & Artelt, 2014). That is, if the I/E model is found, this indicates that postdiction bias is domain-specific. Finally, as a third criterion for domain-generality, postdiction bias should be related to other domain-general characteristics that encompass positive self-appraisals, such as gender and narcissism.

2.1.5 The Present Investigation

Students differ in the extent to which they are able to calibrate their self-appraisals. To examine the incidence of this worrisome trait, this study addresses three questions targeting postdiction bias. First, we analyzed the incidence (i.e., mean level) of postdiction bias as an operationalization of student's ability to be calibrated. We expected a moderate correlation between postdictions and achievement (Zell & Krizan, 2014) as well as a positive postdiction bias indicating overestimations on average. Second, we investigated individual differences in postdiction bias by assessing its predictors. We expected that cognitive abilities, individual and contextual references, as well as evaluative tendencies predicted postdiction bias (as described by the six hypotheses above). Third, we tested the domain-specificity of postdiction bias. Because calibration was described as a stable individual difference, we expected postdiction bias to be domain-general (as indicated by the three criteria described above). Overall, our study focused on external validity: We used the written Abitur exams in English and math as performance criterion in a large sample of German high school students to address these research questions. The Abitur is the final examination in German academic track schools after 13 years of schooling and determines students' chances for university admission. Thus, it is the most important high-stakes exam in German schools.

2.2 Method

2.2.1 Sample

The sample consisted of students who wrote their final school exam (the Abitur) in 2005 in the German federal state of Hamburg (TOSCA-LAU study; see Trautwein, Köller, Lehmann,

& Lüdtke, 2007). The TOSCA-LAU study investigated all students of that age group (in total, 5,566). However, for our investigation, we analyzed a subsample of students. As standardized achievement tests were given in English and math only, we focused our analyses on students who completed their Abitur exam in these subjects. These were students with an advanced course in either English or math. In total, the subsample consisted of 2,461 students (44.21 % of the age group); 1,495 of them had an advanced course in English, 843 students had an advanced course in math, and 123 students had both. The mean age of the subsample was 19.78 (*SD* = 1.03; 49.25% female). The number of valid responses on the variables we used in analyses of the sample can be found in Table 2.1. The measurement was conducted 8 weeks after the Abitur exams and well before students were informed about their results.

Table 2.1

Descriptive Sample Statistics, Number of Valid Responses, Intraclass Correlation Coefficients, and Pairwise Zero-Order Correlations in the Sample (n = 2,461)

Variable		*n*	*M*	*SD*	Skewness	Kurtosis	ICC	Pairwise correlations										
								1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
English exam grade	1	1617	8.88	3.06	-0.01	-0.85	.10											
English postdiction	2	949	9.28	2.56	-0.29	-0.30	.12	.71										
English postdiction bias	3	949	0.09	2.20	-0.01	0.46	.06	-.58	.16									
English self-concept	4	1061	3.05	0.70	-0.42	-0.29	.04	.55	.64	-.05								
English achievement	5	1684	529.51	64.77	-1.06	2.47	.24	.48	.48	-.18	.34							
Math exam grade	6	966	7.74	3.99	-0.04	-0.94	.22	.67	.50	-.23	.29	.45						
Math postdiction	7	523	8.98	3.31	-0.43	-0.40	.16	.50	.57	.00	.22	.41	.84					
Math postdiction bias	8	523	0.36	2.14	0.35	0.80	.11	-.30	.09	.49	-.04	-.10	-.53	.03				
Math self-concept	9	575	3.03	0.70	-0.31	-0.58	.04	.44	.38	-.14	.23	.27	.61	.69	-.06			
Math achievement	10	2461	-0.73	0.70	0.40	-0.22	.17	.37	.34	-.12	.09	.26	.70	.68	-.21	.60		
Narcissism	11	1408	2.93	0.56	-0.04	-0.23	.02	-.01	.13	.16	.14	.09	.12	.21	.10	.26	.07	
Male	12	2461	0.51[a]		-0.03	-2.00	.02	-.13	-.01	.19	-.11	-.01	.07	.18	.16	.13	.20	.14

Note. The sample consists of 1,618 students with an advanced course in English and 966 students with an advanced course in math; 123 students of those had an advanced course in math and English. *n* = number of valid responses. ICC = Intraclass correlation.

[a] Proportion of males.

2.2.2 Procedure and Measures

Written exams of the Abitur in Hamburg took place in February 2005. The TOSCA-LAU study was conducted 8 weeks later. In the context of the study, postdictions, standardized achievement tests, academic self-concept, and narcissism were measured. The testing took place on 2 consecutive days.

Each student had to write Abitur exams in two advanced courses. The exams were the same for all students and graded by independent teachers from different schools. The grades ranged from 0 as the worst to 15 as the best score. This was the usual metric of grades in the last three years of upper secondary education in Germany. Exam grades were normally distributed (see Table 2.1). Postdictions were measured in each advanced course about 8 weeks after the exam with the item "Please estimate the grade you will get on your Abitur exam in this subject." The scale was identical to the scale used for grades. We calculated the difference between the postdictions and the actual grades on the exam as a measure of postdiction bias (Chiu & Klassen, 2009, 2010). Thus, a positive score indicated an overestimation, a negative score indicated an underestimation, and a score close to 0 indicated an accurate postdiction. The exam grades were gathered from the local administrative department.

Standardized low stakes achievement tests (see also Trautwein et al., 2012) were administered in math (Baumert, Bos, & Lehmann, 2000) and English (Köller & Trautwein, 2004). Domain-specific self-concept items were based on established self-concept measures and worded identically in both domains (four items each; cf. Schwanzer, Trautwein, Lüdtke, & Sydow, 2005). Narcissism was measured with six items adapted from the NPI (Raskin & Hall, 1979). The items mostly captured aspects of *grandiose narcissism* (Back et al., 2013). The items were chosen to capture a strong common factor from the NPI (cf. Jonkmann, Becker, Marsh, Lüdtke, & Trautwein, 2012). See Table 2.2 for an overview of the self-concept and narcissism scales and items.

Table 2.2

Scales, Internal Consistency, and Items Used

Scale	Reliability	Item
Academic self-concept	English: α = .84 Math: α = .85	"This subject is one of my best subjects" "I have always been good in this subject" "In this subject, I learn fast" "In this subject, I am very talented"
Narcissism	α = .87	"Altogether, I think I am an extraordinary person" "I have many very special abilities" "Indeed, I am a person with unique talents" "People could learn a lot from me" "I am a person with many special characteristics" "I have the prerequisites to accomplish something special"

Note. The wording of items was identical for the two self-concept scales.

2.2.3 Statistical Procedure

Our statistical analyses consisted of three steps. For the description of the incidence of postdiction bias, we investigated descriptive statistics of postdiction bias of exam grades in math and English. This step informed about mean level tendencies. In addition, we estimated the correlations between the postdictions and exam scores in math and English to inform about rank order accuracy.

In the second step, we tested the predictors of postdiction bias. In order to facilitate the interpretation of regression coefficients, all potential predictors were standardized in the subsample. We tested our hypotheses on the predictors with a number of manifest, multivariate, multilevel regression models using Mplus 7.1 (Muthen & Muthen, 2012). Altogether, we estimated six regression models. First, we tested each predictor separately to investigate its individual effect. That is, self-concept, achievement, gender, and narcissism were tested as separate predictors of postdiction bias. For self-concept and achievement, these models were limited to the same domain (e.g., postdiction bias in English was predicted by self-concept in English). To test for achievement in the respective other domain, we included achievement in both domains at the same time to predict postdiction bias in one domain. Predictors were included simultaneously on the student (i.e., individual) level and on the school (i.e., cluster) level to estimate the total effect of the predictors on both levels. As predictors on the school

level, we used manifest aggregated variables of every variable at the individual level (i.e., self-concept in English and math, achievement in English and math, narcissism, and gender). Variables were aggregated after they were standardized within the sample. In addition, the predictors on the individual level were group-mean-centered (Enders & Tofighi, 2007; Marsh et al., 2009). To test for the contextual effect of a school's average level of achievement after controlling for individual achievement differences, we calculated the difference between the regression coefficient for achievement on the school level and the regression coefficient for achievement on the individual level (Snijders & Bosker, 2011). Finally, we estimated a model with all predictors to assess their predictive value in a joint model. We specified multivariate models to estimate the covariance between postdiction bias in English and math controlling for all predictors. We used level-specific R^2 as provided by Mplus as measures of explained variance. In addition, we also computed the mean squared prediction error suggested by Snijders and Bosker (1994) as an overall measure of explained variance.

In the third step, we evaluated the proposed criteria to decide whether postdiction bias was domain-general. That is, we inspected (1) the zero-order correlation and the correlations between postdiction bias scores in English and math in the regression models, (2) the effect of achievement in a different domain, and (3) the effect of narcissism and gender.

2.2.4 Missing Data

For descriptive statistics and zero-order correlations, we used only students with no missing data. The number of students with valid scores on all variables is indicated in Table 2.1. To handle the missing data in the regression analyses, we used the model-based full information maximum likelihood estimation approach and included all available information in the estimation of the statistical model (Enders, 2010). To improve the quality of this estimation, all variables that were not specified as predictors were used as auxiliary variables and thus correlated to all variables in the model (Enders, 2010; Graham, 2003).

2.3 Results

2.3.1 Incidence of Postdiction Bias

In our first descriptive analysis, we investigated the students' average level of calibration in their postdictions. Table 2.1 gives the relevant descriptive statistics for postdiction bias: The number of students who took their Abitur exam in the respective school subject, the number of students who made a postdiction assessment in that school subject, average grades, average postdictions, average postdiction bias, and the relative accuracy of the postdictions. There were missing data on postdictions for about one third of the students. This proportion of missing data might largely be explained by the different data sources for exam grades and postdictions. Whereas grades were obtained from the administrative department with only a few missing cases, postdictions were assessed during the second day of testing (where more students were missing than on the first day of testing).

In both school subjects, postdiction bias seemed to be slightly positively skewed. In English, average postdiction bias was 0.09, SD = 2.20. In math, average postdiction bias was 0.36, SD = 2.14. That is, in both subjects average postdiction bias was positive but below 1. Note that grades in the German Abitur have a scale that ranges from 0 to 15. The difference from 0 was equal to a standardized effect of d = 0.04 in English and d = 0.17 in math.

To assess not only the direction and distribution of postdiction bias but also the correspondence between postdictions and grades, we estimated the correlation between postdictions and grades (i.e., the relative accuracy). The correlations were r = .71 in English and r = .84 in math, respectively. Overall, students appeared to be highly accurate on average.

2.3.2 Predictors of Postdiction Bias

Descriptive statistics and zero-order correlations for all variables used in these analyses are presented in Table 2.1. There was substantial variability in postdiction bias on the school (cluster) level as indicated by intraclass correlations.

We computed regression analyses simultaneously predicting postdiction bias in English and math. Altogether, we estimated six multivariate multilevel regression models with

postdiction bias in English and math as the dependent variables. Table 2.3 presents the results of the regression analyses. Note that the regression coefficients are partially standardized; that is, the predictors are z-standardized, but the postdiction bias scores are not. Thus, the regression coefficients can be meaningfully interpreted as postdiction bias measured in grade points associated with a score on the predictors that is one standard deviation above the mean.

In English, self-concept did not significantly predict postdiction bias. However, English achievement was a significant negative predictor of postdiction bias (*coef.* = -0.59, *SE* = 0.10). That is, students with lower achievement tended to overestimate their performance. In addition, a school's average level of English achievement was a significant positive predictor after controlling individual achievement differences (contextual effect: *coef.* = 0.43, *SE* = 0.21), indicating that students with equal individual achievement levels tended to overestimate their performance more in schools with higher average achievement (i.e., a social assimilation effect). Math achievement, while controlling for English achievement, had no significant effect on postdiction bias in English contradicting an I/E model. However, narcissism had a positive effect on postdiction bias in English (*coef.* = 0.30, *SE* = 0.07). Finally, males had a postdiction bias that was almost 1 grade point higher than that of females (*coef.* = 0.88, *SE* = 0.13). In the final regression model that included all predictors, the total amount of variance explained in postdiction bias in English was R^2 = .11 on the student level and R^2 = .35 on the school level. The overall explained variance according to Snijders and Bosker (1994) was R^2 = .12.

In math, self-concept predicted postdiction bias only in the final model, which included all predictors (*coef.* = 0.24, *SE* = 0.12). Furthermore, math achievement (*coef.* = -0.58, *SE* = 0.14) and male gender (*coef.* = 0.74, *SE* = 0.18) significantly predicted postdiction bias in math. As in English, male students and students with lower achievement showed overestimations in their postdiction bias. A school's average level of math achievement predicted postdiction bias in math only in the model that included all predictors (contextual effect: *coef.* = 0.81, *SE* = 0.31). As in English, it was a positive effect, indicating a social assimilation effect. However, English achievement and narcissism did not predict postdiction bias in math. In the final regression model including all predictors, the total amount of variance explained in math postdiction bias was R^2 = .14 on the student level and R^2 = .20 on the school level. The overall explained variance according to Snijders and Bosker (1994) was R^2 = .15.

Table 2.3

Multivariate Multilevel Models Predicting Postdiction Bias in English and Math

coef. (SE)	Model					
	1	2	3	4	5	6
Predictors on the student level						
English postdiction bias						
English achievement	-0.58* (0.10)	-0.56* (0.10)				-0.60* (0.12)
Math achievement		-0.03 (0.10)				-0.13 (0.10)
English self-concept			-0.16 (0.09)			0.05 (0.09)
Narcissism				0.29* (0.08)		0.26* (0.08)
Male					0.85* (0.13)	0.85* (0.14)
Math postdiction bias						
Math achievement	-0.63* (0.16)	-0.64* (0.18)				-0.90* (0.22)
English achievement		0.03 (0.17)				0.02 (0.16)
Math self-concept			-0.14 (0.10)			0.24* (0.12)
Narcissism				0.22 (0.12)		0.15 (0.12)
Male					0.60* (0.18)	0.87* (0.16)
Predictors on the school level						
English postdiction bias						
English achievement	-0.17 (0.20)	0.05 (0.25)				0.00 (0.27)
Math achievement		-0.47 (0.30)				-0.50 (0.30)
English self-concept			0.27 (0.20)			0.27 (0.22)
Narcissism				0.45 (0.27)		0.43 (0.28)
Male					0.77 (0.80)	0.76 (0.82)
Math postdiction bias						
Math achievement	-0.11 (0.30)	-0.14 (0.31)				-0.09 (0.31)
English achievement		-0.03 (0.32)				-0.08 (0.31)
Math self-concept			-0.11 (0.32)			-0.27 (0.30)
Narcissism				0.70* (0.35)		0.82* (0.37)
Male					0.31 (0.90)	-0.07 (0.82)
Contextual effects						
English achievement	0.40* (0.20)	0.60* (0.24)				0.60* (0.27)
Math achievement	0.53 (0.33)	0.51 (0.33)				0.81* (0.31)
Bias covariance						
Student level	0.53* (0.13)	0.52* (0.14)	0.52* (0.12)	0.49* (0.13)	0.48* (0.13)	0.53* (0.13)
School level	0.26 (0.25)	0.24 (0.25)	0.23 (0.25)	0.14 (0.27)	0.25 (0.24)	0.03 (0.26)
R^2						
Student level						
English	.05	.05	.00	.02	.04	.11
Math	.08	.08	.00	.01	.02	.14
School level						
English	.03	.12	.05	.12	.04	.34
Math	.00	.01	.01	.15	.00	.20
Snijders and Bosker (1994)						
English	.05	.06	.01	.02	.04	.12
Math	.07	.07	.00	.03	.02	.15

Note. Shown are estimated coefficients (and their standard errors in brackets) in the multivariate multilevel regression models. Regression coefficients were partially standardized because the predictor variables were standardized, but postdiction bias was not. All predictors were specified as correlated. Contextual effects were estimated as the difference between the regression coefficient on the school level minus the regression coefficient on the student level. Coef. =

estimated coefficient; SE = standard error; Bias covariance = standardized covariance between postdiction bias in English and postdiction bias in math. Snijders and Bosker (1994) = An estimation of R^2 according to Snijders and Bosker (1994) that models an aggregate of the proportional reduction in mean squared prediction error of the dependent variables both on the individual and the cluster level.

* $p < .05$.

2.3.3 Postdiction Bias as a Domain-General Characteristic

To test if postdiction bias was rather domain-general or domain-specific, we examined three criteria: (1) correlated postdiction bias scores, (2) an I/E model for postdiction bias, (3) correlations between postdiction bias and domain-general characteristics. First, the zero-order correlation between postdiction bias in English and math was $r = .49$ and the correlations in the regression models controlling for the predictors varied within $.49 < r < .53$. That is, there was a substantial correlation across domains. Second, as can be seen in Table 2.4, we found no evidence for the I/E model, that is, achievement in a different domain did not predict postdiction bias controlling for achievement in the same domain. Third, as mentioned earlier, narcissism predicted postdiction bias in English and gender predicted postdiction bias in English and math. That is, 2 of the 3 criteria indicated that postdiction bias was domain-general.

2.4 Discussion

In our study, we investigated individual differences in students' ability to postdict their performance accurately. Thereby, we focused on the distribution of postdiction bias and mean level accuracy, predictors of postdiction bias, and the question whether postdiction bias was a domain-general construct.

2.4.1 Average Postdiction Bias

Contrary to our expectations and contrary to earlier reports (Dunlosky & Lipko, 2007; Dunning et al., 2004) we found that students postdictions were highly related to their exam performance and that their bias was on average just slightly above 0 with a very small effect size

(Cohen, 1992). That is, students in our sample were well calibrated on average. Students' bias might have depended on contextual factors like the moderators proposed by Zell and Krizan (2014). For instance, the postdictions we investigated were highly *specific*, they were assessed *after the performance*, the Abitur grades may be considered as *objective*, and the *task* was relatively *familiar* to the students. Future studies should examine the effect of these moderators more closely, that is, how they affect correlations between postdictions (or other self-appraisals) and performance as well as mean level postdiction bias. But furthermore, it should be investigated whether these moderators interact with student characteristics (Maki et al., 2005). For example, for students with low achievement, task familiarity might not be very important. But high achieving students could be much less biased in a familiar than in an unfamiliar task.

2.4.2 Predictors of Postdiction Bias

Although students were not biased on average, there was much variation between students. The proposed model of postdiction formation (see Figure 2.1) could explain some of the differences. All three lenses had an effect on postdiction bias, but not all proposed predictors were significant.

In terms of explained variance, the first lens of cognitive ability was the most important one. Achievement in the same domain on the individual student level had the expected negative effect that was also found in earlier studies (Kruger & Dunning, 1999). Thus, the cognitive ability that was necessary to perform well also seemed to be necessary to have a well-calibrated postdiction.

Evidence for the references lens was mixed. First, achievement in a different domain had no effect on postdictions (i.e., there was no I/E model), indicating that students did not compare achievement in the domains internally. Similarly, the domain-specific self-concept had only a small to negligible effect on postdiction bias. It was correlated to postdictions and grades, but not to postdiction bias. However, achievement on the school level had an effect on postdiction bias. That is, students in a school with high average achievement overestimated their performance more, which is in line with the proposed reference effect (Zhao & Linderholm, 2011): In schools with a higher average achievement, peers might perceive the exams easier. This impression could lead students to overestimate their own performance.

Finally, the proposed lens of evaluative tendencies had the expected effect. Narcissism predicted postdiction bias, at least in English. That is, students' tendency to perceive themselves as a grandiose being also inflated their postdictions (de Bruin et al., 2016). Gender differences emerged irrespective of domain in contrast to results of academic self-concept research (Marsh, 1993): males overestimated their performance in English and in math. Possibly, these results indicated a gendered tendency for males to overestimate their performance in general.

Altogether, the results showed that individual differences could be attributed to all three sources: cognitive abilities, internal and contextual references, and evaluative tendencies. Although the effect sizes of the single predictors were not very large in the taxonomy of Cohen (1992), their joint effect in the same terminology would be *moderate*, indicating that individual differences in postdiction bias matter. Adding up the effect of single predictors can possibly identify the most biased students. For instance, male, low achieving students in schools with a high average achievement are probably most biased in their postdictions.

2.4.3 Postdiction Bias as a Domain-General Characteristic

Following the notion that the ability to calibrate is a stable individual difference (Pallier et al., 2002; Stankov et al., 2009), postdiction bias could be a domain-general construct. To evaluate this hypothesis, we tested three criteria: (a) correlated postdiction bias scores, (b) an I/E model of postdiction bias, and (c) relations to narcissism and gender. Postdiction bias scores were correlated (even after controlling for all of our predictors). Furthermore, regarding our second criterion, which would indicate domain-specificity, we observed no I/E model. And finally, postdiction bias was related to both gender and narcissism. Taken together, 2 of the tests indicated that postdiction bias was a domain-general construct and there was no contradiction to this hypothesis. Altogether, these results support the notion that postdiction bias was a domain-general individual difference.

2.4.4 Limitations and Directions for Further Research

In our study, postdictions were assessed about 8 weeks after the Abitur exams. As a consequence, students' postdictions could have become less biased due to a temporal decrease in

overestimation (Sweeny & Krizan, 2012). Thus, investigations of the temporal development of postdictions are necessary to provide more information about this issue.

Another aspect of temporal development is that the students in our sample were at the end of their school careers and were accordingly older and presumably more accurate (Stipek & Mac Iver, 1989). To investigate this topic in more detail, future studies should compare postdiction bias of different age groups.

Although it is likely that factors suggested by Zell and Krizan (2014) led to the accuracy shown in our study, we were not able to make causal claims because we did not manipulate these moderators in our study. However, one could use natural experiments to test these moderators. For example, the task familiarity moderator could be tested by focusing on students' accuracy before and after a school transition.

A limiting factor of our analyses was that we had students' postdictions in only two advanced courses. On the one hand, accuracy might be lower in other subjects. For instance, it could be easier to postdict performance in math (which is a quite objective subject) than in other subjects. On the other hand, having only two subjects limited our analyses of the domain-generality of postdiction bias. Examinations of domain-generality would be richer by investigating more domains per students at the same time. For example, similarity of domains could moderate the correlations of postdiction bias scores across multiple domains (Goetz et al., 2014). Furthermore, features of domains (e.g., objectivity of tests) could moderate accuracy as well (e.g., considering math compared to arts).

2.4.5 Practical Implications

This study provided only first evidence in support of the model of postdiction formation and moreover no causal claims should be made. However, this study may still provide practitioners with some guidance. First of all, students seemed to differ with regard to their ability to calibrate their self-appraisals. That is, some students seem to achieve calibration on their own, but others might require external assistance to reach that goal. Second, the model informs about characteristics of potentially biased students. Teachers could use the model as diagnostic tool to identify students with potentially biased self-appraisals. That is, especially male students, students that have a lower achievement, students that display an exaggerated, narcissistic self-view, and students in high-performing classes may have biased self-appraisals.

Finally, the postdiction model might indicate that different strategies are necessary to help different students to become calibrated depending on the source of bias. If a student lacks the ability to identify shortcomings, more precise feedback might help. But if a student has a narcissistic tendency to overestimate his abilities, this strategy might rather be counterproductive.

2.4.6 Conclusion

By investigating students' individual differences in their ability to postdict their performance in a large sample, we provided a description of the calibration that students show in a high stakes exam. In our case, students were less biased than often observed (Dunning et al., 2004) but this result may be reasonable when considering moderators that maybe affect student accuracy (Zell & Krizan, 2014). According to the proposed model, students' cognitive abilities, internal and contextual references, and evaluative tendencies predicted postdiction bias and moreover, postdiction bias appeared to be a domain-general individual difference. Because we found support for the proposed model of postdiction formation, we can conclude that individual differences in postdiction bias arise as a combination of several individual predictors.

References

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology. Applied*, *13*(2), 57–78. http://doi.org/10.1037/1076-898X.13.2.57

Albert, S. (1977). Temporal comparison theory. *Psychological Review*. http://doi.org/10.1037/0033-295X.84.6.485

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. http://doi.org/10.1016/j.learninstruc.2012.10.003

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, *33*, 92–107. http://doi.org/10.1016/j.learninstruc.2014.04.004

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. a. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. http://doi.org/10.1037/a0034431

Baumert, J., Bos, W., & Lehmann, R. (Eds.). (2000). *TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie—Mathema- tische und naturwissenschaftliche Bildung am Ende der Schullaufbahn [TIMSS/III: Third international mathematics and science study— Students' knowledge of mathematics and sc*. Opladen, Germany: Leske + Budrich.

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

Chiang, E. S., Therriault, D. J., & Franks, B. a. (2010). Individual differences in relative metacomprehension accuracy: Variation within and across task manipulations. *Metacognition and Learning*, *5*(2), 121–135. http://doi.org/10.1007/s11409-009-9052-6

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. http://doi.org/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. http://doi.org/10.1016/j.learninstruc.2008.11.002

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

de Bruin, A. B. H., Kok, E. M., Lobbestael, J., & de Grip, A. (2016). The impact of an online tool for monitoring and regulating learning at university: overconfidence, learning strategy, and personality. *Metacognition and Learning*. http://doi.org/10.1007/s11409-016-9159-5

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. http://doi.org/10.1016/j.learninstruc.2012.01.003

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*(4), 228–232. http://doi.org/10.1111/j.1467-8721.2007.00509.x

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. http://doi.org/10.1111/j.1529-1006.2004.00018.x

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. http://doi.org/10.1177/0146167295213003

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–138. http://doi.org/10.1037/1082-989X.12.2.121

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*. http://doi.org/10.1177/001872675400700202

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(1), 238–44. http://doi.org/10.1037/0278-7393.33.1.238

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718. http://doi.org/10.1037//0278-7393.11.1-4.702

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology*, *99*(4), 715–733. http://doi.org/10.1037/0022-0663.99.4.715

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C., & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology*, *5*(October), 1–14. http://doi.org/10.3389/fpsyg.2014.01153

Graham, J. W. (2003). Adding missing-data-relevant variables to FIML-based structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*. http://doi.org/10.1207/S15328007SEM1001_4

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, *37*(7), 1001–1013. http://doi.org/10.3758/MC.37.7.1001

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. a. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. http://doi.org/10.1037//0022-0663.92.1.160

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. http://doi.org/10.1016/j.jsp.2011.07.001

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: Capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*(1), 25–49. http://doi.org/10.1007/s11409-013-9110-y

Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2012). Personality traits moderate the Big-Fish–Little-Pond Effect of academic self-concept. *Learning and Individual Differences*, *22*(6), 736–746. http://doi.org/10.1016/j.lindif.2012.07.020

Köller, O., & Trautwein, U. (2004). Englischleistungen von Schülerinnen und Schülern an allgemein bildenden und beruflichen Gymnasien. In O. Köller, U. Watermann, U. Trautwein, & O. Lüdtke (Eds.), *Wege zur Hochschulreife in Baden-Württemberg. TOSCA -*

Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien (pp. 285–326). Opladen: Leske und Budrich.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. http://doi.org/10.1037//0096-3445.126.4.349

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, *22*(2), 121–132. http://doi.org/10.1016/j.learninstruc.2011.08.004

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. http://doi.org/10.1037/0022-3514.77.6.1121

Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology*, *92*(1), 152–159. http://doi.org/10.1037/0022-0663.92.1.152

Lundeberg, M. A., Fox, P. W., & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, *86*(1), 114–121. http://doi.org/10.1037/0022-0663.86.1.114

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*(4), 723–731. http://doi.org/10.1037/0022-0663.97.4.7238

Marsh, H. W. (1993). The multidimensional structure of academic self-concept: Invariance over gender and age. *American Educational Research Journal*, *30*(4), 841–860. http://doi.org/10.3102/00028312030004841

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764–802. http://doi.org/10.1080/00273170903333665

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. http://doi.org/10.1348/000709910X503501

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Régner, I., … Gibbons, F. X. (2010). Phantom behavioral assimilation effects: Systematic biases in social comparison choice studies. *Journal of Personality*, *78*(April 2010), 671–710. http://doi.org/10.1111/j.1467-6494.2010.00630.x

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. http://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. http://doi.org/10.1207/s15326985ep2003_1

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. http://doi.org/10.1037/0022-0663.100.3.510

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. http://doi.org/10.1111/j.1467-8721.2009.01628.x

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. http://doi.org/10.1037/a0032459

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. http://doi.org/10.3102/0034654309337522

Muthen, L., & Muthen, B. (2012). *Mplus User's Guide*. *Los Angeles, CA: Muthén & Muthén* (Seventh Ed). Los ANgeles, CA: Muthén & Muthén.

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. http://doi.org/10.1080/00220670209596583

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The Role of Individual Differences in the Accuracy of Confidence Judgments. *The Journal of General Psychology*, *129*(3), 257–299. http://doi.org/10.1080/00221300209602099

Pant, H. A., Stanat, P., Schroeders, U., Roppelt, A., Siegle, T., & Pöhlmann, C. (2012). *The IQB National Assessment Study 2012 - Competencies in Mathematics and the Sciences at the End of Secondary Level I: Summary*. Münster: Waxmann. Retrieved from https://www.iqb.hu-berlin.de/laendervergleich/lv2012/Bericht

Raskin, R. N., & Hall, C. S. (1979). A Narcisstic Personality Inventory. *Psychological Reports*, *45*(2), 590–590. http://doi.org/10.2466/pr0.1979.45.2.590

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, *90*(2), 236–248. http://doi.org/10.1037//0022-0663.90.2.236

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. http://doi.org/10.1016/j.learninstruc.2014.08.001

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. http://doi.org/10.1026/0012-1924.51.4.183

Shanks, L. L., & Serra, M. J. (2014). Domain familiarity as a cue for judgments of learning. *Psychonomic Bulletin & Review*, *21*(2), 445–453. http://doi.org/10.3758/s13423-013-0513-1

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research*. http://doi.org/10.1177/0049124194022003004

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2016). What Happens to the Fish's Achievement in a Little Pond? A Simultaneous Analysis of Class-Average Achievement Effects on Achievement and Academic Self-Concept. *Journal of Educational Psychology*. http://doi.org/10.1037/edu0000135

Stankov, L., Lee, J., & Paek, I. (2009). Realism of Confidence Judgments. *European Journal of Psychological Assessment*, *25*(2), 123–130. http://doi.org/10.1027/1015-5759.25.2.123

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*(3), 521. http://doi.org/10.2307/1130719

Sweeny, K., & Krizan, Z. (2012). Sobering Up: A Quantitative Review of Temporal Declines in Expectations. *Psychological Bulletin*, *139*(3), 702–724. http://doi.org/10.1037/a0029951

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. http://doi.org/10.1037/0022-0663.95.1.66

Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (2007). *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.

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*(3), 763–777. http://doi.org/10.1037/a0027470

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. http://doi.org/10.1207/s15326985ep3902_3

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. http://doi.org/10.1016/S0361-476X(02)00006-1

Zell, E., & Krizan, Z. (2014). Do people have insight Into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. http://doi.org/10.1177/1745691613518075

Zhao, Q., & Linderholm, T. (2008). Adult Metacomprehension: Judgment Processes and Accuracy Constraints. *Educational Psychology Review*, *20*(2), 191–206. http://doi.org/10.1007/s10648-008-9073-8

Zhao, Q., & Linderholm, T. (2011). Anchoring effects on prospective and retrospective metacomprehension judgments as a function of peer performance information. *Metacognition and Learning*, *6*(1), 25–43. http://doi.org/10.1007/s11409-010-9065-1

3

A Well-rounded View: Using an Interpersonal Approach to Predict Achievement by Academic Self-concept and Peer Ratings of Competence

Lösch, T., Lüdtke, O., Robitzsch, A., Kelava, A., Nagengast, B., & Trautwein, U. (submitted). A Well-rounded View: Using an Interpersonal Approach to Predict Achievement by Academic Self-concept and Peer Ratings of Competence. *Contemporary Educational Psychology*.

Abstract

Academic self-concept is a prominent construct in educational psychology that predicts future achievement. Similarly, peer ratings of competence predict future achievement as well. But as both constructs have not been brought together in a multi-rater framework, it is unknown whether academic self-concept will also be advantageous when it is more positive than peers' ratings of competence. In this study, an *interpersonal approach* (Kwan, John, Kenny, Bond, & Robins, 2004) was applied to investigate the interrelation and uniqueness of the two constructs. The interpersonal approach decomposes variance of self-concept ratings into a "method" part that is due to the student as rater (*perceiver effect*), into a "trait" part, that is due to the student's achievement (*target effect*), and an idiosyncratic self-view (*self-enhancement*). In a round-robin design of competence ratings by which each student in a class rated every classmate's competence, a total of 1,549 school students in 87 classes of two age cohorts rated their own math competence and the math competence of their classmates. Three main results emerged. First, the target effect shared substantial variance with self-concept ratings. Second, the target effect was highly correlated to achievement and predicted gains in achievement. Third, self-enhancement had a small positive association with (future) achievement. Altogether, this study shows that self-concept ratings were closely tied to peer ratings of competence and that both constructs shared their positive effect on future achievement.

Keywords: social relations model, self-enhancement, academic self-concept, achievement

3. A Well-rounded View: Using an Interpersonal Approach to Predict Achievement by Academic Self-concept and Peer Ratings of Competence

3.1 Introduction

Academic self-concept is one of the most prominent constructs in educational psychology (Marsh & Craven, 2006; Marsh & Martin, 2011) and meta-analyses have shown that it is a key predictor of academic achievement (Huang, 2011; Valentine, DuBois, & Cooper, 2004). Therein, academic self-concept is referred to as a self-perception of competence in a specific academic domain (Eccles & Wigfield, 1995; Marsh & Shavelson, 1985). Yet in schools, not only students themselves, but also their classmates perceive a student's competence. These *peer ratings of competence* (Gest, Domitrovich, & Welsh, 2005) were found to be important as well, as they consistently predicted a target student's academic self concept and future achievement (Chen, Hughes, Liew, & Kwok, 2010; Denissen, Schönbrodt, van Zalk, Meeus, & van Aken, 2011; Gest, Rulison, Davidson, & Welsh, 2008; Hughes, Dyer, Luo, & Kwok, 2009).

By presenting these two lines of research besides each other, one might question whether a positive academic self-concept is also advantageous when peers do not share this positive perception. That is, students that rate themselves as more competent than peers rate them might be especially high achieving (Marsh & Craven, 2006) – or they might be rather low achieving (de Bruin & van Gog, 2012). So far, research has not connected academic self-concept and peer ratings of competence in a multi-rater perspective (McAbee & Connelly, 2016). Yet, a multi-rater perspective would be necessary to investigate the effect of self-ratings that are not shared with peers on future achievement.

In the present study, we aimed at bringing academic self-concept and peer ratings of competence together, validate them against each other, identify their overlap, and extract an

idiosyncratic self-rating. In order to achieve this goal, we drew on an *interpersonal approach* (Kwan et al., 2004) that has been proposed as a general approach for relating self-ratings to peer ratings. Using a round-robin design for competence ratings (i.e., each student in a class rates every classmate's competence), we show how the interpersonal approach can disentangle self-concept ratings into a "method" component of how a student rated others (*perceiver* effect), a "trait" perception that is shared with peers (i.e., *target effect*), and a non-shared self-perception (i.e., *self-enhancement*). In a longitudinal school achievement study with two age cohorts (fifth and eighth graders), we tested how competence ratings by the self and by peers overlapped and which components of competence ratings were associated with students' actual achievement in math (measured as grades and test scores).

3.1.1 An Interpersonal Approach to Math Self-Concept

In the interpersonal approach, self- and peer perceptions of competence may be studied within one framework. The interpersonal approach originated in research on person perception (Kwan et al., 2004) and is based on Kenny's *social relations model* (SRM; Kenny, Kashy, & Cook, 2006). The SRM is a methodology that can be applied to analyze and describe interpersonal perceptions in various domains. The advantage of this approach is that it allows researchers to disentangle competence ratings into different components as originally outlined by Cronbach (1955).

Within the SRM, individual differences in peer ratings (e.g., how Alex rates Beth and how Beth rates Alex) are attributed to three sources. The prototypical design for investigating interpersonal ratings is a round-robin design (Kenny, 1994): All students in a classroom rate each classmate's competence in math (i.e., peer ratings of math competence; see Figure 3.1 for an illustration). The *perceiver effect* describes a tendency to rate peers as competent (e.g., How competent does Alex think his classmates are in math?). The *target effect* describes a tendency to be rated as competent (e.g., How competent do Alex's classmates think he is in math?). Third, the *relationship effect* describes the unique tendency of one student to rate one other student (e.g., How competent does Alex think that Beth in particular is in math?). In sum, Alex's rating of Beth's competence in math is the result of three distinct components (i.e., perceiver, target, and relationship effects).

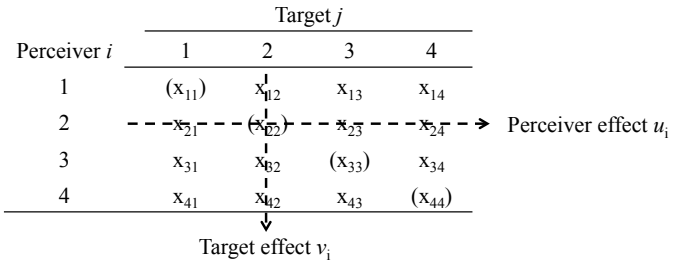

Figure 3.1. Illustration of the target effect and the perceiver effect in a round-robin design with four participants (Alex, Beth, Charlie, and Daisy). Student *i*'s math competence rating of student *j* is represented by x_{ij}. Student *i*'s math self-concept rating is represented by (x_{ii}). The arrows indicate the perceiver or row effects u_i (i.e., how student *i* tends to rate the math competence of others) and the target or column effects v_j (i.e., how others rate the math competence of student *i*).

The interpersonal approach emphasizes that self-ratings and peer ratings may be seen as perceptions of the same underlying trait (Kwan et al., 2004). The same logic is applied in multi-trait multi-method analyses (Campbell & Fiske, 1959): one trait is measured using multiple methods, that is, from the perspective of multiple perceivers. In case of math self-concept, it is a mono-trait multi-method design: The "trait" of (perceived) achievement is measured using the "methods" of self-ratings and classmates' ratings (McAbee & Connelly, 2016). The "trait" part of self-concept ratings refers to the student's achievement (target effect) and is the variance, which is shared with ratings by classmates. The "method" part of self-concept ratings refers to the student's self as rater (perceiver effect) and is the variance, which is shared with ratings by the self about classmates. Accordingly, academic self-concept ratings can be assumed to be a function of target and perceiver effects, but these effects might not be weighted like they are in peer ratings of achievement (see Figure 3.2; Kenny, 1994; McAbee & Connelly, 2016).These notions are shown in a conceptual representation in Figure 3.2. Kenny and West (2010) investigated in a meta-analysis of 24 round-robin studies how the target and perceiver effects were represented in self-ratings (*k* and *q* in Figure 3.2). They found that regression coefficients of self-ratings regressed on target and perceiver effect were significantly positive with an average for *k* of $M = 0.79$, $SD = 0.40$ and an average for *q* of $M = 0.64$, $SD = 0.57$. Thus, they concluded, "there are several pieces of evidence that self-perception and the perception of others are not very different" (p. 208).

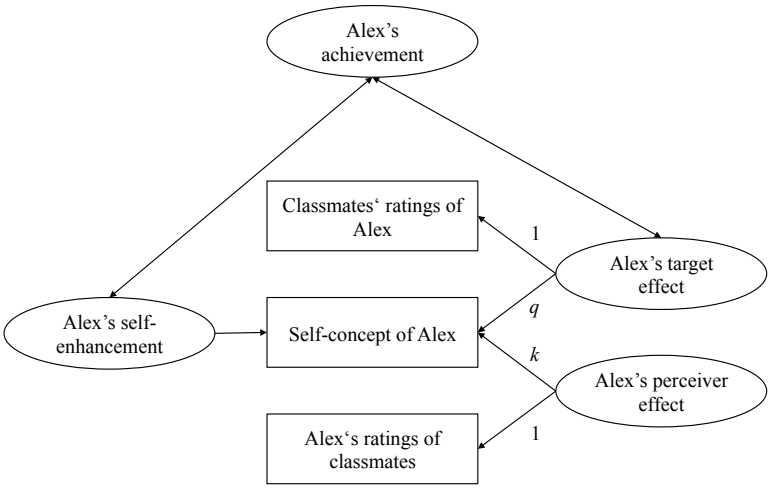

Figure 3.2. Conceptual representation of variance decomposition in the interpersonal approach exemplified from the individual perspective of Alex including objective achievement (Kwan et al., 2004; McAbee & Connelly, 2016).

Finally, in accordance with Kwan et al. (2004), the part of students' math self-concept rating that is not explained by the target and perceiver effects (i.e., the residual) can meaningfully be interpreted as a measure of *self-enhancement* in math (e.g., How competent does Alex think that he in particular is in math?). Thereby, self-enhancement is referred to as the "desire to maintain or increase the positivity (or decrease the negativity) of one's self-concept" (Leary, 2007, p. 319). As can be seen in Figure 3.2, the interpersonal approach disentangles a student's self-concept rating into three different components: a student's perceiver effect, target effect, and self-enhancement.

3.1.2 Associations between Academic Self-concept and Peer Rated Competence

The interpersonal approach connects self-concept ratings with peer ratings of competence (Kwan et al., 2004). Two results corroborate this connection. First, theories of educational psychology describe the potential effect of peer' ratings of competence on students' self-concept (Usher & Pajares, 2008; Wigfield & Eccles, 2000). In line with these notions, several studies could show that peer ratings of competence predicted students academic self-concept controlling for initial academic self-concept (Chen et al., 2010; Gest et al., 2005; Hughes et al., 2009).

Second, both peer ratings of competence and academic self-concept are associated with achievement. It is well known that academic self-concept is reciprocally related to achievement (Valentine et al., 2004). That is, achievement predicts academic self-concept and in turn, academic self-concept predicts achievement (Huang, 2011; Marsh & Craven, 2006). Few studies investigated the association between achievement and peer ratings of competence. Nevertheless, they found a consistent pattern: achievement predicted gains in peer ratings of competence and peer ratings of competence predicted gains in achievement (Chen et al., 2010; Denissen et al., 2011; Gest et al., 2008; Hughes et al., 2009). In sum, both academic self-concept and peer ratings of competence were correlated and they had a similar association with achievement. These results might support the assumption of the interpersonal approach that both ratings are functions of perceiver effect and target effect.

3.1.3 The Effect of Self-enhancement on Academic Achievement

In the framework of the interpersonal approach, the question arises whether the idiosyncratic way of perceiving oneself as high achieving (i.e., self-enhancement) predicts future achievement or if it is exclusively an effect of the shared perception. In educational psychology, there are two theoretical positions on the effect of self-enhancement.

First, prominent motivational theories like expectancy-value theory highlight the driving force of a positive competence rating (including academic self-concept) for achievement (Marsh & Craven, 2006; Wigfield & Eccles, 2000). Accordingly, one would expect self-enhancement to result in higher performance, for instance, as mediated by more effort and greater persistence on a task (Marsh & Martin, 2011; Wigfield & Eccles, 2000). There is some support for a motivational benefit associated with self-enhancement (Côté, Bouffard, & Vezeau, 2014; Praetorius, Kastens, Hartig, & Lipowsky, 2016).

Second, calibration research has highlighted the effect of an accurate competence rating on achievement (Alexander, 2013; de Bruin & van Gog, 2012). A fit between competence ratings and competence (i.e., calibration) is seen as a positive outcome that enables self-regulated learning and thus cognitive control over a study task (Bol & Hacker, 2012; Dunlosky & Lipko, 2007). Accordingly, one would expect that self-enhancement (i.e., a competence rating that is not well-calibrated) should result in an inadequate choice of learning tasks, a lack of suitable study time allocation, and finally, worse subsequent achievement (Metcalfe, 2009). There is some

evidence that a well-calibrated academic self-concept is associated with higher achievement (Chiu & Klassen, 2009, 2010).

3.1.4 The Present Investigation

In the present study, we applied the interpersonal approach to identify the connection between academic self-concept and peer ratings of competence and to test how an idiosyncratic self-rating (i.e., self-enhancement) was associated with achievement. First, we examined the overlap in self-concept ratings and peer rating of competence in math. Second, we tested whether this overlapping variance (i.e., the target effect) or rather self-enhancement would be associated with actual achievement.

We expected a positive correlation between self-concept ratings and the target effect indicating an overlap in variance. Furthermore, we expected that the target effect would be positively associated with achievement, as indicated by prior results regarding academic self-concept and peer ratings of competence (Gest et al., 2008; Marsh & Craven, 2006). Yet based on theoretical and empirical considerations, no clear predictions could be derived for the association between self-enhancement and achievement. One could expect a positive effect (Marsh & Craven, 2006), a negative effect (Chiu & Klassen, 2010), or no effect at all.

We tackled these expectations with two steps. First we studied cross-sectional correlations between achievement and interpersonal competence ratings (as outlined in Figure 3.2). Second, we tested the effect of academic self-concept and the interpersonal components on achievement longitudinally, that is, controlling for prior achievement. Because academic self-concept is a domain-specific self-rating (Marsh & Craven, 2006), and math is one of the most common academic domains in self-concept research, we focused on academic self-concept and peer ratings of competence in math. To add to earlier studies in the field of interpersonal perception (Denissen et al., 2011; Kwan et al., 2004) or peer ratings of competence (Chen et al., 2010; Gest et al., 2005, 2008; Hughes et al., 2009), we used (a) a large sample of school students in two age groups in the fifth and eighth grades over the course of 1 year, (b) domain-specific competence ratings by the self and by peers, and (c) grades and standardized achievement tests as achievement measures.

3.2 Method

3.2.1 Sample

Our study used data from the German TRAIN study, which is a longitudinal school study of two age cohorts that focuses on the achievement, motivation, and well-being of school students who are in the nonacademic school tracks (Jonkmann, Rose, & Trautwein, 2013). We used a subsample from the study that consisted of classes in which the collection of round-robin data was possible. This subsample comprised two cohorts of school students from the nonacademic school track in Saxony (*Mittelschule*). At the first measurement point, the younger cohort was in Grade 5, and the older cohort was in Grade 8. We investigated students' development from the first wave of measurement in 2008 to the second wave of measurement in 2009 (t_1 and t_2).

Altogether, our target sample consisted of 2,094 students of which 1,133 were fifth graders and 961 were eighth graders. We had to exclude 539 students who had no round-robin data (e.g., because they were absent on the day of measurement or they did not have parental consent to participate in the round-robin measurement). The average level of students' participation in the round-robin measurement per classroom was 64.32% (*SD* = 14.73%) for the fifth graders and 57.47% (*SD* = 17.31%) for the eighth graders. Furthermore, we excluded six students in two classes in which fewer than five students participated in the round-robin measurement. This led to a final sample of 1,549 students in 87 classes (46.66% female) with 846 fifth graders in 46 classes (46.33% female, mean age 11.11, *SD* = 0.54) and 703 eighth graders in 41 classes (47.06% female, mean age 14.30, *SD* = 0.65).

3.2.2 Procedure

Measurement was conducted on 2 consecutive days for each wave of measurement. Trained research assistants visited intact classrooms and administered all measures during regular school hours. Only students with parental consent participated in the study. Besides the measures reported below, other measures that were not included in the present study were also administered.

3.2.3 The Social Relations Model

The SRM is a standard method of interpersonal perception (Kenny et al., 2006). It attributes variance in interpersonal ratings (like in a round-robin design, see Figure 3.1) to three sources: the perceiver effect, target effect, and relationship effect. Furthermore, multiple groups (e.g., classes) can differ in their average ratings as represented by a constant term (e.g., Alex and his classmates think of each other that they are more competent than students in another class think of each other). Overall, the SRM may be seen as a mixed-effects model with three random effects (class-average, perceiver effect and target effect) and a residual (relationship effect; Nestler, 2015). Assumptions of variance decomposition in the SRM can be represented by the following equation (Kenny, 1994):

$$x_{ijh} = \mathrm{m}_h + u_{ih} + v_{jh} + \mathrm{e}_{ijh}, \quad (1)$$

where x_{ijh} is the math competence rating that student i (e.g., Alex) assigns to student j (e.g., Beth) in group h (e.g., Alex and Beth's class), m_h is the mean rating within group h, u_{ih} is i's *perceiver effect*, v_{jh} is j's *target effect*, and e_{ijh} is the *relationship effect* of i rating j.

This equation can be adapted to represent academic self-concept ratings within the interpersonal approach as described in the introduction (Kwan et al., 2004). The adaption considers that academic self-concept ratings can be assumed to be a function of target and perceiver effects, but these effects might not be weighted like they are in peer ratings of competence (cf. Kenny, 1994, p. 185). This weighting is represented by the parameters k and q. On the basis of these considerations, Equation 1 can be adapted to represent assumptions about variance decomposition of math self-concept ratings (Kenny & West, 2010):

$$x_{iih} = c_h + k \times u_{ih} + q \times v_{ih} + \mathrm{e}_{iih} \quad (2)$$

where x_{iih} is the math self-concept rating of student i (e.g., Alex), c_h is a constant equal to the mean of all of the math self-concept ratings in group h (e.g., Alex's class), u_{ih} and v_{ih} are the perceiver and target effects of student i, respectively. Furthermore, k and q are regression coefficients that describe the extent to which the target and perceiver effects are reflected in math self-concept ratings (Kenny, 1994). Finally, e_{iih} is the part of student i's math self-concept rating that is not explained by the target and perceiver effects, which is defined as a measure of self-enhancement. This equation is graphically represented in Figure 3.2.

Additionally, in the SRM, the effects' variances and covariances describe rating behaviors on the group level (Kenny, 1994). *Assimilation* (the variance of the perceiver effect)

describes whether a student rates his/her peers' competence in a less differentiated and more uniform way (e.g., Does Alex rate Berta's competence similar to how Alex rates Charlie's competence?). *Consensus* (the variance of target effects) describes whether peers agree in their rating of a student (e.g., Does Berta rate Alex's competence like Charlie rates Alex's competence?). Taken together, assimilation and consensus describe the degree of differentiation in interpersonal math competence ratings (Malloy, Sugarman, Montvilo, & Ben-Zeev, 1995).

Reciprocity describes how the reciprocal ratings of students are correlated. There are two reciprocity measures. *Dyadic reciprocity* (correlation between the reciprocal ratings of two students) increases as the similarity in two students' ratings of each other's math competence increases (e.g., Does Alex rate Berta's competence like Berta rates Alex's competence?). *Generalized reciprocity* (the correlation between the target and perceiver effects) is higher when students who rate others as competent are also rated as competent in turn (e.g., Are Alex's ratings of his classmates' competence similar to his classmates' competence ratings of him?). Reciprocity was not the focus or our investigation, yet it needs to be controlled for valid estimates of target and perceiver effects.

3.2.4 Measures

Interpersonal competence ratings. Students' academic self-concept in math and peer ratings of math competence were assessed in a round-robin design. Students rated every classmate and themselves on the item "This student is good in math," which could be answered on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*very much*). We regarded the self-rating as a measure of academic self-concept in math (i.e., math self-concept rating). This was supported by a manifest correlation of $r = .53$ for the fifth graders and $r = .64$ for the eighth graders between the round-robin item and the usual four-item measure of academic self-concept in math that was also administered in this study (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005). Furthermore, Gogol et al. (2014) demonstrated that academic self-concept could be measured with sufficient psychometric quality using only single items. Altogether, we analyzed data from 13,255 dyads (where at least one student rated one other student) of which 7,309 were fifth graders and 5,946 were eighth graders. On average, fifth graders rated competence of $M = 15.60$, $SD = 5.74$ classmates and eighth graders rated competence of $M = 15.36$, $SD = 6.02$

classmates. On the basis of this round-robin data, we estimated students' target and perceiver effects (see the statistical procedure below).

Target effect. The target effect was an aggregate estimated within the SRM of how classmates rated a target student. The target effect differed from the mean of all ratings by classmates, as it was estimated while controlling for students' perceiver effect. High values indicated that the student was rated positively by classmates on average, low values indicated that the student was rated negatively by classmates on average.

Perceiver effect. The perceiver effect was an aggregate estimated within the SRM of how a target student rated classmates. The perceiver effect differed from the mean of all ratings by the student, as it was estimated while controlling for students' target effect. High values indicated that the student rated classmates positively on average, low values indicated that the student rated classmates negatively on average.

Self-enhancement. After estimating the target and perceiver effects, we computed students' self-enhancement as recommended by Church et al. (2014) and Kenny and West (2010). The computation of self-enhancement can be expressed by rearranging Equation 2:

$$e_{iih} = x_{iih} - c_h - k \times u_{ih} - q \times v_{ih}, \quad (3)$$

where x_{iih} is the math self-concept rating of student i, c_h is a constant equal to the mean of all math self-concept ratings in a class, u_{ih} and v_{ih} are the perceiver and target effects of student i, and e_{iih} is student i's self-enhancement in math. Furthermore, k and q are regression coefficients that describe the extent to which the target and perceiver effects are reflected in the math self-concept rating (Kenny & West, 2010). That is, self-enhancement is computed as the residual of the math self-concept rating regressed on an individual's target and perceiver effects.

Math achievement. Achievement was measured with grades and performance on a standardized test at t_1 and t_2. Grades were obtained from the school administration. When no grades could be obtained from these sources, grades reported by students were used (which was done in 20 cases). We used the most recent final grades in math at the respective measurement point, that is, at t_1, the final grades from Grades 4 and 7, respectively, and at t_2, the final grades from Grades 5 and 8, respectively. For the analyses, we recoded all grades so that high scores represented good grades.

The second measure of achievement was a standardized achievement test in math that consisted of curricularly valid items (Rose et al., 2013). The test was administered in a rotated

test booklet design (Mislevy, Beaton, Kaplan, & Sheehan, 1992) that allowed us to scale across the waves of measurement (Rose et al., 2013). We used WLE estimates of students' test performance. Their marginal reliability (which is comparable to Cronbach's α) was .77 at t_1 and .74 at t_2. The tests were administered during the 2 days of measurement at t_1 and t_2.

3.2.5 Statistical Procedure

We conducted all statistical analyses for both cohorts in identical ways. Our analyses consisted of three steps. In the first step, the basic parameters (i.e., variances and covariances) of the social relations model (SRM) for the interpersonal competence ratings were estimated. The SRM requires specialized software or code (e.g., Kenny, 1994; Nestler, 2015; Schönbrodt, Back, & Schmukle, 2012). We chose to use a flexible Bayesian approach (Lüdtke, Robitzsch, Kenny, & Trautwein, 2013) that allowed for an integration of achievement measures as covariates when estimating the SRM parameters (see below). We investigated assimilation and consensus as the relative variances of the target and perceiver effects as is typically done in SRM research (Kenny, 1994). In addition, we investigated dyadic reciprocity and generalized reciprocity. For all these parameters, we calculated the mean of the posterior distribution (expected a posteriori; EAP) as the point estimate and the 95% Bayesian credibility interval (BCI) as the interval estimate of the parameters.

For the Bayesian estimation, we used the software WinBUGS accessed via R and the package R2winBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; R Core Team, 2015; Sturtz, Ligges, & Gelman, 2005). In accordance with Lüdtke et al. (2013), we chose noninformative priors for all parameters. Thus, the parameter estimates were based primarily on observed data and can be interpreted like maximum-likelihood estimates (Nestler, 2015). We estimated the SRM model separately in the two cohorts.

To correct for unreliability of the target and perceiver effects (see Bonito & Kenny, 2010) when estimating the associations with the achievement measures and the math self-concept ratings, we extended the SRM approach by applying the multiple imputation procedure from the missing data literature (Yang & Seltzer, 2015; see also Enders, 2010). That is, we generated multiple imputations (also known as plausible values; see Asparouhov & Muthén, 2010) for the target and perceiver effects. To do this, we specified a background model in the estimation of the SRM by including all grades, achievement tests, and the math self-concept ratings as predictors

of the individual target and perceiver effects. We used this approach to correctly estimate how the target and perceiver effects were correlated with the achievement measures and the math self-concept ratings in the following analyses (see Yang & Seltzer, 2015). For additional statistical analyses, we drew 50 plausible values from the posterior distributions of the target and perceiver effects (see Bodner, 2008). We aggregated statistical estimates by applying Rubin's rule (Rubin, 1987) via the R packages miceadds for the correlations (Robitzsch, 2015) and procedures to handle multiply imputed data integrated in Mplus 7.1 for regression analyses (Muthen & Muthen, 2012). In the regression analyses, we corrected standard errors for the clustered data (students clustered in classes).

In the second step, we estimated cross-sectional associations between the measures. To identify overlap in variance between self-concept ratings and peer ratings of competence in math, we calculated bivariate correlations between self-concept ratings with target and perceiver effects. Furthermore, we regressed math self-concept ratings on target and perceiver effect for estimates of parameters k and q (see Figure 3.2 and Equation 2). Lastly, we calculated correlations between achievement measures with math self-concept ratings, target effect, perceiver effect, and self-enhancement to identify which component of competence ratings was associated with achievement.

In the third step, we predicted gains in math achievement over the course of 1 year from the math self-concept ratings, the perceiver effect, the target effect, and self-enhancement. To test for these effects, we specified two multivariate regression models in Mplus 7.1 (Muthen & Muthen, 2012). We used multivariate analyses because the achievement measures were assumed to be correlated. The t_2 achievement measures (i.e., grades and test scores) were the dependent variables that were regressed on the respective predictors and the matching t_1 achievement measure. In the first model, we used math self-concept ratings as predictor and in the second model we used perceiver effect, target effect, and self-enhancement as predictor. The second model was intended to show which component of math self-concept ratings predicted achievement.

3.3 Results

3.3.1 Variance Decomposition in the SRM

In the first step we decomposed variance of interpersonal ratings in terms of assimilation, consensus, and reciprocity. Following, we summarize the main results of the SRM estimation. In Supplement 1, the results are described in more detail.

Students showed substantive assimilation and consensus. That is, students differed interindividually with regard to their ratings of their peers' math competence as indicated by assimilation. Yet at the same time, they showed substantial consensus, which allowed for a meaningful interpretation of target effects. However, there were differences between cohorts. The eighth graders showed a lower level of assimilation and a higher level of consensus. That is, their ratings were more differentiated and classmates agreed more with each other.

3.3.2 Cross-sectional Associations

In Table 3.1, we investigated how perceiver and target effects were represented in math self-concept ratings using correlations and regression parameters k and q. The parameters k and q show the degree to that self-concept ratings are a function of perceiver and target effect (see Figure 3.2). Values equal to 1 would indicate that perceiver effect and target effect are weighed in math self-concept ratings as they are weighed in peer ratings of competence.

For the fifth graders, the two effects were equally associated with math self-concept ratings with $r = .33$ for the perceiver effect and $r = .31$ for the target effect, $z = 0.46$, $p = .647$. Entering the perceiver effect and the target effect in a regression simultaneously yielded $k = 0.70$, $p < .001$ (perceiver effect as predictor) and $q = 0.69$, $p < .001$ (target effect as predictors). For the fifth graders, the explained variance in math self-concept ratings by perceiver and target effect was $R^2 = 0.22$. For the eighth graders, the correlation between perceiver effect and math self-concept ratings was $r = .26$. For the eight graders compared with the fifth graders, the target effect had a higher correlation with math self-concept ratings ($r = .51$, $z = 4.74$, $p < .001$). Entering the perceiver effect and the target effect in a regression simultaneously yielded $k = 0.84$, $p < .001$ (perceiver effect as predictor) and $q = 0.94$, $p < .001$ (target effect as predictors). For the

eighth graders, the explained variance in math self-concept ratings by perceiver and target effect was R^2 = 0.37. Overall, the results showed that math self-concept ratings were similarly a function of perceiver and target effect as peer ratings were a function of target and perceiver effect.

Table 3.1

Weighting of Perceiver Effect and Target Effect in Math Self-concept Ratings

	Fifth graders (n = 846)		Eighth graders (n = 703)	
	r	b (SE)	r	b (SE)
Perceiver effect (k)	.33*	0.70* (0.08)	.26*	0.84* (0.09)
Target effect (q)	.31*	0.69* (0.08)	.51*	0.94* (0.07)
R^2		0.22		0.37

Note. r = correlation coefficient; b = unstandardized regression coefficient of predicting math self-concept ratings by perceiver effect and target effect as explained by Kenny and West (2010); SE = standard error. The expressions k and q refer to Equations (2) and (3) as well as Figure 3.2.

In Table 3.2, we aimed to disentangle the cross-sectional association between math self-concept ratings and achievement as indicated in Figure 3.2. Accordingly, we present correlations between math self-concept ratings and perceiver effect, target effect, and self-enhancement with the achievement measures at t_1. In both cohorts, the two achievement measures were significantly correlated with each other with a moderate effect size (r = .48 and r = .40, respectively), indicating that grades and test scores had some overlap but also measured different aspects of achievement.

As a basis for further investigations, we tested whether math self-concept ratings were associated with achievement (as it is usually found). Math self-concept ratings were correlated with both achievement measures with stronger associations for the eighth graders. Especially for the eighth graders, the correlation between grades and the math self-concept ratings (r = .45) was higher than the correlation between the achievement test and math self-concept ratings (r = .35, z = 2.23, p = .026). This indicated that the older students seemed to rely more on their grades than

on the standardized achievement test to form their math self-concept. Overall, math self-concept ratings were correlated with achievement, so we could investigate which component would show this correlation as well.

Table 3.2

Correlations between Achievement at t_1 with Math Self-Concept Ratings and their Components

	Fifth graders (n = 846)		Eighth graders (n = 703)	
	Test	Grade	Test	Grade
Grade	.48*		.40*	
Math self-concept ratings	.23*	.28*	.35*	.45*
Perceiver effect	-.10*	-.06	-.11*	.02
Target effect	.52*	.56*	.56*	.64*
Self-enhancement	.12*	.11*	.15*	.09*

Note. Test = Performance on the standardized math achievement test at t_1; Grade = Final grade in math in the previous year (i.e., at t_1), grades were recoded so that high scores represent high achievement.

* $p < .05$.

The perceiver effect was negatively correlated with achievement on the standardized test in both cohorts (e.g., $r = –.10$ for the fifth graders). That is, students with high achievement had a slight tendency to rate their peers as less competent. This correlation was significant only for the achievement test but not for grades.

Conversely, we expected the target effect representing shared competence ratings to correlate positively with actual achievement. The target effect was strongly associated with both achievement measures. In fact, the associations between the target effect and the achievement measures were stronger than the associations between the math self-concept ratings and the achievement measures. For instance, for fifth graders, the correlation between the target effect and the achievement test was $r = .52$ and the correlation between the math self-concept ratings and the achievement test was $r = .23$, $z = 7.02$, $p < .001$. Similarly, the correlations between the target effect and the achievement measures were stronger than the interrelation between the two

achievement measures. The correlations between the target effect and achievement were even higher for the eighth graders (e.g., $r = .56$ for the correlation between the target effect and the achievement test and $r = .64$ for the correlation between the target effect and math grades). Overall, the target effect was highly correlated with students' achievement.

Finally, we tested whether self-enhancement was positively or negatively associated with achievement. Self-enhancement had a relatively small but significant positive association with achievement. For instance, the correlations between self-enhancement and the achievement test were $r = .12$ for fifth graders and $r = .15$ for eighth graders, $z = 0.59$, $p = .553$. There were only minor differences between cohorts and achievement measures. That is, higher achieving students also had a more positive self-enhancement, which supported the propositions put forward by motivational theories.

3.3.3 Predicting Math Achievement

Finally, we addressed the question of which part of math self-concept ratings would predict achievement longitudinally. Table 3.3 shows the results of the regression models in which we predicted achievement at t_2 by math self-concept ratings, the target effect, the perceiver effect, self-enhancement, and achievement at t_1.

As expected, the math self-concept ratings positively predicted an increase in achievement for the fifth graders ($b = 0.06$, $p = .012$ for the achievement test and $b = 0.16$, $p < .001$ for grades) and the eighth graders ($b = 0.11$, $p = .004$ for the achievement test and $b = 0.26$, $p < .001$ for grades). Students who rated themselves as more competent had higher gains in achievement 1 year later.

Table 3.3

Prediction of Achievement at t_2 by Math Self-Concept Ratings and their Components while Controlling for Achievement at t_1

	Achievement measure							
	Fifth graders (n = 846)				Eighth graders (n = 703)			
	Test t_2	Grade t_2	Test t_2	Grade t_2	Test t_2	Grade t_2	Test t_2	Grade t_2
	Model 1a		Model 2a		Model 1b		Model 2b	
Test t_1	0.65* (0.02)		0.57* (0.03)		0.55* (0.04)		0.43* (0.04)	
Grade t_1		0.47* (0.03)		0.21* (0.03)		0.49* (0.05)		0.24* (0.06)
Math self-concept ratings	0.06* (0.02)	0.16* (0.03)			0.11* (0.04)	0.26* (0.04)		
Perceiver effect			-0.02 (0.03)	0.00 (0.04)			-0.03 (0.03)	0.02 (0.03)
Target effect			0.16* (0.04)	0.51* (0.04)			0.29* (0.04)	0.56* (0.05)
Self-enhancement			0.03 (0.03)	0.08* (0.02)			0.02 (0.04)	0.10* (0.03)
R^2	.44	.28	.46	.44	.35	.42	.41	.56

Note. The table displays the standardized regression coefficients and standard errors from the multivariate regression models for predicting achievement at t_2; Test = Performance on the standardized math achievement test; Grade = Final grade in math in the year of measurement; grades were recoded so that high scores represent high achievement.

* $p < .05$.

The interpersonal approach allowed for a closer examination of these effects. The perceiver effect did not predict achievement at t_2 over and above achievement at t_1. By contrast, the target effect had a positive effect on achievement at t_2 over and above achievement at t_1. This effect was especially pronounced for grades (b = 0.51, $p < .001$ for the fifth graders and b = 0.56, $p < .001$ for the eighth graders). That is, students who were rated as competent in math had higher gains in achievement after 1 year. Finally, self-enhancement did not predict achievement increases in the standardized test, but it positively predicted an improvement in grades (b = 0.08, $p = .001$ for the fifth graders and b = 0.10, $p = .001$ for the eighth graders). Students who overestimated their competence had a slight improvement in their grades after 1 year. This result is in line with the propositions of motivational theories regarding self-enhancement, yet the effect was small. Overall, in interpreting the effects of the components of math self-concept ratings, the target effect emerged as the component of academic self-concept that most likely drove its positive effect on achievement gains.

3.4 Discussion

In this study, we investigated academic self-concept in math from an interpersonal perspective to disentangle associations between self-concept ratings, peer ratings of competence, and achievement in math. We studied (a) overlap in variance between self-concept ratings and peer ratings of competence in math (b) the cross-sectional correlations between math self-concept ratings, the perceiver effect, the target effect, self-enhancement, and achievement and (c) how math self-concept ratings, the perceiver effect, the target effect, and self-enhancement predicted future math achievement. Math self-concept ratings had moderate to strong associations with both the target and perceiver effects and furthermore with achievement. In addition, the target effect was particularly associated with achievement. Both the math self-concept ratings and the target effect predicted achievement gains. Self-enhancement had a small positive association with achievement and had a small positive effect on achievement gains.

3.4.1 Associations between Components of Competence Ratings and Achievement

By using the interpersonal approach, we were able to connect academic self-concept to peer ratings of competence. We adapted the approach of Kwan et al. (2004) and posited that not only peer ratings of competence but also academic self-concept would be a function of target and perceiver effects. Our results substantiated this notion: In their meta-analysis of round-robin studies, Kenny and West (2010) found that regression coefficients to predict self-ratings by perceiver and target effect were on average $k = 0.79$ for the perceiver effect and $q = 0.64$ for the target effect. In our study, both parameters had a similar or larger size (for fifth graders: $k = 0.70$, $q = 0.69$; for eighth graders: $k = 0.84$, $q = 0.94$), indicating that self-concept ratings were similarly a function of perceiver and target effect as peer ratings of compentence were a function of target and perceiver effect.

Furthermore, the components were differently correlated with achievement. The perceiver effect was rather unrelated to achievement and rather represented students' tendency to rate competence. In terms of a multi-trait multi-method framework (Campbell & Fiske, 1959), the perceiver effect might thus be labeled a "method factor" of interpersonal ratings (cf. McAbee & Connelly, 2016). In contrast, the target effect was highly correlated to achievement. It represented a perception of who is competent in a class that was shared between students themselves and their classmates. In terms of multi-trait multi-method analyses, the target effect might thus be called a "trait factor" of interpersonal ratings that highly correlated with its behavioral measure, that is, actual achievement. Finally, students' self-enhancement had a small correlation with achievement. That is, students' specific self-ratings were associated with achievement as well, but with a smaller effect size.

Overall, our results contribute to the construct validity of academic self-concept and peer ratings of competence. Both constructs overlapped in their variance and mainly this shared variance was associated with achievement. Self-enhancement as a students idiosyncratic self-rating had incremental validity with regard to the association with achievement.

3.4.2 The Target Effect as Predictor of Achievement

Our study further supports existing evidence that positive competence ratings predict improvements in achievement. We found that math self-concept predicted achievement, but moreover, especially the target effect predicted future achievement over and above prior achievement. Therein, the target effect represented a positive rating that was shared between students themselves and their peers. This result may connect research on academic self-concept (Huang, 2011; Marsh & Craven, 2006; Marsh & Martin, 2011; Valentine et al., 2004) and on peer ratings of competence (Chen et al., 2010; Gest et al., 2008; Hughes et al., 2009) that both show a positive effect of competence ratings on achievement gains. Moreover, students that were perceived as higher achieving by their parents (e.g., Castro et al., 2015) or by teachers (e.g., Jussim & Harber, 2005) showed improvements in achievement as well. This pattern of results – including our study - could indicate, that positive ratings by peers, parents, teachers, and students themselves are expressions of a shared underlying trait in a multi-rater framework that represents a positive competence perception (McAbee & Connelly, 2016). Overall, the positive effect of the target effect on achievement controlling for prior achievement contributes to the predictive validity of the interpersonal approach.

3.4.3 Age Differences in Competence Ratings

Our results showed that correlations between math self-concept ratings, target effect and achievement were stronger for the eighth graders than for the fifth graders. That is, competence ratings of different sources fell more in line with each other and with objective achievement (cf. Denissen, Zarrett, & Eccles, 2007). Although students' academic self-concept is known to decrease over time in school (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Watt, 2004), academic self-concept may at the same time be seen a getting more realistic in terms of its correlation with objective achievement (cf. Stipek & Mac Iver, 1989). So far, studies have shown this trend for self-ratings of motivation and achievement. Our study adds to this research because it indicates that this increase in correlations with achievement seems to extend to peer ratings of competence (Malloy, Yarlas, Montvilo, & Sugarman, 1996).

3.4.4 The Effect of Self-enhancement on Achievement

In the introduction, we presented two theoretical accounts that proposed different effects of self-enhancement. On the one hand, self-enhancement might improve achievement through the driving force of a positive self-view (Marsh & Craven, 2006; Wigfield & Eccles, 2000). On the other hand, self-enhancement might hamper achievement through a lack of cognitive control over the learning process. Our results support the first explanation.

However, the effect of self-enhancement was rather small. This might be the result of the two proposed effects canceling each other out. Self-enhancement might result in higher motivation and thus more persistence in learning tasks. But due to a suboptimal choice of learning behavior, the learning might be inefficient. These two effects might result in the overall small net effect we observed (for a similar effect see Murayama & Elliot, 2012).

Moreover, in our study, we investigated the effect of self-enhancement on a larger scale than is usually the case in calibration research (de Bruin & van Gog, 2012). It is unclear whether the effect of self-enhancement across a period of 1 year in schools is the same as the effect of self-enhancement over a period of 1 hour in a lab setting. That is, there is a need for more research to compare these two settings and explain why differential results might be expected.

3.4.5 Limitations and Directions for Future Research

One limitation concerns the measurement of self-concept ratings and peer ratings of achievement in math. Due to time constraints in the study (Jonkmann et al., 2013), both were measured using one item for each target students, that is, one student rated her- or himself with one item and each classmate with one item. There is evidence that single items may have sufficient psychometric quality regarding self-reports (Gogol et al., 2014) and peer reports (Denissen, Geenen, Selfhout, & van Aken, 2008). Nevertheless, reliability and validity of the ratings and the derived measures may have suffered from using only single items. Furthermore, with math self-concept measured with only one item, self-enhancement was confounded with measurement error (Kenny, 1994; Kwan et al., 2004). This confounding may have led to an underestimation of the effect of self-enhancement. Altogether, in future studies it would be advisable to measure both, self-concept ratings as well as peer ratings of competence with more than one item.

Another limitation regards the sample. Although the sample was relatively large, it was nevertheless restricted. The study came from the non-academic school track from one federal

state of Germany. Due to the school track, students were lower achieving and potentially they had a lower motivation than students from the academic school track. Moreover, students in the federal state of Saxony could differ from students in other federal states and they potentially differed from students outside of Germany. Altogether, it is not clear whether results would generalize to other school tracks and to other countries or cultures. That is, more research is needed that investigates round-robin ratings in diverse school settings. Nevertheless, our results are fairly consistent with prior studies in self-concept research and research on peer ratings of competence (Gest et al., 2008; Marsh & Craven, 2006), which could be seen as an indicator of generalizability.

Finally, a major use of the interpersonal approach might lie in an extended investigation of academic self-concept effects from a multi-rater perspective. There is a vast literature on the structure of academic self-concept (Marsh & Craven, 2006), gender differences in academic self-concept (Jansen, Schroeders, & Lüdtke, 2014), and effects that contribute to the formation of academic self-concept such as the *big-fish-little-pond effect* (BFLPE; e.g., Pinxten et al., 2015) or the *internal/external frame of reference model* (e.g., Niepel, Brunner, & Preckel, 2014). It might be fruitful to examine these effects within an interpersonal conceptualization of academic self-concept. Peer ratings might be similarly subject to reference effects like academic self-concept. To better understand reference effects, it might be interesting to contrast reference effects between self- and peer ratings.

3.4.6 Conclusion

Altogether, the interpersonal approach connects academic self-concept and peer ratings of competence. Because we could show that the target effect correlated substantially with math self-concept ratings and predicted gains in achievement, the target effect may be seen as perceived competence “trait” that underlies peer ratings of competence as well as academic self-concept

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. http://doi.org/10.1016/j.learninstruc.2012.10.003

Asparouhov, T., & Muthén, B. (2010). *Plausible values for latent variables using Mplus*. *Mplus Technical Report*. Retrieved from http://www.statmodel.com/download/Plausible.pdf

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(4), 651–675. http://doi.org/10.1080/10705510802339072

Bol, L., & Hacker, D. J. (2012). Calibration research: where do we go from here? *Frontiers in Psychology*, *3*(July), 229. http://doi.org/10.3389/fpsyg.2012.00229

Bonito, J. A., & Kenny, D. A. (2010). The measurement of reliability of social relations components from round-robin designs. *Personal Relationships*, *17*(2), 235–251. http://doi.org/10.1111/j.1475-6811.2010.01274.x

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. http://doi.org/10.1037/h0046016

Castro, M., Expósito-Casas, E., López-Martín, E., Lizasoain, L., Navarro-Asencio, E., & Gaviria, J. L. (2015). Parental involvement on student academic achievement: A meta-analysis. *Educational Research Review*, *14*, 33–46. http://doi.org/10.1016/j.edurev.2015.01.002

Chen, Q., Hughes, J. N., Liew, J., & Kwok, O.-M. (2010). Joint contributions of peer acceptance and peer academic reputation to achievement in academically at-risk children: Mediating processes. *Journal of Applied Developmental Psychology*, *31*(6), 448–459. http://doi.org/10.1016/j.appdev.2010.09.001

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. http://doi.org/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. http://doi.org/10.1016/j.learninstruc.2008.11.002

Church, A. T., Katigbak, M. S., Mazuera Arias, R., Rincon, B. C., Vargas-Flores, J. D. J., Ibáñez-Reyes, J., … Ortiz, F. a. (2014). A four-culture study of self-enhancement and adjustment using the social relations model: Do alternative conceptualizations and indices make a difference? *Journal of Personality and Social Psychology*, *106*(6), 997–1014. http://doi.org/10.1037/a0036491

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*, *8*, 415–434. http://doi.org/10.1111/bjep.12045

Cronbach, L. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, *52*(3), 177–193. http://doi.org/10.1037/h0044919

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. http://doi.org/10.1016/j.learninstruc.2012.01.003

Denissen, J. J. A., Geenen, R., Selfhout, M., & van Aken, M. A. G. (2008). Single-item big five ratings in a social network design. *European Journal of Personality*, *22*(1), 37–54. http://doi.org/10.1002/per.662

Denissen, J. J. A., Schönbrodt, F. D., van Zalk, M., Meeus, W. H. J., & van Aken, M. A. G. (2011). Antecedents and consequences of peer-rated intelligence. *European Journal of Personality*, *25*(2), 108–119. http://doi.org/10.1002/per.799

Denissen, J. J. A., Zarrett, N. R., & Eccles, J. S. (2007). I Like to Do It, I'm Able, and I Know I Am: Longitudinal Couplings Between Domain-Specific Achievement, Self-Concept, and Interest. *Child Development*, *78*(2), 430–447. http://doi.org/10.1111/j.1467-8624.2007.01007.x

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*(4), 228–232. http://doi.org/10.1111/j.1467-8721.2007.00509.x

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. http://doi.org/10.1177/0146167295213003

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology*, *97*(3), 337–346. http://doi.org/10.1037/0022-0663.97.3.337

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology*, *44*(3), 625–636. http://doi.org/10.1037/0012-1649.44.3.625

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., … Preckel, F. (2014). "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*(3), 188–205. http://doi.org/10.1016/j.cedpsych.2014.04.002

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, *49*(5), 505–528. http://doi.org/10.1016/j.jsp.2011.07.001

Hughes, J. N., Dyer, N., Luo, W., & Kwok, O.-M. (2009). Effects of peer academic reputation on achievement in academically at-risk elementary students. *Journal of Applied Developmental Psychology*, *30*(2), 182–194. http://doi.org/10.1016/j.appdev.2008.12.008

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. http://doi.org/10.1111/1467-8624.00421

Jansen, M., Schroeders, U., & Lüdtke, O. (2014). Academic self-concept in science: Multidimensionality, relations to achievement measures, and gender differences. *Learning and Individual Differences*, *30*, 11–21. http://doi.org/10.1016/j.lindif.2013.12.003

Jonkmann, K., Rose, N., & Trautwein, U. (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen*. Tübingen: Projektbericht an die Kultusministerien der Länder.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns

and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. http://doi.org/10.1207/s15327957pspr0902_3

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self-and other perception: A meta-analysis. *Personality and Social Psychology Review*, *14*(2), 196–213. http://doi.org/10.1177/1088868309353414

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*(1), 94–110. http://doi.org/10.1037/0033-295X.111.1.94

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology*, *58*(1), 317–344. http://doi.org/10.1146/annurev.psych.58.110405.085658

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. http://doi.org/10.1037/a0029252

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337. http://doi.org/10.1023/A:1008929526011

Malloy, T. E., Sugarman, D. B., Montvilo, R. K., & Ben-Zeev, T. (1995). Children's interpersonal perceptions: A social relations analysis of perceiver and target effects. *Journal of Personality and Social Psychology*, *68*(3), 418–426. http://doi.org/10.1037/0022-3514.68.3.418

Malloy, T. E., Yarlas, A., Montvilo, R. K., & Sugarman, D. B. (1996). Agreement and accuracy in children's interpersonal perception: A social relations analysis. *Journal of Personality and Social Psychology*, *71*(4), 692–702. http://doi.org/10.1037/0022-3514.71.4.692

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163.

http://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. http://doi.org/10.1348/000709910X503501

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. http://doi.org/10.1207/s15326985ep2003_1

McAbee, S. T., & Connelly, B. S. (2016). A Multi-Rater Framework for Studying Personality: The Trait-Reputation-Identity Model. *Psychological Review*. http://doi.org/10.1037/rev0000035

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. http://doi.org/10.1111/j.1467-8721.2009.01628.x

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161. http://doi.org/10.1111/j.1745-3984.1992.tb00371.x

Murayama, K., & Elliot, A. J. (2012). The competition–performance relation: A meta-analytic review and test of the opposing processes model of competition and performance. *Psychological Bulletin*, *138*(6), 1035–1070. http://doi.org/10.1037/a0028324

Muthen, L., & Muthen, B. (2012). *Mplus User's Guide*. *Los Angeles, CA: Muthén & Muthén* (Seventh Ed). Los ANgeles, CA: Muthén & Muthén.

Nestler, S. (2015). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*. http://doi.org/10.1007/s11336-015-9474-9

Niepel, C., Brunner, M., & Preckel, F. (2014). The longitudinal interplay of students' academic self-concepts and achievements within and across domains: Replicating and extending the reciprocal internal/external frame of reference model. *Journal of Educational Psychology*, *106*(4), 1170–1191. http://doi.org/10.1037/a0036307

Pinxten, M., Wouters, S., Preckel, F., Niepel, C., De Fraine, B., & Verschueren, K. (2015). The formation of academic self-concept in elementary education: A unifying model for external and internal comparisons. *Contemporary Educational Psychology*, *41*, 124–132. http://doi.org/10.1016/j.cedpsych.2014.12.003

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? *Zeitschrift Für*

Entwicklungspsychologie Und Pädagogische Psychologie, *48*(1), 14–26. http://doi.org/10.1026/0049-8637/a000140

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Robitzsch, A. (2015). miceadds: Some additional multiple imputation functions, especially for "mice." Retrieved from http://cran.r-project.org/package=miceadds

Rose, N., Jonkmann, K., Hübner, N., Sälzer, C., Lüdtke, O., & Nagy, G. (2013). Durchführung und methodische Grundlagen der TRAIN-Studie. In K. Jonkmann, N. Rose, & U. Trautwein (Eds.), *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen*.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9780470316696

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: an R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–70. http://doi.org/10.3758/s13428-011-0150-4

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. http://doi.org/10.1026/0012-1924.51.4.183

Stipek, D., & Mac Iver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*(3), 521. http://doi.org/10.2307/1130719

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, *12*(21), 1–16. Retrieved from http://link.aps.org/doi/10.1103/PhysRevLett.44.1404

Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, *78*(4), 751–796. http://doi.org/10.3102/0034654308321456

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. http://doi.org/10.1207/s15326985ep3902_3

Watt, H. M. G. (2004). Development of Adolescents' Self-Perceptions, Values, and Task

Perceptions According to Gender and Domain in 7th- through 11th-Grade Australian Students. *Child Development*, *75*(5), 1556–1574. http://doi.org/10.1111/j.1467-8624.2004.00757.x

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

Yang, J. S., & Seltzer, M. (2015). Handling measurement error in predictors using a multilevel latent variable plausible values approach. In J. R. Harring, L. M. Stapelton, & S. N. Beretvas (Eds.), *Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications* (pp. 295–333). Information Age Publishing.

3.5 Supplement

Table S3.1 shows the SRM variance and covariance estimates. Students' assimilation and consensus are usually evaluated by inspecting relative variances (Kenny, 1994; Kenny & West, 2010). A higher relative perceiver variance indicates assimilation, whereas a higher relative target variance indicates consensus. The relative relationship variance is termed uniqueness, but this was not the focus of our study. As a rule of thumb, a relative variance that is larger than .10 expresses sufficient variation in the respective component (Kenny, 1994).

Table S3.1

Estimates of the SRM Parameters

	Fifth graders (n = 846)		Eighth graders (n = 703)	
	Est.	95% interval	Est.	95% interval
Assimilation	0.25	[0.23, 0.27]	0.16	[0.14, 0.18]
Consensus	0.22	[0.20, 0.25]	0.37	[0.34, 0.40]
Uniqueness	0.53	[0.51, 0.55]	0.47	[0.44, 0.49]
Dyadic reciprocity	0.05	[0.02, 0.08]	0.04	[0.01, 0.07]
Generalized reciprocity	-0.03	[-0.11, 0.06]	-0.10	[-0.18, 0.00]

Note. SRM = Social relations model; Est. = point estimate; assimilation is the relative variance of perceiver effects; consensus is the relative variance of target effects; uniqueness is the relative variance of relationship effects.

Among fifth graders, assimilation was quite high with an estimate of 0.25 and slightly larger than consensus with an estimate of 0.22. Students differed interindividually with regard to their ratings of their peers' math competence. However, for the eighth graders compared with the fifth graders, assimilation was smaller with an estimate of 0.16, and consensus was larger with an estimate of 0.37. That is, the eighth graders differentiated more between their peers' math competence, and they agreed more in their math competence ratings.

Dyadic reciprocity describes the correlation between reciprocal ratings. A higher dyadic reciprocity reflects the idea that two students rate each other's competence similarly. The dyadic

reciprocity was rather small with an estimate of 0.05 for the fifth graders and an estimate of 0.04 for the eighth graders; that is, they were not systematically associated.

Generalized reciprocity describes the correlation between the target effect and the perceiver effect across students. Generalized reciprocity was rather small for the fifth graders with an estimate of -0.03. For the eighth graders, the effect was slightly more negative with a point estimate of -0.10 but with a confidence interval that included 0. That is, there was a slight tendency for students who were rated as competent by their peers to rate their peers as less competent in turn. Altogether, the variance proportions as well as dyadic and generalized reciprocity were comparable in size to what is often found for peer ratings of personality traits (Kenny, 1994).

4

Perspective Matters: The Internal/External Frame of Reference Model for Self- and Peer Ratings of Achievement

Lösch, T., Kelava, A., Nagengast, B., Trautwein, U., & Lüdtke, O. (submitted). Perspective Matters: The Internal/External Frame of Reference Model for Self- and Peer Ratings of Achievement. *Learning and Instruction*.

Abstract

The internal/external frame of reference (I/E) model posits that students' academic self-concept in one domain (e.g., math) is positively associated with achievement in the same domain, but negatively associated with achievement in a comparison domain (e.g., English). Whereas academic self-concept encompasses self-ratings of achievement, peer academic reputation (PAR) encompasses peer ratings of the same achievement. Accordingly, it was hypothesized that the I/E model also applies to PAR. This hypothesis was tested in a sample of 850 seventh grade students in 47 classes in math and English. Therein, PAR was measured in a round robin design where students rated each classmate's level of achievement. We tested I/E models using academic self-concept and PAR as outcomes. Results supported the I/E model for academic self-concept but not for PAR. There, the effects across domains were positive rather than negative. Thus, whereas academic self-concept was characterized by contrasts between domains, PAR was not.

Keywords: academic self-concept, peer academic reputation, I/E model, dimensional comparison theory, social relations model

4. Perspective Matters: The Internal/External Frame of Reference Model for Self- and Peer Ratings of Competence

4.1 Introduction

When students rate their achievement in a school subject (e.g., math), there are many possible comparison standards they can use. The *internal/external frame of reference model* (I/E model; Marsh, 1986) highlights the effect of dimensional comparisons between students' achievement in math and their achievement in a verbal subject (e.g., English). For instance, students that are better in math than in English will have a more positive rating of their math achievement, and likewise a more negative rating of their English achievement. There is ample empirical support for the I/E model when it comes to self-ratings in the form of *academic self-concept* (Möller & Marsh, 2013; Möller, Müller-Kalthoff, Helm, Nagy, & Marsh, 2016; Möller, Pohlmann, Köller, & Marsh, 2009).

Whereas the I/E model is empirically well-supported when it comes to academic self-concept, it is unclear whether the I/E model also applies to peer-rated achievement, measured as *peer academic reputation* (PAR; Gest, Domitrovich, & Welsh, 2005). Based on the view that the I/E model appears to be a pervasive feature of domain-specific constructs, Möller, Helm, Müller-Kalthoff, Nagy, and Marsh (2015) proposed the hypothesis that dimensional comparisons are a reference for the formation of domain-specific peer ratings as well. So far, evidence for this hypothesis is mixed because experimental studies support it, but observational studies do not (Möller et al., 2015). However, these observational studies asked only single perceivers to rate a target student's achievement (Marsh, Smith, & Barnes, 1984; Pohlmann, Möller, & Streblow, 2004). Because of idiosyncratic rating behaviors, this method may have obscured effects within the I/E model.

In the present study, we investigated whether the I/E model applies not only to self- but also peer-rated achievement. We tried to overcome difficulties of earlier studies by using a more

rigorous design to assess PAR. That is, we used a sample of 850 seventh grade students who not only rated themselves, but also all their classmates with regard to achievement in English and math. In the resulting round-robin design, it is possible to control for idiosyncratic rating behaviors, which should allow for a more reliable test of the I/E model for PAR.

4.1.1 The Internal/External Frame of Reference Model for Academic Self-concept

Academic self-concept is referred to as one's self-perception of academic competence in a specific domain or school subject (Marsh & Shavelson, 1985). Academic self-concept has proved to be a very powerful construct as it is associated with several important outcomes. For instance, academic self-concept is predictive of academic choices (Parker et al., 2012), academic effort (Trautwein, Lüdtke, Schnyder, & Niggli, 2006), and academic achievement (Marsh & Craven, 2006).

The *internal/external frame of reference model* (I/E model; see Figure 4.1) describes a central mechanism in the formation of academic self-concept (Möller et al., 2009). The I/E model was originally proposed with regard to verbal and math self-concept and describes two simultaneous processes (Marsh, 1986). In the I/E model, students' self-concept in English is positively associated with achievement in English. But controlling for this association, self-concept in English is negatively associated with achievement in math (with analogous results for math self-concept). This pattern of results was supported in a meta-analysis by Möller et al. (2009) including 69 studies with a combined sample of 125,308 participants. The associations in the I/E model result in a profile of perceived strengths and weaknesses with only weak correlations between self-concepts, although the correlations between corresponding achievement scores are moderate to high (Möller & Marsh, 2013). Recently, findings regarding the I/E model were used as a basis to put forward dimensional comparison theory (DCT; Möller & Marsh, 2013), which proposes that dimensional comparisons (i.e., the internal part of the I/E model) are a pervasive feature of domain-specific ratings and evaluations. Accordingly, the I/E model might also apply to PAR (Möller et al., 2015).

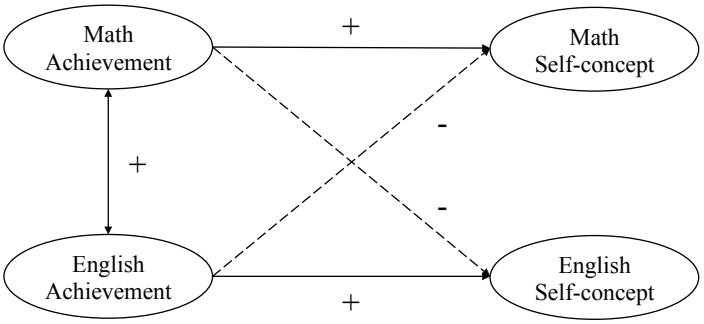

Figure 4.1. The internal/external frame of reference model (I/E model) according to Möller et al. (2009). Achievement in one domain has a positive effect on self-concept in the same domain (+), but a negative effect on self-concept in the comparison domain (-).

4.1.2 Peer Ratings of Achievement

Whereas academic self-concept describes a student's self-perception of their own achievement, *peer academic reputation* (PAR; Gest et al., 2005) describes how classmates perceive a target student's achievement. To measure PAR, multiple students in a classroom rate the achievement of a target student, and these ratings are then aggregated. Similar to general academic self-concept (Brunner et al., 2010), PAR has often been measured as a domain-general construct (i.e., not specific to English or math; Chen, Hughes, Liew, & Kwok, 2010; Gest et al., 2005; Gest, Rulison, Davidson, & Welsh, 2008; Hughes, Dyer, Luo, & Kwok, 2009). Nevertheless, for both academic self-concept and other self-perceptions of achievement, meta-analyses show that the correlation between self-perception and an achievement criterion is higher when both are measured at the same level of specificity (Möller et al., 2009; Valentine, DuBois, & Cooper, 2004): For instance, the correlation between domain-specific self-concept and domain-specific achievement is higher than the correlation between general self-concept and domain-specific achievement. Similarly to academic self-concept, domain-specific PAR may be more closely associated with domain-specific achievement. Moreover, studies by Marsh and colleagues (Marsh, 1986; Marsh et al., 1984) suggest that peer ratings of achievement might be Relatively few studies on PAR exist, yet they show a consistent pattern of PAR predicting multiple indicators of positive academic development. Gest et al. (2005, 2008) found that PAR was associated with academic achievement, effort, and general academic self-concept. Importantly, they found that PAR predicted school grades over and above earlier grades over the

course of several years (and similarly, that grades predicted PAR over and above earlier PAR). Similarly, Hughes et al. (2009) found that PAR predicted teacher-rated effort, teacher-rated achievement in math and reading, and performance on a reading test (but not on a math test) one year later, controlling for prior values of all measures including prior achievement. Chen et al. (2010) added to these results by finding that PAR predicted achievement in a standardized test three years later, controlling for prior achievement as well as IQ and economic status. Therein, the effect of PAR on achievement was mediated by general academic self-concept and engagement. Finally, Denissen et al. (2011) found in research on university students that PAR predicted and was predicted by grades on exams in a cross-lagged model. Moreover, PAR predicted student dropout, whereas self-rated ability did not. Overall, considering the evidence on the predictive validity of PAR so far, it might be important to understand how PAR is formed.

4.1.3 The I/E Model in Peer Perception

Möller et al. (2015, p. 430) hypothesized that "dimensional comparisons influence evaluations of and by other people and groups." This hypothesis is based on the idea that ratings (including achievement ratings) are always made with regard to some frame of reference (Kahneman & Miller, 1986). Two notions support the idea that dimensional comparisons within the I/E model might be a frame of reference for PAR. First, a major cause underlying the I/E model seems to be dissimilarity between domains, which should be equally valid in self- and peer perception. Second, the I/E model might describe a "routine procedure" (Möller & Marsh, 2013, p. 550) that happens in a wide range of domain-specific ratings and evaluations between dissimilar domains. In the following paragraphs, we provide a more detailed outline of these arguments.

First, it is proposed that the I/E model is caused by characteristics of academic domains. According to the multidimensional model of self-concept (Marsh & Shavelson, 1985), academic domains can be arranged on a continuum ranging from verbal to mathematical. Dimensional comparisons between domains that are further apart on the continuum (e.g., math and English) should have stronger effects than dimensional comparisons between near domains (e.g., math and physics). One experiment (Helm, Mueller-Kalthoff, Nagy, & Möller, 2016) and a number of observational studies supported this reasoning and found no effects or even positive effects for near comparisons (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2014; Marsh,

Lüdtke, et al., 2015; Möller, Streblow, Pohlmann, & Köller, 2006; Xu et al., 2013). Similarly, students who believed in a negative correlation between mathematical and verbal abilities made stronger dimensional comparisons (Möller, Streblow, & Pohlmann, 2006).

Second, evidence so far suggests that the I/E model might apply to domain-specific ratings (in dissimilar domains) in general (Möller et al., 2016). So far, the I/E model has been applied to interest (Schurtz, Pfost, Nagengast, & Artelt, 2014), academic emotions (Goetz, Frenzel, Hall, & Pekrun, 2008), intrinsic motivation (Marsh, Abduljabbar, et al., 2015), and perceptions of the learning environment (Arens & Möller, 2016). Additionally, intervention effects targeting motivation in one domain might have negative effects on motivation in a comparison domain (Gaspard et al., 2016). In a study on group perception, Yzerbyt, Kervyn, and Judd (2008) found results similar to the I/E model: groups that were perceived as high in "warmth" were perceived as low in "competence" and vice versa (Möller et al., 2015).

In line with this reasoning, two experimental studies support the I/E model with regard to peers' and teachers' perceptions of student achievement. Dickhäuser (2005) studied whether teachers' ratings of students' achievement were affected by dimensional comparisons in an experiment using vignettes. If a vignette described a student as low-achieving in math, teachers rated that student's verbal ability as higher than when a vignette described the student as high-achieving in math. This result indicates that teachers used dimensional comparisons when rating the student. Also using vignettes, results by Möller (2005) indicate that students also use dimensional comparisons to rate the ability of (hypothetical) peers.

However, evidence for the I/E model in terms of the perception of others is not conclusive (Möller et al., 2015): Observational studies in school contexts do not support the I/E model with regard to the perception of others (Marsh, 1986). Dai (2002) found no support for the cross-domain effect of the I/E model in parents' ratings of their children's competence. Marsh et al. (1984) tested the I/E model for inferred self-concepts. In their study, they let single peers infer the self-concept of one target student each, and found no negative effects across domains. Finally, in a German study, Pohlmann et al. (2004) asked single peers to infer a target student's self-concept, and also let teachers infer the self-concepts of all their students. The I/E model was not supported for either teachers' or peers' inferred self-concept.

However, it is unclear to what extent the results of these observational studies regarding inferred self-concepts generalize to PAR. Whereas PAR describes how peers perceive the

achievement of a target student, inferred self-concepts describe how peers perceive the self-concept of a target student. Moreover, whereas PAR was measured on the basis of multiple classmates' ratings of a target student (Gest et al., 2005), the studies mentioned above measured inferred self-concepts by asking single students to rate one target student (Marsh et al., 1984; Pohlmann et al., 2004). If only single perceivers rate a target, these ratings may be confounded with idiosyncratic tendencies, leading to inflated correlations between ratings (Nestler & Back, 2015). Studies on interpersonal perception employing multiple perceivers per target show that about 20% of variance in interpersonal ratings is accounted for by idiosyncratic rating tendencies (termed *relationship effect*; Hoyt & Kerns, 1999; Kenny, 1994). Having multiple perceivers assess each target controls for these idiosyncratic tendencies. For instance, studies that used vignettes (and supported the I/E model with regard to ratings by others) were based on multiple perceivers per target (Dickhäuser, 2005; Möller, 2005). Overall, by investigating PAR instead of inferred self-concepts and by taking ratings from more than one perceiver per target, student results for the I/E model might be different compared to earlier studies.

4.1.4 The Present Investigation

In this study, we compared the I/E model for self-concept in math and English with the I/E model for PAR in math and English. For self-concepts, we expected to find positive effects within domains and negative effects across domains, in line with the I/E model. Following the hypothesis by Möller et al. (2015), we also expected the same for PAR.

We conducted the study by reanalyzing data from a large-scale school achievement study with math and English as a foreign language as domains. The I/E model should apply to these domains in a German sample, because the domains are located on either extreme of the continuum of verbal and mathematical domains (Marsh, Lüdtke, et al., 2015; Möller, Streblow, Pohlmann, et al., 2006). In earlier observational studies, inferred self-concepts were investigated rather than PAR, and moreover, only single perceivers were asked to rate each target student. To avoid these issues, we focused our investigation on PAR, which was assessed in a round-robin design where students rated every classmate (see Figure 4.2 for an example). Furthermore, we used a large sample of 850 school students and achievement measures consisting of school grades and standardized achievement tests.

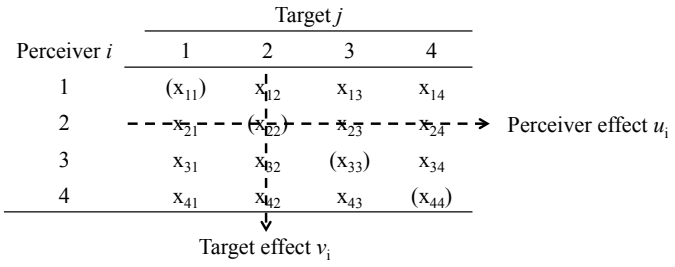

Figure 4.2. Illustration of an univariate round robin design with four participants (Alex, Beth, Charlie, and Daisy). The achievement rating of student *i* about student *j* is represented by x_{ij}. The self-concept of student *i* is represented by (x_{ii}). The arrows indicate effects that are estimated in the social relations model (Kenny, 1994): the perceiver or row effects u_i (i.e., how student *i* tends to rate the achievement of others) and the target or column effects v_j (i.e., how others rate the achievement of student *i*). The target effect can be used as a measure of PAR (Denissen et al., 2011).

4.2 Method

4.2.1 Sample

This study uses data from the German TRAIN study, which is a longitudinal school study focusing on achievement, motivation, and well-being among students in non-academic school tracks (Jonkmann, Rose, & Trautwein, 2013a). We used a subsample of the study consisting of classes in which round-robin data were collected. This subsample comprised students from the non-academic school track in Saxony (*Mittelschule*). We used data from the third measurement wave in 2010, when students were in the seventh grade. Earlier studies have shown that older (and also better acquainted) students' peer ratings of competence are more differentiated and more accurate than the peer ratings of younger (and less closely acquainted) students (Malloy, Sugarman, Montvilo, & Ben-Zeev, 1995; Malloy, Yarlas, Montvilo, & Sugarman, 1996). Thus, we chose the measurement wave in which students have known each other for somewhat more than two years to ensure reliable and valid measurement of PAR.

Altogether, our target sample comprised 1,083 students in 48 classes. We had to exclude 233 students who had no round robin data (23 were absent at the time of measurement and 210

did not have parental consent to participate in the round robin measurement). On average, 85.39 % (*SD* = 14.76 %) of students in each classroom participated in the round-robin design. Furthermore, we excluded one class in which only five students participated in the round robin measurement. This led to a final sample of 850 students in 47 classes (46.52 % female, mean age 13.08, *SD* = 0.51). Measurement was conducted on two consecutive days. Trained research assistants visited intact classrooms and administered all measures during regular school hours.

4.2.2 Measures

Academic self-concept. Academic self-concept in math and English was measured with a German version of the SDQ including items like "I am good in math/English" (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005). The scale consisted of four items per domain and had a reliability of α = .85 in math and α = .80 in English.

Interpersonal achievement ratings. Peer ratings of math and English achievement were assessed in a round-robin design (see Figure 4.2 for an example). Each student had to rate all classmates with the items "This student is good at math" and "This student is good at English", which could be answered on a 7-point Likert scale ranging from *not at all* to *very much so*. Altogether, we analyzed data from 7,662 dyads (where at least one student rated one other student). Based on this round-robin data, we estimated PAR in math and English as students' target effect (see the statistical procedure below).

Achievement. Achievement was measured by grades and performance on a standardized achievement test. Grades were obtained from the school administration. We used the previous year's final grades in math and English, that is, at the end of sixth grade. For the analyses, we recoded all grades (which were reported in line with the German grade metric), so that high scores represented good grades.

Second, achievement was measured with standardized achievement tests in math and English. The math test consisted of curricularly valid items (Rose et al., 2013). The test was administered according to a rotated test booklet design (Mislevy, Beaton, Kaplan, & Sheehan, 1992) that allowed for scaling across measurement waves (Rose et al., 2013). We used WLE estimates of students' test performance. Their marginal reliability (which is comparable to Cronbachs α) was .70. The English test was a listening comprehension test with curricularly valid items (Jonkmann, Rose, & Trautwein, 2013b). The test was the same for all students and

consisted of 25 items with a reliability of $\alpha = .71$. We used the sum score of the test as a measure of English achievement. Both tests were administered during the two days of measurement.

4.2.3 Statistical Procedure

Our analyses consisted of two steps. First, we estimated PAR in math and English. Second, we analyzed I/E models for academic self-concept and PAR.

We estimated PAR using a bivariate *social relations model* (SRM; Kenny, Kashy, & Cook, 2006). The SRM is a methodology to analyze and describe interpersonal perceptions in various domains. The univariate SRM is well known in research on interpersonal perception. However, to investigate the I/E model, we needed to model competence ratings in two domains simultaneously. Thus, we drew on the bivariate extension of the SRM (Kenny, 1994). The prototypical design for investigating the SRM is a round-robin design (Kenny, 1994). Variance in ratings is attributed to individual differences stemming from three sources. The *perceiver effect* describes a tendency to rate peers as competent (e.g., How competent does Alex think his classmates are in math?). The *target effect* describes a tendency to be rated as competent (e.g., How competent do Alex' classmates think he is in math?). Finally, the *relationship effect* describes the unique tendency of one student to rate one other student, and cannot be separated from measurement error (e.g., How competent does Alex think that Beth in particular is in math?). The target effect can be used as a measure of students' PAR (Denissen et al., 2011; Nestler & Back, 2015).

We applied a Bayesian approach to estimate the bivariate SRM due to its flexibility (Lüdtke, Robitzsch, Kenny, & Trautwein, 2013). We used the software WinBUGS accessed via R and the package R2winBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; R Core Team, 2015; Sturtz, Ligges, & Gelman, 2005). Following Lüdtke et al. (2013), we chose non-informative priors for all parameters. Thus, parameter estimates were mainly based on observed data and can be interpreted similarly to maximum-likelihood estimates (Nestler, 2015). For parameter estimates of the SRM analysis, see Table S4.1 in the supplement.

To control for unreliability in the estimation of target effects (which were estimated in the SRM; see Bonito & Kenny, 2010), and their associations with the achievement measures and the academic self-concept in the following analyses, we extended the SRM approach by applying the method of plausible values from the missing data literature (Yang & Seltzer, 2015; see also

Enders, 2010). That is, we generated plausible values (see Asparouhov & Muthén, 2010) for the target effect. To do this, we specified a background model within the SRM estimation by including grades, achievement tests and academic self-concept in both domains as predictors for the target effect in both domains. We used this approach to correctly estimate the correlation between the target effect and the achievement measures and academic self-concept in the following analyses (see Yang & Seltzer, 2015). For further statistical analyses, we drew 50 plausible values from the posterior distribution of the target effect (see Bodner, 2008). We aggregated statistical estimates using Rubin's rule (Rubin, 1987) via the R package miceadds for the correlations (Robitzsch, 2015) and via procedures to handle multiple imputations implemented in Mplus 7.1 for the I/E models (Muthen & Muthen, 2012).

In the second step, we investigated several I/E models for academic self-concept and PAR as structural equation models. In the basic I/E model, math self-concept is regressed on math and English achievement and simultaneously English self-concept is regressed on English and math achievement (Möller et al., 2009). A negative regression coefficient across domains (e.g., English self-concept on math achievement) indicates dimensional comparison effects. In the basic model, we used grades and standardized tests as achievement measures. We adapted this basic specification and additionally used the target effect (as a measure for PAR) as outcome. Altogether, we specified four I/E models: two basic models with self-concept as the outcome and either grades or achievement test as the predictor, and two models with target effect as the outcome and either grades or achievement test as the predictor. We specified self-concept as measured by the self-concept items. We controlled for unreliability in achievement test scores by specifying them as measured with a single indicator and constrained their residual variance to 1 - their reliability (.30 in math and .29 in English). The I/E models were estimated using Mplus 7.1 (Muthen & Muthen, 2012). We adjusted standard errors to account for the clustered data structure with students clustered in classes. For all models, we report standardized coefficients.

4.3 Results

4.3.1 Zero Order Correlations

Manifest correlations of all investigated variables are presented in Table 4.1. Both achievement measures were correlated with the corresponding self-concept and PAR, for

instance, in math, with $r = .36$ for the correlation between grades and self-concept and $r = .74$ for the correlation between grades and PAR. Both students themselves and their peers were relatively accurate. Furthermore, self-concept and PAR were correlated as well (e.g., $r = .40$ in math). This result indicates that students had a moderate to high level of self-other agreement with their peers.

Table 4.1

Zero-order Correlations between Achievement, Self-concept, and PAR in Math and English

	Math				English			
	1.	2.	3.	4.	5.	6.	7.	8.
Math								
1. Test								
2. Grade	.41*							
3. Self-concept	.35*	.36*						
4. PAR	.48*	.74*	.40*					
English								
5. Test	.40*	.30*	.14*	.40*				
6. Grade	.28*	.57*	.10*	.63*	.42*			
7. Self-concept	.10*	.18*	.17*	.32*	.27*	.43*		
8. PAR	.40*	.69*	.27*	.94*	.44*	.71*	.40*	

Note: Test = Performance on the standardized achievement test; Grade = Final grade for the previous year, grades were recoded so that higher scores represent higher achievement; PAR = Peer academic reputation.

* $p < .05$

Across domains, the correlations between the same constructs measured in math and English showed an interesting pattern. Achievement measures across domains had a moderate to high correlation, for instance, grades in math and English were correlated with $r = .57$. As is usually found, the correlation between self-concepts in math and English was substantially lower with $r = .17$. These patterns point to dimensional comparison effects with contrasts for self-ratings. However, PAR across domains were highly correlated, with $r = .94$. This pattern indicated no dimensional contrasts for peer ratings.

4.3.2 I/E Models for Self-Ratings and Peer Ratings

In Figure 4.3, we present results for the I/E models. Overall, self-concept in one domain was positively associated with achievement in the same domain. The usual I/E model emerged with grades as predictors: A good grade in math was related to a higher self-concept in math (with a coefficient of 0.53, *SE* = 0.04, $p < .001$) but a lower self-concept in English, controlling for English grades (with a coefficient of -0.14, *SE* = 0.06, $p = .027$). Similarly, a good grade in English was related to a higher self-concept in English (with a coefficient of 0.54, *SE* = 0.06, $p < .001$) but a lower self-concept in math, controlling for grades in math (with a coefficient of -0.19, *SE* = 0.05, $p < .001$). However, there were no effects across domains with the achievement tests as predictors. The difference between achievement tests and grades might indicate that grades were more salient for the students than the achievement tests.

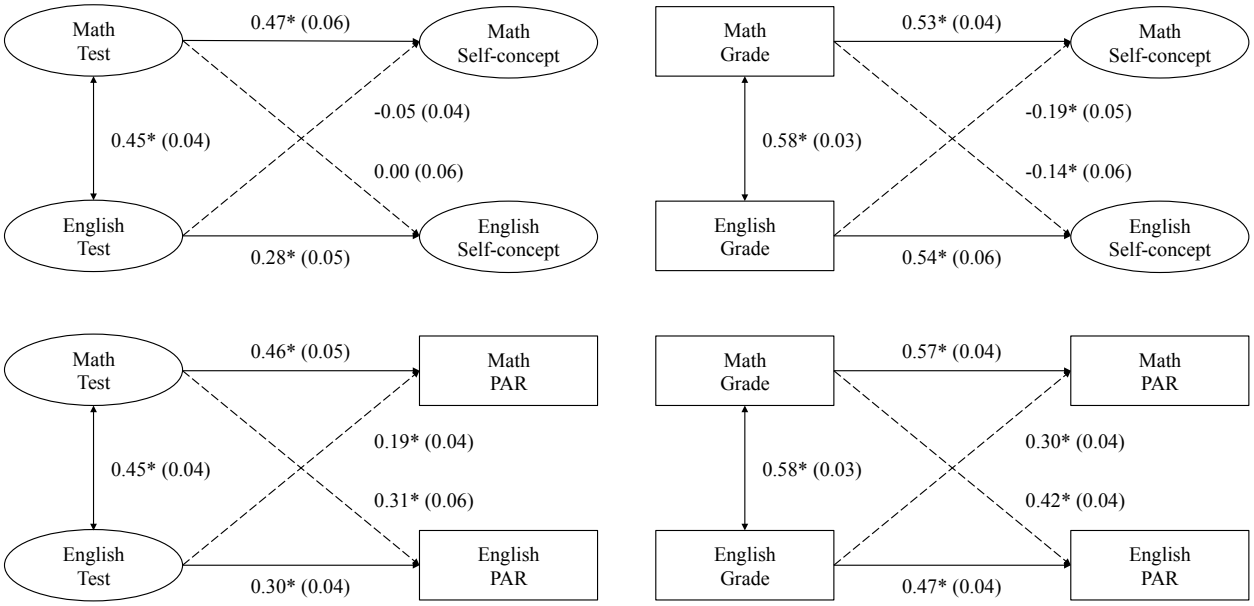

Figure 4.3. Results of the I/E models. The upper row shows I/E models with self-concept as outcome, the lower row shows PAR as outcome. On the left, standardized tests are used as predictors, and on the right, grades are used as predictors. Shown are standardized coefficients (and their standard errors in brackets). Coefficients of dashed lines refer to dimensional comparisons; * $p < .05$

A different picture emerged for PAR. As was the case for self-concept, achievement in one domain was positively associated with PAR in that domain for both achievement measures. For instance, the coefficient of math grades predicting PAR in math was 0.61, $SE = 0.05$, $p < .001$. Contrary to the results for self-concept, achievement in one domain was also positively associated to PAR in the other domain. For instance, the coefficient of math grades predicting PAR in English was 0.41, $SE = 0.04$, $p < .001$. Apparently, students with higher achievement in math were rated as more competent in both math and English. Similarly, students with higher achievement in English were rated as more competent in both English and math. Thus, the internal frame of reference did not emerge. Achievement in general was associated with a more positive rating by peers, which might indicate that peers perceived differences in achievement as smaller than they actually were. Comparing achievement tests and grades as predictor, the coefficients were higher for the grades.

4.4 Discussion

In this study, we investigated whether the I/E model applies to peer ratings of competence as well as to self-concept ratings. By reanalyzing data from a large-scale school achievement study, we investigated the I/E model for academic self-concept and PAR in math and English. To overcome issues associated with earlier studies, we measured PAR using a round-robin design in which students rated each of their classmates. Our results supported the I/E model for self-concept. That is, grades in one domain (e.g., English) had a positive effect on self-concept in the corresponding domain (e.g., English), but a negative effect on self-concept in the non-corresponding domain (e.g., math). However, the I/E model was not supported for PAR as outcome. Grades in one domain had a positive effect on PAR in the corresponding domain but also a positive effect on PAR in the non-corresponding domain.

4.4.1 Theoretical Implications

The results of our study imply that PAR is based not on a dimensional, but rather on a social frame of reference (Marsh, 1986). That is, the effect within domains indicates that peers could accurately sort the rank order of their classmates' achievement in both domains. Yet the

absence of effects across domains in the I/E model indicates that peers did not use a dimensional frame of reference.

The absence of negative cross-domain effects in the I/E models for PAR might have implications for the supposed causes of dimensional comparisons. One proposed cause in the I/E model relates to students' beliefs about the negative interrelatedness of mathematical and verbal ability (Möller, Streblow, & Pohlmann, 2006). This belief was assessed with items like "there are only a few people who are equally talented in math and German" (Möller et al., 2006, p. 61), which explicitly addresses the negative interrelatedness of domains in other individuals. Another proposed cause for the I/E model was said to be the relative distance between domains on the continuum from verbal to mathematical domains (Helm et al., 2016; Jansen et al., 2015; Marsh, Lüdtke, et al., 2015). Both potential causes refer to characteristics of domains. In turn, domain characteristics seem to be independent of the target of a competence rating. Thus, they should be equally applicable to the I/E model for peer ratings. However, as we did not find evidence for the I/E model in peer ratings, this raises the question of what is different between these two perspectives, that is, what are the moderating mechanisms that allow the outlined causes to work in self- but not in peer perceptions of competence.

Targeting this question, our results might imply that peer perception is less differentiated than self-perception, that is, PAR seems to have a different and less multidimensional organization compared to academic self-concept. Earlier studies on the I/E model in peer perception often found no cross-domain effects (Marsh et al., 1984; Pohlmann et al., 2004). We even found positive cross-domain effects. This effect is similar to effects observed for near comparisons on the continuum from verbal to mathematical domains (Jansen et al., 2015; Marsh et al., 2014; Marsh, Lüdtke, et al., 2015; Möller, Streblow, Pohlmann, et al., 2006; Xu et al., 2013). Thus, it might be that the distance between verbal and mathematical domains is smaller with regard to peer perception than with regard to self-perception. PAR might be rather unidimensional and measure general perceived competence in school. If that were to be the case, the I/E model might apply to different dimensions for peer perception than for self-perception. That is, the I/E model might not apply to academic dimensions like math and English, but to more general dimensions like general academic competence and, for instance, trustworthiness (Möller & Savyon, 2003).

But there are also indicators that PAR might be seen as domain-specific. Marsh et al. (1984) found that inferred self-concepts were similarly domain-specific to self-concepts. In our study, the correlation for PAR across domains was markedly higher than the correlation for achievement across domains. At the same time, however, PAR was highly correlated with achievement within domains – at least regarding grades. Thus, although the dimensional structure of PAR does not seem to match the dimensional structure of achievement exactly, PAR was highly correlated with domain-specific achievement. Overall, more studies of PAR that collect peer ratings in more than just two domains are necessary to draw conclusions about the dimensionality of peer perception.

Finally, our study included a much larger number of perceivers per target student than earlier studies investigating inferred self-concepts (Marsh et al., 1984; Pohlmann et al., 2004). Regarding the interpretation of results, we were able to confirm earlier findings for the I/E model with regard to peer perception. We were also able to support this conclusion with a stronger methodology precluding the biases that may occur due to single perceivers. Moreover, going beyond earlier studies, we found positive effects across domains. Future studies should follow up on this result and investigate whether it can be replicated for different samples.

4.4.2 Limitations and Directions for Future Research

One limitation of our study concerns the measurement of PAR. Peer ratings of achievement were collected in a grid structure with classmates ordered in rows and competence ratings in columns. Furthermore, the item format is confounded with the outcome: Academic self-concept was measured in a traditional format, whereas PAR was measured within the grid. This item format may have evoked or diminished frame of reference effects. However, it is unclear whether this grid format has an effect and in which direction this effect might work. Future studies should systematically observe the effects of such item formats to exclude it as a potential alternative explanation. Nevertheless, we observed similar effects as Pohlmann et al. (2004) and Marsh et al. (1984), who assessed inferred self-concepts in a traditional questionnaire design. That is, the effect of the grid format is not likely to be very strong.

Moreover, our sample may have characteristics that limit the generalizability of results. In the federal state of Saxony (where the data was collected), there are two school tracks, and our sample comprised students from the lower track. These students were lower-achieving and might

have also been potentially less motivated. This may affect the generalizability of our results and may have led to lower reliability and validity of our measures, especially considering that providing peer ratings for achievement might have been arduous. That being said, results from the SRM models (see table S4.1 in the supplement) indicated a sufficient amount of variation in ratings to allow for a useful interpretation of PAR (see Kenny, 1994). Nevertheless, with regard to PAR, there is a need for more studies in diverse samples to shed more light on the generalizability of results.

4.4.3 Conclusion

In this study, we sought to understand whether and how the I/E model would apply to peer perception as compared to self-perception as one specific test of dimensional comparison theory (Möller et al., 2015; Möller & Marsh, 2013). We found negative cross-domain effects for self-ratings but positive cross-domain effects for peer ratings in I/E models with math and English. Because of these results, it seems that self-perception of competence is dominated by contrasts between verbal and mathematical domains, but peer perception of competence is not.

References

Arens, A. K., & Möller, J. (2016). Dimensional comparisons in students' perceptions of the learning environment. *Learning and Instruction*, *42*, 22–30. http://doi.org/10.1016/j.learninstruc.2015.11.001

Asparouhov, T., & Muthén, B. (2010). *Plausible values for latent variables using Mplus*. *Mplus Technical Report*. Retrieved from http://www.statmodel.com/download/Plausible.pdf

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(4), 651–675. http://doi.org/10.1080/10705510802339072

Bonito, J. A., & Kenny, D. A. (2010). The measurement of reliability of social relations components from round-robin designs. *Personal Relationships*, *17*(2), 235–251. http://doi.org/10.1111/j.1475-6811.2010.01274.x

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, *102*(4), 964–981. http://doi.org/10.1037/a0019644

Chen, Q., Hughes, J. N., Liew, J., & Kwok, O.-M. (2010). Joint contributions of peer acceptance and peer academic reputation to achievement in academically at-risk children: Mediating processes. *Journal of Applied Developmental Psychology*, *31*(6), 448–459. http://doi.org/10.1016/j.appdev.2010.09.001

Dai, D. Y. (2002). Incorporatingparent Perceptions: A Replication and Extension Study of the Internal-External Frame of Reference Model of Self-Concept Development. *Journal Of Adolescent Research*, *17*(6), 617–645. http://doi.org/10.1177/074355802237467

Denissen, J. J. A., Schönbrodt, F. D., van Zalk, M., Meeus, W. H. J., & van Aken, M. A. G. (2011). Antecedents and consequences of peer-rated intelligence. *European Journal of Personality*, *25*(2), 108–119. http://doi.org/10.1002/per.799

Dickhäuser, O. (2005). Teachers' inferences about students' self-concepts – the role of dimensional comparison. *Learning and Instruction*, *15*(3), 225–235. http://doi.org/10.1016/j.learninstruc.2005.04.004

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.

Gaspard, H., Dicke, A.-L., Flunger, B., Ha fner, I., Brisson, B. M., Trautwein, U., & Nagengast, B. (2016). Side Effects of Motivational Interventions? Effects of an Intervention in Math Classrooms on Motivation in Verbal Domains. *AERA Open*, *2*(2). http://doi.org/10.1177/2332858416649168

Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology*, *97*(3), 337–346. http://doi.org/10.1037/0022-0663.97.3.337

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology*, *44*(3), 625–636. http://doi.org/10.1037/0012-1649.44.3.625

Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33*(1), 9–33. http://doi.org/10.1016/j.cedpsych.2006.12.002

Helm, F., Mueller-Kalthoff, H., Nagy, N., & Möller, J. (2016). Dimensional Comparison Theory: Perceived Subject Similarity Impacts on Students Self-Concepts. *AERA Open*, *2*(2), 1–9. http://doi.org/10.1177/2332858416650624

Hoyt, W. T., & Kerns, M.-D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, *4*(4), 403–424. http://doi.org/10.1037//1082-989X.4.4.403

Hughes, J. N., Dyer, N., Luo, W., & Kwok, O.-M. (2009). Effects of peer academic reputation on achievement in academically at-risk elementary students. *Journal of Applied Developmental Psychology*, *30*(2), 182–194. http://doi.org/10.1016/j.appdev.2008.12.008

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and Assimilation Effects of Dimensional Comparisons in Five Subjects: An Extension of the I/E Model. *Journal of Educational Psychology*, Advance online publication. http://doi.org/10.1037/edu0000021

Jonkmann, K., Rose, N., & Trautwein, U. (2013a). Hörverstehen in Englisch: Bundesland- und Schulformunterschiede und eine quasi-experimentelle Untersuchung der Nachhaltigkeit des Grundschulenglischunterrichts. In K. Jonkmann, N. Rose, & U. Trautwein (Eds.), *Tradition*

und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen.

Jonkmann, K., Rose, N., & Trautwein, U. (2013b). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen*. Tübingen: Projektbericht an die Kultusministerien der Länder.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153. http://doi.org/10.1037//0033-295X.93.2.136

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. http://doi.org/10.1037/a0029252

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337. http://doi.org/10.1023/A:1008929526011

Malloy, T. E., Sugarman, D. B., Montvilo, R. K., & Ben-Zeev, T. (1995). Children's interpersonal perceptions: A social relations analysis of perceiver and target effects. *Journal of Personality and Social Psychology*, *68*(3), 418–426. http://doi.org/10.1037/0022-3514.68.3.418

Malloy, T. E., Yarlas, A., Montvilo, R. K., & Sugarman, D. B. (1996). Agreement and accuracy in children's interpersonal perception: A social relations analysis. *Journal of Personality and Social Psychology*, *71*(4), 692–702. http://doi.org/10.1037/0022-3514.71.4.692

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. http://doi.org/10.3102/00028312023001129

Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., Nagengast, B., … Abu-Hilal, M. M. (2015). The Internal/External Frame of Reference Model of Self-

Concept and Achievement Relations: Age-Cohort and Cross-Cultural Differences. *American Educational Research Journal*, *52*(1), 168–202. http://doi.org/10.3102/0002831214549453

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. http://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology*, *39*(4), 326–341. http://doi.org/10.1016/j.cedpsych.2014.08.003

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

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. http://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. *Australian Journal of Psychology*, *36*(3), 367–386. http://doi.org/10.1080/00049538408255318

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161. http://doi.org/10.1111/j.1745-3984.1992.tb00371.x

Möller, J. (2005). Paradoxical effects of praise and criticism: Social, dimensional and temporal comparisons. *British Journal of Educational Psychology*, *75*(2), 275–295. http://doi.org/10.1348/000709904X24744

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional Comparisons and Their Consequences for Self-Concept, Motivation, and Emotion. In *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 6, pp. 430–436). Elsevier. http://doi.org/10.1016/B978-0-08-097086-8.26092-3

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

120(3), 544–560. http://doi.org/10.1037/a0032459

Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The generalized internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research*. http://doi.org/10.14786/flr.v4i2.169

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. http://doi.org/10.3102/0034654309337522

Möller, J., & Savyon, K. (2003). Not Very Smart, thus Moral: Dimensional Comparisons Between Academic Self-Concept and Honesty. *Social Psychology of Education*, *6*(2), 95–106. http://doi.org/10.1023/A:1023247910033

Möller, J., Streblow, L., & Pohlmann, B. (2006). The belief in a negative interdependence of math and verbal abilities as determinant of academic self-concepts. *British Journal of Educational Psychology*, *76*(1), 57–70. http://doi.org/10.1348/000709905X37451

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education*, *21*(4), 467–487. http://doi.org/10.1007/BF03173515

Muthen, L., & Muthen, B. (2012). *Mplus User's Guide*. *Los Angeles, CA: Muthén & Muthén* (Seventh Ed). Los ANgeles, CA: Muthén & Muthén.

Nestler, S. (2015). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*. http://doi.org/10.1007/s11336-015-9474-9

Nestler, S., & Back, M. D. (2015). Using Cross-Classified Structural Equation Models to Examine the Accuracy of Personality Judgments. *Psychometrika*, 1–23. http://doi.org/10.1007/s11336-015-9485-6

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: a multicontext study. *Developmental Psychology*, *48*(6), 1629–42. http://doi.org/10.1037/a0029167

Pohlmann, B., Möller, J., & Streblow, L. (2004). Zur Fremdeinschätzung von Schülerselbstkonzepten durch Lehrer und Mitschüler. *Zeitschrift Für Pädagogische Psychologie*, *18*(3/4), 157–169. http://doi.org/10.1024/1010-0652.18.34.157

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Robitzsch, A. (2015). miceadds: Some additional multiple imputation functions, especially for "mice." Retrieved from http://cran.r-project.org/package=miceadds

Rose, N., Jonkmann, K., Hübner, N., Sälzer, C., Lüdtke, O., & Nagy, G. (2013). Durchführung und methodische Grundlagen der TRAIN-Studie. In K. Jonkmann, N. Rose, & U. Trautwein (Eds.), *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen*.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/9780470316696

Schurtz, I. M., Pfost, M., Nagengast, B., & Artelt, C. (2014). Impact of social and dimensional comparisons on student's mathematical and English subject-interest at the beginning of secondary school. *Learning and Instruction*, *34*, 32–41. http://doi.org/10.1016/j.learninstruc.2014.08.001

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. http://doi.org/10.1026/0012-1924.51.4.183

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, *12*(21), 1–16. Retrieved from http://link.aps.org/doi/10.1103/PhysRevLett.44.1404

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, *98*(2), 438–456. http://doi.org/10.1037/0022-0663.98.2.438

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. http://doi.org/10.1207/s15326985ep3902_3

Xu, M. K., Marsh, H. W., Hau, K.-T., Ho, I. T., Morin, A. J. S., & Abduljabbar, A. S. (2013). The internal/external frame of reference of academic self-concept: Extension to a foreign language and the role of language of instruction. *Journal of Educational Psychology*, *105*(2), 489–503. http://doi.org/10.1037/a0031333

Yang, J. S., & Seltzer, M. (2015). Handling measurement error in predictors using a multilevel latent variable plausible values approach. In J. R. Harring, L. M. Stapelton, & S. N. Beretvas (Eds.), *Advances in Multilevel Modeling for Educational Research: Addressing Practical Issues Found in Real-World Applications* (pp. 295–333). Information Age Publishing.

Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation Versus Halo: The Unique Relations Between the Fundamental Dimensions of Social Judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1110–1123. http://doi.org/10.1177/0146167208318602

Supplement: SRM Results

Table S4.1

Estimates of the SRM Parameters of the Bivariate Model for Math and English

	Est.	95% interval
Variance proportion math		
Perceiver variance proportion (assimilation)	.20	[.18, .22]
Target variance proportion (consensus)	.33	[.30, .35]
Relationship variance Proportion (uniqueness)	.47	[.45, .49]
Variance proportion English		
Perceiver variance proportion (assimilation)	.24	[.22, .27]
Target variance proportion (consensus)	.28	[.26, .30]
Relationship variance Proportion (uniqueness)	.48	[.46, .50]
Dyadic correlations		
Dyadic reciprocity math	.04	[.01, .06]
Dyadic reciprocity English	.06	[.03, .08]
Dyadic correlation math-English within	.57	[.56, .58]
Dyadic correlation math-English between	.05	[.04, .08]
Individual correlations		
Generalized reciprocity math	-.02	[-.10, .06]
Generalized reciprocity English	-.06	[-.14, .01]
Generalized perceiver correlation math-English	.90	[.88, .91]
Generalized target correlation math-English	.94	[.92, .95]
Generalized between reciprocity math-English	-.03	[-.10, .06]
Generalized between reciprocity English-math	-.05	[-.13, .03]

Note: SRM = Social relations model; Est. = point estimate; variance proportion = variance proportion of the respective component in relation to the total variance of interpersonal ratings; Dyadic correlations are correlations on the dyad level; individual correlations are correlations on the individual level; Generalized reciprocity math-English (English-math) = correlation of the perceiver effect in math (English) to the target effect in English (math).

5

General Discussion

5. General Discussion

5.1 Summary of Results

In three studies, this thesis investigated the triangular connections between self-appraisals, peer appraisals, and achievement. Study 1 addressed the link between self-appraisals and achievement. A sample of academic track students in Grade 13 gave postdictions about their final exams in math and English. On average, students were well calibrated. Individual differences in postdiction bias (i.e., overestimations in postdictions) was predicted by low achievement in the same domain, a high level of a school's average achievement, male gender, and high narcissism. Overall, postdiction bias appeared to be a domain-general characteristic. Study 2 connected all three constructs of the triangle: self-appraisals, peer appraisals, and achievement. Using a sample of non-academic track students in Grade 5 and Grade 8, the link between math self-concept and math achievement was studied in an interpersonal approach. Within the interpersonal approach, students' self-concept was disentangled into an appraisal shared with classmates and an appraisal that was only held be students themselves (self-enhancement). Self-enhancement showed only a small positive association with achievement. Conversely, there was a substantial overlap in variance between self-appraisals and peer appraisals. Moreover this overlap was highly correlated to current achievement and predictive of future achievement. These results indicated that especially a shared positive achievement appraisal might lead to better achievement. Finally, Study 3 targeted the connection between peer appraisals and achievement in a multidimensional perspective. We compared the internal/external frame of reference (I/E) model for academic self-concept with the I/E model for peers' appraisal of achievement (i.e., PAR) in a sample of non-academic track students in the domains of math and English in Grade 7. The typical I/E model emerged for the academic self-concept, but not for PAR. Conversely, for PAR, positive and not negative cross-domain effects emerged. Thus, it seemed that academic self-concept was characterized by contrasting between domains, but peer appraisals were not. Main characteristics of the three studies are summarized in Table 5.1.

Table 5.1

Summary of Studies

	Study 1: Postdiction bias	Study 2: Interpersonal approach	Study 3: Perspective matters
School grade	Grade 13	Grade 5 and Grade 8	Grade 7
School track	Academic track	Non-academic track	Non-academic track
Federal state	Hamburg	Saxony	Saxony
Domains	Math English	Math	Math English
Achievement measures	Final exam (Abitur), standardized test	Grade at the end of the year, standardized test; both in two consecutive years	Grade at the end of the year, standardized test
Classification of self-appraisal	Criterion-based Past-oriented Task-specific	Absolute Past-oriented Domain-specific	Absolute Past-oriented Domain-specific
Calibration estimates (correlation coefficients)	Postdiction (in math/English) with… …exam: .84/.71 … test: .68/.48	Self-concept (in Grade 5/8) with… …test: .23/.25 …grade: .28/.45	Self-concept (in math/English) with… …test: .35/.27 …grade: .36/.43
	Self-concept (in math/English) with… …exam: .61/.55 …test: .60/.34	PAR (in Grade 5/8) with… …test: .52/.56 …grade: .56/.64	PAR (in math/English) with… …test: .48/.44 …grade: .74/.71

Following, I present theoretical implications of this thesis in reference to the major chapters of the theoretical framework. Next, I will outline practical implications of the three studies, discuss strengths and limitations of the thesis, and conclude by referring back to the student introduced in the introduction of the present thesis.

5.2 Theoretical Implications

5.2.1 Self-appraisals: Insights from the Classification Model

In the introduction, I presented a classification model of self-appraisals (Jansen & Lösch, in prep.). This model categorizes results found for different self-appraisals as a function of self-appraisals' *frame of reference*, *specificity*, and *time orientation*. The studies of this thesis gave new insights, which can inform further development of this model.

Results of Study 1 can be interpreted in light of the classification model, because postdictions are brought in a new research context. So far, postdictions were often used in lab-like settings (Baars, Vink, van Gog, de Bruin, & Paas, 2014) or without considering personality or reference effects as correlates (Hacker, Bol, Horgan, & Rakow, 2000). In this study, postdictions were assessed in a setting typical for self-concept research allowing for comparisons between both constructs (Marsh, Trautwein, Lüdtke, & Köller, 2008). Conceptually, both constructs differ: Postdictions can be classified as criterion-referenced, task-specific, and past-oriented. Contrary, academic self-concept can be classified as absolute-referenced, domain-specific, and past-oriented (Bong & Skaalvik, 2003). In line with these conceptual differences, empirical results differed as well: Contrary to what is found for academic self-concept (Möller & Marsh, 2013), we did not find the *internal/external frame of reference* (I/E) model for postdiction bias. An effect similar to the *big-fish-little-pond* (BFLPE) effect represented by a school's average level of achievement had the opposite direction as for academic self-concept (Marsh, Trautwein, et al., 2008). Finally, postdictions were accompanied by higher calibration estimates. Overall, this pattern of results might be indicative of effects elicited by the *frame of reference* and *specificity* dimensions proposed in the classification model.

Similarly, results of Study 2 may be informed by the classification model. In calibration research, it is generally supposed that calibration leads to better achievement (Bol & Hacker, 2012; de Bruin & van Gog, 2012; Thiede, Anderson, & Therriault, 2003). Put otherwise, especially overestimations of one's achievement are seen as detrimental for future achievement (Dunlosky & Rawson, 2012). Study 2 tested this claim for academic self-concept over the course of one year. Yet, no detrimental effect of an overly positive self-concept on achievement could be found. That is, the dimensions of the classification model might have moderated the effect of

calibration on achievement. Of course, this interpretation should be handled with care, as not only self-appraisals but also many other study characteristics differed, for example, calibration was often studied in a much smaller and more controlled setting. That is, more thorough investigations targeted at the classification model are necessary to clarify the association between calibration and achievement and the moderating role of the classification dimensions.

Finally, Studies 2 and 3 might be seen as adding another dimension to the classification model, which might be *perspective*. Appraisals from the perspective of the self differed from appraisals from the perspective of others. Adding this dimension could inform research on collaborative learning and the collaborative regulation of study behaviors (Grau & Whitebread, 2012; Iiskala, Vauras, Lehtinen, & Salonen, 2011; Kollar & Fischer, 2010; Volet, Vauras, & Salonen, 2009). For instance, there could be differences between peer appraisals as we used them (absolute-referenced, domain-specific, past-oriented) and peer appraisals that ask to estimate how well peers will answer questions about a given text (criterion-referenced, task-specific, future-oriented).

5.2.2 Self-appraisals and Achievement: Calibration in Schools

All three studies of this thesis targeted calibration to some extent Thus, the studies may inform about characteristics of calibration, that is, the association between students' (self-) appraisals and actual achievement.

Overall, calibration estimates indicated at least a moderate level of accuracy. Almost all effect sizes of calibration are in excess of the meta-analytic effect size of $r = .29$ identified by Zell and Krizan (2014), indicating that students in schools seem on average to be better calibrated than samples of previous meta-analyses. The statement that students are biased and not aware of their abilities (Dunning, Heath, & Suls, 2004) might not be appropriate. It might be more useful to start with the notion that "the mind typically reflects, rather than produces social reality" (Jussim, Harber, Crawford, Cain, & Cohen, 2005, p. 85), meaning that appraisals should be supposed to be rather accurate.

Nevertheless, in Study 1, we determined conditions for this accuracy. We investigated postdiction bias, which was the difference between postdictions of exam grades and the actual exam grades. This difference was predicted by cognitive abilities, personality traits, and contextual factors. Earlier studies interpreted such deviance scores exclusively as a lack in ability

to correctly appraise one's achievement (Kruger & Dunning, 1999) or exclusively as motivational tendency to present oneself as overly positive (Dufner et al., 2012). Yet, according to our results, both explanations seem to be valid. That is, difference scores between a self-appraisal and an achievement criterion may indicate both: a lack of ability for accurate self-appraisals and the motivation to self-enhance. To be able to interpret these difference scores in a straightforward way, it would be helpful to quantify the relative importance of cognition (the ability to be calibrated) and motivation (especially self-enhancement) in self-appraisals. In order to quantify these contributions, valid measures for ability and motivation to be calibrated need to be employed. However, depending on the cause of lacking calibration, different interventions are necessary. That is, lacking ability should be treated differently than lacking motivation to be accurate.

Relatedly, the ability to calibrate should not only affect self-appraisals, but also peer appraisals. That is, the ability to calibrate should result in a similar accuracy for the self and for others. Yet, we found that students were more accurate for others than for the self, which might be seen as indication that students are potentially able to calibrate accurately but have evaluative tendencies to increase their self-appraisals (Alicke & Sedikides, 2009).

In Studies 1 and 2, we studied the association between calibration and achievement. In both studies, the association between achievement and postdiction bias or self-enhancement had a small effect size. That is, the importance of having calibrated self-appraisals for achieving academic success claimed in self-regulated learning (Alexander, 2013; Bol & Hacker, 2012; de Bruin & van Gog, 2012) was less evident in our studies. Yet, calibration is especially said to influence choices, for instance, regarding what material should be learned (Metcalfe, 2009). Self-enhancement in academic self-concept measured in Study 2 might not affect small-scale regulation choices that would increase achievement. Conversely, as self-concept influenced choices regarding educational trajectories (Parker et al., 2012), these choices might be affected by self-enhancement (for instance, which vocational career to take). That is, self-enhancement might lead to suboptimal choices regarding educational trajectories and accordingly affect long-term economic outcomes. For instance, choosing a study subject that relies a lot on math (e.g., engineering) although one is rather low achieving but self-enhancing in math, could lead to less satisfaction with the study or dropout of university.

Finally, the result of Study 2 that self-enhancement had no further effect on achievement might be interesting to note with regard to academic self-concept theory. Our study may suggest that especially a substantively positive self-concept predicts future achievement. When peers shared a positive appraisal, there was a positive effect on achievement. When peers did not share a positive appraisal, the effect was much smaller. This pattern may shed a different light on reference effects affecting academic self-concept. Often, self-concept is seen as "an important mediating construct that facilitates the attainment of other desirable psychological and behavioral outcomes" (Marsh & Hau, 2003). In case of the BFLPE, the class-average achievement is suggested to affect achievement mediated by academic self-concept (e.g., Marsh, Kuyper, Morin, Parker, & Seaton, 2014). Yet it is somewhat puzzling, that class-average achievement was found to be positively associated with individual achievement (Marsh et al., 2010) – a result that recently was also found longitudinally (Stäbler, Dumont, Becker, & Baumert, 2016). That is, there was a negative effect of class-average achievement on academic self-concept, but a positive effect of class-average achievement on students' achievement. This effect might be interpreted in accordance with Study 2 and the TRI model: the BFLPE might only affect self-enhancement, that is, the idiosyncratic, not shared self-perception. In turn, self-enhancement had only a small effect on future achievement. In contrast, the shared part of academic self-concept (represented by the target effect) might not be affected by the BFLPE. Thus, no negative effect on achievement might follow the BFLPE. Overall, the mediating effect of academic self-concept appears to be less clear in light of recent studies and our results.

Considering this thesis with regard to the presented dichotomy of the motivational and cognitive approach, this thesis rather favors ideas of the motivational approach. Limiting self-appraisals to cognitive mechanisms seems to paint a picture too narrow to fit to schools. Self-appraisals in all three studies were positively associated with achievement and only small or no negative effects of overestimations could be shown. Nevertheless, this thesis uses designs that are typical for the motivational approach. For a comprehensive picture, it is necessary to test motivational accounts in designs of the cognitive approach as well.

5.2.3 Self-appraisals and Peer Appraisals: Similarities and Differences

Self-appraisals and peer appraisals of achievement (e.g., PAR) might be seen as aspects of the same latent construct, as they are conceptually similar and have similar correlates and

outcomes. For example, both have reciprocal associations with achievement (Gest, Rulison, Davidson, & Welsh, 2008; Marsh & Martin, 2011). The underlying latent construct could be labeled *perceived* achievement as presented within the framework of the *trait-reputation-identity* (TRI) model (Kenny & West, 2010; McAbee & Connelly, 2016). This thesis presented some similarities and differences between self-appraisals and peer appraisals of achievement corroborating the validity of the TRI model for appraisals. In the introduction, we proposed that validity of the TRI model would be supported by correlations between self-appraisals and peer appraisals, similar correlates, and similar formation mechanisms of both constructs.

Overall, there was a shared variance regarding achievement and appraisals of achievement. That is, correlations were at least moderate in size when considering self-other agreement, other-other agreement (i.e., consensus), the correlation between self-appraisals and achievement, and the correlation between peer appraisals and achievement. Furthermore, self-appraisals and peer appraisals predicted future achievement. Regarding all criteria proposed by Funder (1995), a level of accuracy was achieved that is at least equal to levels of accuracy achieved in other personality traits (Back & Kenny, 2010). Furthermore, these aligning appraisals indicate a consensual rank-order. Self-other agreement and other-other agreement could only be achieved because students ranked their peers and themselves in a similar order in the round-robin designs (cf. Kwan, John, Robins, & Kuang, 2008). Thus, these correlations indicate that self-appraisals and peer appraisals are formed with a similar social reference (Marsh, Seaton, et al., 2008).

Yet, there were also systematic differences between self-appraisals and peer appraisals with regard to their correlates and formation mechanisms. In Studies 2 and 3, peer appraisals correlated higher with achievement than self-appraisals correlated with achievement. On the one hand, this might be attributable to information aggregated in peer appraisals. Measures of peer appraisals (e.g., PAR) were based on a much higher amount of information than would be possible for self-appraisals. On the other hand, peer appraisals may be based on different criteria than self-appraisals. That is, self-enhancement might underlie the formation of self-appraisals with no similar strong equivalent for peer appraisals (Vazire, 2010). Thus, self-enhancement might have reduced correlations between self-appraisals and achievement.

Another difference between peer and self-appraisals might be their dimensionality as identified in Study 3. The positive cross-domain effects in the I/E models in Study 3 contradicted

dimensional comparison effects in PAR. Similarly, PAR was highly correlated across domains. Overall, these results might indicate that the dimensional structure of peer appraisals might be different from the dimensional structure of self-appraisals. Möller and colleagues (Möller, Helm, Müller-Kalthoff, Nagy, & Marsh, 2015; Möller, Müller-Kalthoff, Helm, Nagy, & Marsh, 2016; Möller & Marsh, 2013) noted that the motive to know one's strengths and weaknesses might be a cause of the I/E model. Since this motive is absent for peer appraisals, high correlations of PAR between domains might have resulted. In inferred self-concepts, Marsh and colleagues found evidence that the dimensional structure was invariant across perspectives (Marsh, 1986; Marsh, Smith, & Barnes, 1984). Yet it might be fruitful to reexamine this finding with regard to appraisals of achievement.

Altogether, the adapted TRI model (McAbee & Connelly, 2016) as presented in the introduction might be a useful representation of appraisals of achievement in an interpersonal context. It displays an overlap in variance between self-appraisals and peer appraisals. Furthermore, this shared part seems to correlate higher with current and future achievement than the self-appraisal alone. In our studies, differences appeared between self-appraisals and peer appraisals but they may be explained with self-motives of self-enhancement and knowing one's strengths and weaknesses (Leary, 2007).

5.3 Practical Implications

The topic of calibration in actual school contexts is relevant for teachers' behavior: Should they rather foster optimistic or realistic self-appraisals? The best answer that can be given so far is that it depends on the self-appraisal. Regarding specific self-appraisals like JOLs it seems best to foster realistic self-appraisals as shown in earlier research (Bol & Hacker, 2012). Yet, this claim does not seem to hold for academic self-concept. There is massive evidence that academic self-concept and similar self-appraisals are beneficial for a positive development (Guo, Marsh, Parker, Morin, & Yeung, 2015; Marsh & Craven, 2006; Trautwein et al., 2012; Usher & Pajares, 2008). Study 2 is in line with other studies that found either no effects of an overly positive self-concept (Praetorius, Kastens, Hartig, & Lipowsky, 2016) or even found positive effects of overly positive self-appraisals (Bouffard & Narciss, 2011). Evidence for a negative effect of self-concept calibration stems largely from cross-sectional data (Chiu & Klassen, 2009, 2010). Overall, evidence for a negative effect of overly positive self-concept is rather weak.

Thus, the best advice for practitioners so far would be to foster a positive level with regard to general self-appraisals, but foster a realistic level of self-appraisals with regard to specific self-appraisals. However, no study so far investigated this composition of different self-appraisals explicitly. Especially experimental evidence should inform this advice.

A very different perspective on this result may be that teachers should systematically address and reduce the accuracy of students' appraisals of achievement. For instance, assuming that PAR has an effect on academic development controlling for prior achievement, this also implies that children who are seen as less achieving will also achieve less in the future (Gest et al., 2008). This connection might be worrisome, as it might prevent some students from developing ideally. That is, teachers might be encouraged to prevent a negative PAR for students. Alternatively, teachers could especially foster self-appraisals of students with negative PAR, because self-appraisals might function as mediators between PAR and achievement. Increasing self-appraisals could thus work as countermeasure.

5.4 Strengths and Limitations

A major strength of this thesis is the integration of similar constructs originating from different areas of psychology, which are often studied in isolation. Capitalizing on the classification model (Jansen & Lösch, in prep.), the interpersonal approach (Kwan, John, Kenny, Bond, & Robins, 2004) as well as the TRI model approach (McAbee & Connelly, 2016), several forms of appraisals can be studied in a joint framework. This approach allows comparing appraisals studied in several branches of educational, social, and personality psychology and thus may foster research with a broader scope. Ideally, this approach may contribute to the "important task" that was proposed by Wigfield and Eccles (2000, p. 72) to "examine more closely how similar and different these various measures are".

Moreover, Studies 2 and 3 in this thesis were among the first to use the novel Bayesian approach to the SRM (Lüdtke, Robitzsch, Kenny, & Trautwein, 2013). The SRM is a standard method of personality psychology (Back & Kenny, 2010; Kenny, West, Malloy, & Albright, 2006; Nestler, Grimm, & Schönbrodt, 2015) that is associated with the drawback of being relatively difficult to employ (Schönbrodt, Back, & Schmukle, 2012). Up until now, only few studies in educational psychology have picked up this approach. Yet, our studies may be an example that using the SRM might lead to interesting conclusions. As (methodological) research

in SRM is still ongoing and developing (Nestler, 2015; Nestler, Geukes, Hutteman, & Back, 2016; Nestler et al., 2015) its employment in educational psychology might be a fruitful endeavor.

Finally, strengths that all three studies share are the samples and the measures used. The samples are very large and designed to be representative of certain populations. Additionally, there are very few studies with a similar sample size using either postdictions or round-robin measures of peer appraisals. Finally, the achievement measures were carefully designed (Jonkmann, Rose, & Trautwein, 2013; Trautwein, Köller, Lehmann, & Lüdtke, 2007). Therefore, we were able to use highly valid and reliable achievement measures accompanied by externally valid school grades.

On the other hand, the samples pose some limitations to conclusions. Samples came from two federal states in Germany. Accordingly, implications might be limited to the respective regions. Moreover, the studies targeted only certain age groups. Additionally, our samples were restricted in terms of school tracks. We had older students from the academic track and younger students from the non-academic track but no mixtures of these properties. Overall, these sample characteristics might limit the generalizability of results and call for investigation of these patterns using more diverse samples in terms of federal state, age, and school track.

Another concern regards the measurement of appraisals. Postdictions in Study 1 and peer appraisals in Studies 2 and 3 were measured with one item each. Although self-appraisals and peer appraisals might be measured sufficiently with one item (Denissen, Geenen, Selfhout, & van Aken, 2008; Gogol et al., 2014), a more extensive measurement would be useful to ensure reliability. Especially regarding Study 1, it is difficult to assess postdictions with more than one item, yet it could be worthwhile to develop creative approaches that allow latent modeling of constructs.

Furthermore, we only investigated achievement and appraisals of achievement in two domains, that is, English and math. This choice of domains is reasonable against the background of the continuum of domains assumed in the Marsh/Shavelson model of academic self-concept (Marsh & Shavelson, 1985). Here, domains are assumed to lie on a continuum from verbal to mathematical domains where math and English may represent endpoints of that continuum (Jansen, Schroeders, Lüdtke, & Marsh, 2015; Marsh et al., 2015; Marsh, Kuyper, Seaton, et al., 2014). But some results indicated that domains can be differentiated on multiple dimensions and

that differences on these dimensions affect correlations between appraisals (Goetz et al., 2014; Gogol, Brunner, Preckel, Goetz, & Martin, 2016). Recently, in a study on the I/E model, the school subjects of history and politics did not quite fit in the unidimensional model of domains (Arens, Möller, & Watermann, 2016). Thus, results could be moderated by the studied domain.

A difficulty that arises in analyzing discrepancy scores (like self-enhancement and postdiction bias) is their interpretation (Edwards, 2001). Neither the criterion achievement nor the appraisal of achievement were measured without error. For instance, a student' self-appraisal could be "right" and the achievement might be "wrong". Similarly, the positive effect of academic self-concept on future achievement in the reciprocal effects model (Marsh & Martin, 2011) could result because previous achievement was not measured validly. That is, the effect of self-concept on achievement could be the result of imperfect achievement measures. However, as described above, the achievement measures were developed to be highly valid and reliable. Thus this concern might not weigh to heavy.

Finally, all three studies are observational and do not allow claims about causality and observed effects could work in another direction, bidirectional, or could be the result of a third variable. Furthermore, Studies 1 and 3 are cross-sectional. An experimental or intervention design would be useful to establish validity regarding effects of calibration or regarding the causation underlying the link between peers' appraisals and a student's achievement.

5.5 Directions for Future Research

5.5.1 Building on the Interpersonal Approach of Academic Self-concept

A major contribution of this thesis is the interpersonal approach to academic self-concept that outlines how academic self-concept might be intertwined with peer appraisals. Study 3 was a first application of this general idea as it investigated how components of the interpersonal approach were affected within the I/E model. As a further extension, more well known effects in research on academic self-concept could be studied from an interpersonal perspective.

One approach would be to study gender differences. Gender differences in academic self-concept are well known (Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Marsh et al., 2012; Watt, 2004) and they are said to affect gender differences in career aspirations (Parker et al., 2012; Watt et al., 2012). In light of the importance of these differences, it would be interesting to

note in which part of the interpersonal approach gender differences appear, that is, whether they only emerge for the self or for others as well. Moreover, Lüdtke et al. (2013) introduced an approach to investigate moderators within the SRM. This possibility could be used to investigate gender as moderator, that is, to examine, whether there are gender differences regarding gender of target, perceiver, or both.

Another approach would be to target longitudinal development of interpersonal appraisals. For academic self-concept, there is evidence for a general decline over school years (Nagy et al., 2010). Again, it would be interesting to see whether this decline is also apparent for peer appraisals. Recently, Nestler et al. (2016) proposed a longitudinal growth-curve approach to the SRM. This approach could be utilized to tackle interpersonal developments in appraisals and their association with achievement.

Finally, this thesis adds evidence that PAR is a predictor of students' academic development (Chen, Hughes, Liew, & Kwok, 2010; Denissen, Schönbrodt, van Zalk, Meeus, & van Aken, 2011; Gest, Domitrovich, & Welsh, 2005; Gest et al., 2008; Hughes, Dyer, Luo, & Kwok, 2009). It could be interesting to investigate the formation of PAR, that is, examine which information and references peers use in the formation of PAR. Furthermore, PAR could be related to popularity, yet the direction is unclear and could work in either direction (Leising, Erbs, & Fritz, 2010; Rentzsch, Schröder-Abé, & Schütz, 2012; Rentzsch, Schütz, & Schröder-Abé, 2011). Moreover, the academic setting should be incorporated into research on PAR, that is, whether teacher behavior or other class characteristics influence or moderate effects regarding PAR. For instance, teacher behavior that focuses on social comparisons might affect the correlation between objective achievement and PAR. Additionally, it would be interesting to further identify mediators of how PAR affects achievement. Chen et al. (2010) identified potential mediators like effort, yet their results need to be replicated. Finally, it would be worthwhile to identify whether PAR might be causal for a positive development of achievement. Previous studies argued that PAR is rather the cause than a marker of a positive development (Gest et al., 2008; Hughes et al., 2009). Yet for the similar case of teacher expectations, the expectancy effect is said to be rather weak (Jussim & Harber, 2005). In addition to the observational studies, it could thus be an approach to investigate an effect of PAR experimentally.

5.5.2 Targeting Calibration

Another branch of implications for future studies targets students' calibration. So far, it appears that students should be rather optimistic regarding their self-concept and rather realistic regarding their more specific self-appraisals. One approach would be to test this implication experimentally. Students could be given feedback that varies in positivity on two levels of specificity that is, that is on a specific level of postdictions and a more general level of academic self-concept. There were studies that manipulated feedback regarding specific self-appraisals which showed that realistic feedback was superior to overly positive or overly negative feedback (Försterling & Morgenstern, 2002; Kim, Chiu, & Zou, 2010). Yet, these studies would profit from additionally considering the motivational benefit of having a positive academic self-concept. Overall, such a study would combine accounts of motivationally focused and cognitively focused notions about self-appraisals.

Furthermore, in personality and social psychology, self-enhancement is often studied with regard to its effect on intrapersonal and interpersonal adaptation (Dufner et al., 2012; Dufner, Reitz, & Zander, 2014). This might also be especially interesting in the domain of schools where measures of achievement are available allowing for a valid operationalization of self-enhancement. An effect of self-enhancement might be tested on self-esteem or social self-concept. One specific outcome could be others perception: Leary (2007) outlined that self-enhancement might serve as a function to increase others' appraisal of oneself. This could be studied using round-robin data within schools, where longitudinal data might be especially interesting (Rentzsch & Schroder-Abe, 2014). This notion might be similar to testing whether self-enhancement might lead to an inflated appraisal of others about one's achievement (Lamba & Nityananda, 2014).

5.6 Conclusion

In the beginning, I introduced the sorcerer's apprentice. The apprentice clearly was not well calibrated and overestimated his magical abilities. As a consequence of his faulty self-appraisal, he had to face dire consequences. Therefore, this poem seems to illustrate the positions obtained by calibration researchers (Alexander, 2013; Bol & Hacker, 2012; de Bruin & van Gog, 2012; Dunlosky & Lipko, 2007; Metcalfe, 2009) and some personality and social psychologists

(Colvin & Block, 1994; Dunning et al., 2004; Wilson, 2009) who portray the consequences of overestimations as failure. However, as many studies in the tradition of calibration research, the poem had a quite short timeframe. It is unknown what happened afterwards.How did the apprentice continue learning and did he successfully finish his sorcerer training? At least, the poem indicates that the apprentice is highly motivated and eager to do magic, which might in the end be beneficial for learning (Marsh & Craven, 2006; Wigfield & Eccles, 2000). By living through the grave mistake the apprentice made, he might have learned much more than if he had not tried the magic in the first place. Based on the results of this thesis, one could suppose that the apprentice is not too bad in his calibration and his overestimations might not be too problematic for him in the end. Yet, to get to know more about his magical abilities, it would probably be a good idea to ask his fellow apprentices.

References

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction*, *24*, 1–3. http://doi.org/10.1016/j.learninstruc.2012.10.003

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*(1), 1–48. http://doi.org/10.1080/10463280802613866

Arens, A. K., Möller, J., & Watermann, R. (2016). Extending the internal/external frame of reference model to social studies: Self-concept and achievement in history and politics. *Learning and Individual Differences*, *51*, 91–99. http://doi.org/10.1016/j.lindif.2016.08.044

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, *33*, 92–107. http://doi.org/10.1016/j.learninstruc.2014.04.004

Back, M. D., & Kenny, D. A. (2010). The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass*, *4*(10), 855–870. http://doi.org/10.1111/j.1751-9004.2010.00303.x

Bol, L., & Hacker, D. J. (2012). Calibration research: where do we go from here? *Frontiers in Psychology*, *3*(July), 229. http://doi.org/10.3389/fpsyg.2012.00229

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

Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive biases in self-evaluation of academic competence: Introduction. *International Journal of Educational Research*, *50*(4), 205–208. http://doi.org/10.1016/j.ijer.2011.08.001

Chen, Q., Hughes, J. N., Liew, J., & Kwok, O.-M. (2010). Joint contributions of peer acceptance and peer academic reputation to achievement in academically at-risk children: Mediating processes. *Journal of Applied Developmental Psychology*, *31*(6), 448–459. http://doi.org/10.1016/j.appdev.2010.09.001

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. http://doi.org/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. http://doi.org/10.1016/j.learninstruc.2008.11.002

Colvin, C. R., & Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation. *Psychological Bulletin*, *116*(1), 3–20. http://doi.org/10.1037/0033-2909.116.1.3

de Bruin, A. B. H., & van Gog, T. (2012). Improving self-monitoring and self-regulation: From cognitive psychology to the classroom. *Learning and Instruction*, *22*(4), 245–252. http://doi.org/10.1016/j.learninstruc.2012.01.003

Denissen, J. J. A., Geenen, R., Selfhout, M., & van Aken, M. A. G. (2008). Single-item big five ratings in a social network design. *European Journal of Personality*, *22*(1), 37–54. http://doi.org/10.1002/per.662

Denissen, J. J. A., Schönbrodt, F. D., van Zalk, M., Meeus, W. H. J., & van Aken, M. A. G. (2011). Antecedents and consequences of peer-rated intelligence. *European Journal of Personality*, *25*(2), 108–119. http://doi.org/10.1002/per.799

Dufner, M., Denissen, J. J. A., 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*(3), 537–572. JOUR. http://doi.org/10.1111/j.1467-6494.2011.00742.x

Dufner, M., Reitz, A. K., & Zander, L. (2014). Antecedents, Consequences, and Mechanisms: On the Longitudinal Interplay between Academic Self-Enhancement and Psychological Adjustment. *Journal of Personality*, 1–41. http://doi.org/10.1111/jopy.12128

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*(4), 228–232. http://doi.org/10.1111/j.1467-8721.2007.00509.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. http://doi.org/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. http://doi.org/10.1111/j.1529-1006.2004.00018.x

Edwards, J. R. (2001). Ten Difference Score Myths. *Organizational Research Methods*, *4*(3), 265–287. http://doi.org/10.1177/109442810143005

Försterling, F., & Morgenstern, M. (2002). Accuracy of self-assessment and task performance: Does it pay to know the truth? *Journal of Educational Psychology*, *94*(3), 576–585. http://doi.org/10.1037//0022-0663.94.3.576

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*(4), 652–670. http://doi.org/10.1037/0033-295X.102.4.652

Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology*, *97*(3), 337–346. http://doi.org/10.1037/0022-0663.97.3.337

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology*, *44*(3), 625–636. http://doi.org/10.1037/0012-1649.44.3.625

Goetz, T., Haag, L., Lipnevich, A. A., Keller, M. M., Frenzel, A. C., & Collier, A. P. M. (2014). Between-domain relations of students' academic emotions and their judgments of school domain similarity. *Frontiers in Psychology*, *5*(October), 1–14. http://doi.org/10.3389/fpsyg.2014.01153

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., … Preckel, F. (2014). "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, *39*(3), 188–205. http://doi.org/10.1016/j.cedpsych.2014.04.002

Gogol, K., Brunner, M., Preckel, F., Goetz, T., & Martin, R. (2016). Developmental Dynamics of General and School-Subject-Specific Components of Academic Self-Concept, Academic Interest, and Academic Anxiety. *Frontiers in Psychology*, *7*(MAR). http://doi.org/10.3389/fpsyg.2016.00356

Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and*

Instruction, *22*(6), 401–412. http://doi.org/10.1016/j.learninstruc.2012.03.003

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J. S., & Yeung, A. S. (2015). 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://doi.org/10.1016/j.lindif.2015.01.008

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. a. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. http://doi.org/10.1037//0022-0663.92.1.160

Hughes, J. N., Dyer, N., Luo, W., & Kwok, O.-M. (2009). Effects of peer academic reputation on achievement in academically at-risk elementary students. *Journal of Applied Developmental Psychology*, *30*(2), 182–194. http://doi.org/10.1016/j.appdev.2008.12.008

Iiskala, T., Vauras, M., Lehtinen, E., & Salonen, P. (2011). Socially shared metacognition of dyads of pupils in collaborative mathematical problem-solving processes. *Learning and Instruction*, *21*(3), 379–393. http://doi.org/10.1016/j.learninstruc.2010.05.002

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, *73*(2), 509–527. http://doi.org/10.1111/1467-8624.00421

Jansen, M., & Lösch, T. (n.d.). Integrating perspectives on the self-appraisal of academic abilities: A classification model.

Jansen, M., Schroeders, U., Lüdtke, O., & Marsh, H. W. (2015). Contrast and Assimilation Effects of Dimensional Comparisons in Five Subjects: An Extension of the I/E Model. *Journal of Educational Psychology*, Advance online publication. http://doi.org/10.1037/edu0000021

Jonkmann, K., Rose, N., & Trautwein, U. (2013). *Tradition und Innovation: Entwicklungsverläufe an Haupt- und Realschulen in Baden-Württemberg und Mittelschulen in Sachsen - Abschlussbericht für die Länder Baden-Württemberg und Sachsen*. Tübingen: Projektbericht an die Kultusministerien der Länder.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, *9*(2), 131–155. http://doi.org/10.1207/s15327957pspr0902_3

Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies*, *6*(1), 85–102. http://doi.org/10.1075/is.6.1.07jus

Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self-and other perception: A meta-analysis. *Personality and Social Psychology Review*, *14*(2), 196–213. http://doi.org/10.1177/1088868309353414

Kenny, D. A., West, T. V, Malloy, T. E., & Albright, L. (2006). Componential Analysis of Interpersonal Perception Data. *Personality and Social Psychology Review*, *10*(4), 284–294. http://doi.org/10.1207/s15327957pspr1004_1

Kim, Y.-H., Chiu, C.-Y., & 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. http://doi.org/10.1037/a0020555

Kollar, I., & Fischer, F. (2010). Peer assessment as collaborative learning: A cognitive perspective. *Learning and Instruction*, *20*(4), 344–348. http://doi.org/10.1016/j.learninstruc.2009.08.005

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. http://doi.org/10.1037/0022-3514.77.6.1121

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*(1), 94–110. http://doi.org/10.1037/0033-295X.111.1.94

Kwan, V. S. Y., John, O. P., Robins, R. W., & Kuang, L. L. (2008). Conceptualizing and assessing self-enhancement bias: a componential approach. *Journal of Personality and Social Psychology*, *94*(6), 1062–77. http://doi.org/10.1037/0022-3514.94.6.1062

Lamba, S., & Nityananda, V. (2014). Self-Deceived Individuals Are Better at Deceiving Others. *PLoS ONE*, *9*(8), e104562. http://doi.org/10.1371/journal.pone.0104562

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology*, *58*(1), 317–344. http://doi.org/10.1146/annurev.psych.58.110405.085658

Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, *98*(4), 668–682. JOUR.

Retrieved from 10.1037/a0018771

Lüdtke, O., Robitzsch, A., Kenny, D. A., & Trautwein, U. (2013). A general and flexible approach to estimating the social relations model using Bayesian methods. *Psychological Methods*, *18*(1), 101–119. http://doi.org/10.1037/a0029252

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal*, *23*(1), 129–149. http://doi.org/10.3102/00028312023001129

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., … Parker, P. (2012). Factorial, Convergent, and Discriminant Validity of TIMSS Math and Science Motivation Measures: A Comparison of Arab and Anglo-Saxon Countries. *Journal of Educational Psychology*, *105*(1), 108–128. http://doi.org/10.1037/a0029907

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. http://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K.-T. (2003). Big-Fish--Little-Pond effect on academic self-concept: A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, *58*(5), 364–376. http://doi.org/10.1037/0003-066X.58.5.364

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. http://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Kuyper, H., Seaton, M., Parker, P. D., Morin, A. J. S., Möller, J., & Abduljabbar, A. S. (2014). Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation. *Contemporary Educational Psychology*, *39*(4), 326–341. http://doi.org/10.1016/j.cedpsych.2014.08.003

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

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement:

Relations and causal ordering. *British Journal of Educational Psychology*, *81*(1), 59–77. http://doi.org/10.1348/000709910X503501

Marsh, H. W., Seaton, M., Kuyper, H., Dumas, F., Huguet, P., Régner, I., … Gibbons, F. X. (2010). Phantom behavioral assimilation effects: Systematic biases in social comparison choice studies. *Journal of Personality*, *78*(April 2010), 671–710. http://doi.org/10.1111/j.1467-6494.2010.00630.x

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. http://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, *20*(3), 107–123. http://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Smith, I. D., & Barnes, J. (1984). Multidimensional self-concepts: Relationships with inferred self-concepts and academic achievement. *Australian Journal of Psychology*, *36*(3), 367–386. http://doi.org/10.1080/00049538408255318

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. http://doi.org/10.1037/0022-0663.100.3.510

McAbee, S. T., & Connelly, B. S. (2016). A Multi-Rater Framework for Studying Personality: The Trait-Reputation-Identity Model. *Psychological Review*. http://doi.org/10.1037/rev0000035

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. http://doi.org/10.1111/j.1467-8721.2009.01628.x

Möller, J., Helm, F., Müller-Kalthoff, H., Nagy, N., & Marsh, H. W. (2015). Dimensional Comparisons and Their Consequences for Self-Concept, Motivation, and Emotion. In *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 6, pp. 430–436). Elsevier. http://doi.org/10.1016/B978-0-08-097086-8.26092-3

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. http://doi.org/10.1037/a0032459

Möller, J., Müller-Kalthoff, H., Helm, F., Nagy, N., & Marsh, H. W. (2016). The generalized

internal/external frame of reference model: An extension to dimensional comparison theory. *Frontline Learning Research*. http://doi.org/10.14786/flr.v4i2.169

Nagy, G., Watt, H. M. G., Eccles, J. S., Trautwein, U., Lüdtke, O., & Baumert, J. (2010). The Development of Students' Mathematics Self-Concept in Relation to Gender: Different Countries, Different Trajectories? *Journal of Research on Adolescence*, *20*(2), 482–506. http://doi.org/10.1111/j.1532-7795.2010.00644.x

Nestler, S. (2015). Restricted maximum likelihood estimation for parameters of the social relations model. *Psychometrika*. http://doi.org/10.1007/s11336-015-9474-9

Nestler, S., Geukes, K., Hutteman, R., & Back, M. D. (2016). Tackling longitudinal round-robin data: A Social Relations Growth Model, (September).

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*(2), 272–295. http://doi.org/10.1002/per.1997

Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: a multicontext study. *Developmental Psychology*, *48*(6), 1629–42. http://doi.org/10.1037/a0029167

Praetorius, A.-K., Kastens, C., Hartig, J., & Lipowsky, F. (2016). Haben Schüler mit optimistischen Selbsteinschätzungen die Nase vorn? *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *48*(1), 14–26. http://doi.org/10.1026/0049-8637/a000140

Rentzsch, K., & Schroder-Abe, M. (2014). Self-Enhancement 2.0: An Integrated Approach to Measuring Dyadic Self-Enhancement at Two Levels. *Social Psychological and Personality Science*, *6*(3), 251–258. http://doi.org/10.1177/1948550614558634

Rentzsch, K., Schröder-Abé, M., & Schütz, A. (2012). Being Called a "Streber": The Roles of Personality and Competition in the Labelling of Academically Oriented Students. *European Journal of Personality*, *27*(5), n/a-n/a. http://doi.org/10.1002/per.1884

Rentzsch, K., Schütz, A., & Schröder-Abé, M. (2011). Being Labeled Nerd: Factors That Influence the Social Acceptance of High-Achieving Students. *The Journal of Experimental Education*, *79*(2), 143–168. http://doi.org/10.1080/00220970903292900

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: an R package for social relations analyses based on round-robin designs. *Behavior Research Methods*, *44*(2), 455–70. http://doi.org/10.3758/s13428-011-0150-4

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2016). What Happens to the Fish's Achievement in a Little Pond? A Simultaneous Analysis of Class-Average Achievement Effects on Achievement and Academic Self-Concept. *Journal of Educational Psychology*. http://doi.org/10.1037/edu0000135

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. http://doi.org/10.1037/0022-0663.95.1.66

Trautwein, U., Köller, O., Lehmann, R., & Lüdtke, O. (2007). *Schulleistungen von Abiturienten: Regionale, schulformbezogene und soziale Disparitäten*. Münster: Waxmann.

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*(3), 763–777. http://doi.org/10.1037/a0027470

Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, *78*(4), 751–796. http://doi.org/10.3102/0034654308321456

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. JOUR. Retrieved from 10.1037/a0017908

Volet, S., Vauras, M., & Salonen, P. (2009). Self- and Social Regulation in Learning Contexts: An Integrative Perspective. *Educational Psychologist*, *44*(4), 215–226. http://doi.org/10.1080/00461520903213584

Watt, H. M. G. (2004). Development of Adolescents' Self-Perceptions, Values, and Task Perceptions According to Gender and Domain in 7th- through 11th-Grade Australian Students. *Child Development*, *75*(5), 1556–1574. http://doi.org/10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., Shapka, J. D., Morris, Z. a, Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation,

educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*(6), 1594–1611. http://doi.org/10.1037/a0027838

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

Wilson, T. D. (2009). Know Thyself. *Perspectives on Psychological Science*, *4*(4), 384–389. http://doi.org/10.1111/j.1745-6924.2009.01143.x

Zell, E., & Krizan, Z. (2014). Do people have insight Into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. http://doi.org/10.1177/1745691613518075