University of Trier
Faculty I – Department of Psychology

Students' Achievement Motivation During the School Career: Development, Correlates, and Effects of Ability Grouping

Dissertation submitted to the University of Trier, Faculty I – Department of Psychology to obtain the degree of doctor rerum naturalium (Dr. rer. nat.) in Psychology.

Supervisors:
Prof. Dr. Franzis Preckel
University of Trier, Chair of Giftedness Research and Education

Prof. Dr. Michael Schneider
University of Trier, Chair of Educational Psychology

Submitted by:
M. Sc. Vsevolod Scherrer
on September 3, 2019

Dissertationsort: Trier

Danksagung

Ich möchte mich folgend bei den Menschen bedanken, die zum Erfolg dieser Dissertation beigetragen haben. Zunächst gilt mein besonderer Dank meiner Betreuerin Prof. Dr. Franzis Preckel und meinem Betreuer Prof. Dr. Michael Schneider für die sehr angenehme und produktive Zusammenarbeit.

Ebenso danke ich meiner Familie, Ma, Pa, Vasilij, Dima, Lena und den Großeltern, für die Unterstützung währen des Studium und der Promotion. Einen erheblichen Anteil am Erfolg der Promotion hatten auch meine jetzigen und ehemaligen Kollegen/ Kolleginnen: Isabelle, Lena, Moritz, Karin, Katharine, Henrike, Rachel, Svenja, Markus, Petra, Mireille, Andrea, Elena, Kaitlyn, Martina, Christine, Ella und Alex. Natürlich danke ich auch allen meinen Freunden, Markus, Danielle, Marie, Valentin, Jasmin, Nena, Dirk, Lisa und Laila, die ich hier nicht alle namentlich benennen kann. Danke oder aber auch spasibo!

Students' Achievement Motivation During the School Career: Development, Correlates, and Effects of Ability Grouping

Abstract

Theoretical and empirical research assumes a negative development of student achievement motivation over the course of their school careers (i.e., mean-level declines of achievement motivation). However, the exact magnitude of this motivational change remains elusive and it is unclear whether different motivational constructs show similar developmental trends. Furthermore, it is unknown whether motivational declines are related to a particular school stage (i.e., elementary, middle, or high school) or the school transition, and which additional changes are associated with motivational decreases (e.g., changes in student achievement). Finally, previous research has remained inconsistent regarding the question whether ability grouping of students helps prevent motivational declines or results in additional motivational "costs" for students.

This dissertation presents three articles that were designed to address these research questions. In Article 1, a meta-analysis based on 107 independent longitudinal studies investigated student mean-level changes in self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals from first to 13th grade. Article 2 comprised two longitudinal studies with German adolescents (Study: n = 745 students assessed in four waves in grades 5-7; Study 2: n = 1420 students assessed in four waves in grades 5-8). Both longitudinal studies investigated the separate and the joint development of achievement goals, interest, and achievement in math. In Article 3, a longitudinal study (n = 296 high-ability students assessed in four waves in grades 5-7) investigated the effects of full-time ability grouping on student development of academic self-concept and achievement in math.

The meta-analysis revealed significant decreases in math and language academic self-concept, intrinsic motivation, and mastery and performance-approach goals, whereas no significant changes in self-esteem, general academic self-concept, academic self-efficacy, and performance-avoidance goals were found. Interestingly, motivational declines were not related to school stage or school transition. In Article 2, decreases in interest and mastery, performance-approach, and performance-avoidance goals were indicated by both longitudinal studies. Development of mastery and performance-approach goals was positively related or unrelated to development in interest and achievement, whereas development of performance-avoidance goals was negatively related or unrelated to development of interest and achievement. Finally, the longitudinal study in Article 3 revealed no significant change in student academic self-concept in math over time. Ability grouping showed no positive or negative effects on student academic self-concept. However, high-ability students that were grouped together demonstrated greater gains in their achievement than high-ability students in regular classes.

Index of Publication

This Dissertation is divided into six chapters. In Chapter 1, the general topic is introduced and the central rationale of the present research is given. In Chapter 2, the theoretical background is presented. Chapter 3 contains an original research article, which is published in *Review of Educational Research.* Chapter 4 comprises an original research article that is currently under review in *Developmental Psychology*. Chapter 5 presents an original research article that is published in *Child Development*. Chapter 6 contains a general discussion of all presented articles. The author of this dissertation is the first author of Article 1 and 2 and co-author of Article 3. His contributions to the articles are presented below.

Chapter 3 Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research, 89,* 211-258. doi:10.3102/0034654318819127

Vsevolod Scherrer was the leading developer of the concept behind Article 1 (with support of Franzis Preckel). He collected the data, conducted the statistical analyses, and wrote the first draft of the article.

Chapter 4 Scherrer, V., & Preckel, F., Schmidt, I., & Elliot, A., J. (under review). Development of achievement goals and their relation to academicinterest and achievement in adolescence: A review of the literature and two longitudinal studies. *Developmental Psychology*.

Vsevolod Scherrer was the leading developer of the concept behind Article 2 (with support of Franzis Preckel and Andrew Elliot). He conducted the statistical analyses (with support of Isabelle Schmidt), and wrote the first draft of article 2.

Chapter 5 Preckel, F. , Schmidt, I. , Stumpf, E. , Motschenbacher, M. , Vogl, K. , Scherrer, V., & Schneider, W. (2019). High-ability grouping: Benefits for gifted students’ achievement development without costs in academic self-concept. *Child Development, 90,* 1185-1201. doi:10.1111/cdev.12996

Vsevolod Scherrer conducted and reported the propensity score matching analyses and participated in the general statistical analyses of Article 3.

Table of Content

Chapter 1. Introduction

Working as a lecturer at Trier University and holding courses on *intelligence*, I often noticed a deep frustration in some of my students when they realized that general intelligence is the best predictor of school grades. As it turned out, these students assumed that effort, persistence, and something that they called *motivation* - concepts these students considered malleable - determine school achievement and not “fixed” or “cold” concepts, such as intelligence. At first glance, their assumptions seemed to be seriously questioned by the empirical fact that intelligence is the best predictor of school grades. However, just as it is true that meta-analytic findings suggest a high correlation between intelligence scores and school grades (ρ= .54: Roth et al., 2015), these findings do not contradict the assumption that, to some extent, school achievement is also determined by effort, persistence, and motivation. In my courses, I noticed that this frustration quickly disappeared when students realized that even with a high correlation of ρ= .54, 69% of school grades variance remains unexplained by intelligence. That is, the portion of school grades variance that is not explained by intelligence is considerably larger than the portion that is. In fact, empirical research has shown that motivational constructs are able to explain school grades beyond intelligence (Steinmayr & Spinath, 2009). Finally, a smaller correlation between motivational factors and grades in comparison to intelligence and grades just means that, on average, differences in motivation are less predictive of grades than differences in intelligence. This does not rule out the possibility that motivation plays an even larger role than intelligence in particular cases. For example, it is possible that at the end of my course, a highly intelligent student purposely decides to submit a blank sheet instead of the course work to demonstrate that sometimes motivation can influence 100 percent of the (academic achievement) outcome.

Recognizing the power of motivation for school achievement and assuming that motivation is more variable than intelligence; the question arises, whether students’ motivation is treated responsibly. Regarding intelligence, it is well known that each additional school year benefits students’ IQ (Ceci, 1991; Ritchie & Tucker-Drob, 2017). However, the same cannot be said about the development of students’ achievement motivation. On the contrary, theoretical approaches and longitudinal research suggest a decline in students’ achievement motivation over the course of their school careers (Eccles, et al., 1993; Gottfried, Fleming, & Gottfried, 2011; Wigfield & Wagner, 2005). Thus, frustration about findings on student motivation should be related to the fact that students lose motivation over the course of their school careers and not to the fact that motivation is less predictive of achievement than intelligence.

The motivational decline is often explained by stage-environment fit theory (Eccles et al., 1993) according to which, a mismatch between adolescents' developing needs and the opportunities afforded by the school environment fosters a decline in student achievement motivation. However, precise statements about the magnitude of the decline and whether this change is more pronounced at certain grade levels (e.g., in adolescence after middle school transition) remain elusive, as no meta-analysis has examined motivational development during the school career to date. Furthermore, it remains unknown, whether different motivational constructs show the same or different developmental trends. This is because different predictions regarding different constructs can be derived from previous theoretical and empirical research. For example, regarding students' achievement goals, various predictions can be assumed in accordance with stage-environment fit theory. It has been assumed that the myriad of changes and challenges faced by adolescents during the adolescent period lead to a shift from mastery goal pursuit to performance goal pursuit (Anderman & Midgley, 1997; Eccles et al., 1993; Nicholls, 1989). However, one alternative is that the stage-environment fit mismatch results in a decreased commitment to competence per se, so that not only mastery goal pursuit, but also performance-approach and performance-avoidance goal pursuit decrease (Paulick, Waterman, & Nückles, 2013). Furthermore, it remains unclear which factors contribute to motivational decline and whether this decline is related to declines in additional constructs, such as school achievement.

Student grouping and lesson adjustment according to students' abilities seems to be a valid method to reduce the mismatch between students' needs and opportunities in accordance with stage-environment fit theory. However, in the literature, no consensus has been reached regarding the utility of ability grouping for the development of student achievement motivation and achievement (Dai, 2004; Mash et al., 2008; Seaton, Marsh, & Craven, 2009).

The remaining open research questions regarding student motivational development can be summarized as follows:

1. How does student achievement motivation develop during the school career?
2. What are the correlates of achievement motivation development?
    2a. Which variables predict achievement motivation development?
    2b. Which changes are associated with changes in achievement motivation?
3. How does ability grouping affect student motivational development?

This dissertation comprises three articles that address these research questions. The first article consists of a meta-analysis of longitudinal studies investigating student development on various motivational constructs. The second article presents two longitudinal studies investigating the (joint) development of achievement goals, interest, and achievement in German students. The third article consists a new longitudinal study investigating whether grouping high-ability students into special classes is related to their development of achievement motivation and achievement. Table 1 depicts which research questions were addressed by which article and displays which particular motivational constructs were investigated respectively.

Table 1. *Overview of Research Questions and Motivational Constructs Investigated by the Reported Articles*

Articles	Investigated motivational constructs	Research question 1: How does student achievement motivation develop during the school career?	Research question 2: What are the correlates of achievement motivation development?		Research question 3: How does ability grouping affect student motivational development?
			a. Which variables predict achievement motivation development?	b. Which changes are associated with changes in achievement motivation?	
Article 1	-Self-esteem -Academic self-concept -Self-efficacy -Intrinsic motivation -Mastery goals -PAP goals -PAV goals	✓	✓		
Article 2	-Interest -Mastery goals -PAP goals -PAV goals	✓	✓	✓	
Article 3	-Academic self-concept	✓		✓	✓

Note. PAP = Performance-approach. PAV = Performance-avoidance.

Chapter 2. Theoretical Background

2.1 Why do Students Learn? Theories of Achievement Motivation

Before investigating students' motivational development throughout their school careers, it is necessary to define what is meant by the term achievement motivation[1]. Students provide a wealth of different answers when asked why they learn for school (Buff, 2001; Urhahne, 2008). Students learn, because school grades are important, because they want to master the material, because of an extrinsic reward, because it is fun, because their parents want them to learn, because they want to be one of the best in their class, because they do not want to embarrass themselves in front of others, because they want to avoid a punishment, because they think that they are good in school, and because they expect to succeed. In short, students learn for school because they are motivated.

The term "motivation" is derived from the Latin word "movere" which means "to move". Thus, motivation is what gets us to move (Rudolph, 2009). A student without motivation does not move and does not learn. More precisely, motivation is a psychological process that initiates, organizes, sustains, controls, and evaluates purposeful actions of individuals (Grassinger, Dickhäuser, & Dresel, 2019). This process cannot be directly observed, but indirectly estimated by various indicators, such as previous behavior or self-reports. Although various reasons for learning can be summarized under the term (achievement) motivation (Murphy, & Alexander, 2000), this does not mean that the reasons have the same origin, nor that they work in the same way or even the same direction. In fact, various combinations of different reasons and interactions between these reasons are possible. For example, intrinsically oriented students learn because they enjoy the learning process, whereas extrinsically oriented students learn because they expect positive consequences. Interestingly, extrinsic rewards might undermine the free choice and intrinsic enjoyment of doing a task (Deci, 1971; Deci, Koestner, & Ryan, 1999)[2].

From antique times until today, researchers have developed various distinct (but sometimes overlapping) theories to describe, explain, structure, quantify, and predict motivational processes and their outcomes in humans (and animals). Long before modern

[1] Note that in Article 1 the term *competence motivation* is used instead of the term *achievement motivation* following the arguments of Elliot, Dweck, and Yeager (2017, p. 3) according to which the term *competence* offers a more precise definition of motivation than the term *achievement*.

[2] Some researchers doubt the universality of the undermining effect of extrinsic rewards on intrinsic motivation (e.g., Cameron & Perce, 1994; Eisenberger & Cameron, 1996; Lepper & Henderlong, 2000).

psychology, the Greek philosopher Epicurus developed the hedonistic school, which has been labeled as one of the first motivation theories (Rudolph, 2009). According to Epicurus' view, all actions can be explained by two expected conditions for oneself: the pleasure that one wants to maximize and the displeasure or pain that one wants to minimize. Most of the reasons why students learn can already be roughly categorized in terms of the hedonistic school. For example, enjoying the learning process, the desire to master the material, to be one of the best in the class, or to achieve an extrinsic reward, can all be classified into the pleasure condition. The desire to avoid either an embarrassment in front of others or a punishment can be classified into the displeasure condition. Furthermore, Epicurus already distinguished between physiological and mental pleasure conditions (Rudolph, 2009), which allow a further distinction between the reasons why students learn. An extrinsic reward is rather attributable to the physiological pleasure condition, whereas enjoying the learning process is truly a mental pleasure condition. Obviously, the motivational reasons why students learn show further distinctions that cannot be satisfactorily differentiated by the hedonistic school. For example, according to the hedonistic school, intrinsic interest in doing a task and the desire to master a task would be defined as a mental pleasure, respectively. However, these reasons can be further distinguished at both the theoretical and empirical level (Eccles & Wigfield, 2002; Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Korn & Elliot, 2016). In fact, several modern achievement motivation theories are required to sufficiently describe and allocate the reasons why students learn.

Modern achievement motivation theories focus on intentional and conscious motivational processes and suggest motivational constructs that represent the different reasons why students learn (Grassinger, Dickhäuser, & Dresel, 2019). For example, in achievement goal theory, the performance-avoidance goal is a construct that represents the desire to avoid embarrassment in front of others (Elliot & Hulleman, 2017). In the present research, several motivational constructs were investigated: self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, interest, and achievement goals (mastery, performance-approach, and performance-avoidance goals). Self-esteem reflects the overall attitude toward one's self (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995). Academic self-concept refers to a self-evaluation of one's ability in general or in a specific academic domain (Marsh & Craven, 1997; Marsh & Shavelson, 1985; Shavelson et al., 1976). Academic self-efficacy refers to beliefs about one's future performance on a specific task and the ability to regulate one's own learning activities and to master the task (Bandura, 1993; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Pajares, 1996). Note that self-efficacy and self-concept show great similarity, as both constructs refer to personal competence beliefs (Wigfield & Eccles, 2002). However, they can be distinguished at both the theoretical (Bong & Skaalvik, 2003) and the empirical level (Ferla, Valcke, & Cai, 2009; Scherer, 2013). Intrinsic motivation is a non-drive-based motivational propensity for behavior that does not require reinforcement for its maintenance (Deci & Ryan, 1985). Interest is characterized by an appreciation of the target (of interest), autonomy regarding interaction with the target, and positive emotions during the interaction (Schiefele, 1996). Intrinsic motivation and interest show strong conceptual similarities, as both constructs refer to enjoyment during engagement with a task and are often used interchangeably in the literature (e.g., Rawsthorne & Elliot, 1999; Baranik et al., 2010). Thus, differences between intrinsic motivation and interest are not further distinguished in the main articles of this dissertation[3]. Finally, achievement goals are the purpose for engaging in a competence-relevant behavior (Elliot & Hulleman, 2017). Table 1 depicts which exact constructs were investigated in which of the reported articles (see Chapter 1).

Although various theorists have made an effort to integrate motivational constructs from other theories into their own frameworks (e.g. Eccles & Wigfield, 2002), there is still no undisputed theory that is able to integrate all of the mentioned motivational constructs in a consistent way (Spinath, 2008). As the investigated constructs are central constructs within the structural model of self-concept, the achievement goals theory, the expectancy-value theory, and the self-determination theory, a summary of these theories is provided in the following. In addition, the construct self-esteem is introduced, as self-esteem was also investigated in the present research. Note that self-esteem is related to motivation, but is not a measure of motivation itself. Finally, the volition model of action phases (Heckhausen & Heckhausen, 2006) is introduced and the investigated constructs are allocated in accordance with Urhahne’s (2008) attempt to integrate constructs from several motivational theories within the volition model of action phases.

2.1.1. Structural Model of Self-Concept

In educational research, student academic self-concept is one of the most investigated predictor variables of achievement outcomes (e.g., Guay, Marsh, & Boivin, 2003; Huang, 2011). Individuals’ self-concept refers to the perceptions and beliefs about oneself in general (e.g., “I am a successful person”) or the perceptions of one’s ability in a specific area (e.g., “I am good at math”; Marsh, 1990; Shavelson, Hubner, & Stanton, 1976). The self-concept

[3] Some researchers have discussed a further differentiation between *intrinsic motivation* and *interest* on the theoretical (Renninger, 2000) and empirical level (Reeve, 1989). Note that *intrinsic motivation* and *interest* are subsumed under the term *intrinsic motivation* in Article 1 and under term *interest* in Article 2.

allows individuals to categorize a vast amount of information about themselves and to evaluate this information with regard to others individuals and other perceptions of oneself (Shavelson, Hubner, & Stanton, 1976). According to structural model of self-concept (e.g., Marsh, 1990), self-concept is multifaceted and hierarchical, with the general self-concept at the apex of the model and specific self-concepts, such as the academic self-concept, at a subordinate levels. In the literature, a positive academic self-concept is posited as a desirable outcome that is positively associated with student effort (e.g., Skaalvik & Rankin, 1995), cognitive strategy use and self-regulation (Pintrich & De Groot, 1990), mastery achievement goals (Bong, 2001; Roeser, Midgley, & Urdan, 1996), intrinsic motivation (Gottfried, 1990), and achievement (Huang, 2011).

According to the internal/external frame of reference model (Marsh, 1986), students' self-concepts are formed by internal and external comparison processes. Internal comparison processes (dimensional comparisons) refer to the comparison of one's ability on different domains (e.g., one's ability in math versus one's ability in the German language). Internal comparison processes result in self-concept decreases on the worst domain and self-concept increases on the best domain (Möller & Marsh, 2013). External comparison processes (social comparisons) refer to the comparison of one's ability on one domain with others' abilities on the same domain (e.g., one's ability in math versus the class average ability in math). Dependent on the ability of the social reference group of one domain, external comparison processes can increase or decrease one's own self-concept on the same domain.

2.1.2. Achievement Goal Theory.

Achievement goal theory is a further prominent approach to explain how students think, feel, and behave in school settings. According to this theory, achievement goals are competence-relevant purposes for engaging in behavior (Elliot & Hulleman, 2017). Depending on the underlying achievement goal model, a different number of achievement goals are postulated. In the initial model of achievement goals, a dichotomous differentiation between mastery and performance goals was proposed. Mastery goals focus on the development of task competence, whereas performance goals focus on the demonstration of competence relative to others (Dweck, 1986; Nicholls, 1984). In the trichotomous model, performance goals were additionally split into performance-approach and performance-avoidance goals. Performance-approach goals focus on trying to demonstrate a higher level of competence relative to others, while performance-avoidance goals focus on trying to avoid demonstrating incompetence relative to others (Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997). Furthermore, the 2 x 2 model split mastery goals into mastery-approach and mastery-avoidance goals in addition to the maintained differentiation of performance-approach and performance-avoidance goals. Mastery-approach goals focus on trying to approach intrapersonal or task-based competence, whereas mastery-avoidance goals focus on trying to avoid intrapersonal or task-based incompetence (Elliot, 1999; Pintrich, 2000a). Finally, based on the specific standard of evaluation, mastery-approach and mastery-avoidance goals were additionally divided into intrapersonal and task-based goals in the 3 x 2 model (Elliot, Murayama, & Pekrun, 2011). Intrapersonal mastery goals (approach or avoidance) focus on intrapersonal competence and refer to the self as the standard of evaluation ("I want to do better than before"). Task-based mastery goals (approach or avoidance) focus on task-based competence and refer to an absolute criterion as the standard of evaluation ("I want to be as good as possible"). Note that these more recent expanded achievement goal models do not make previous models obsolete and that usually just a subset of specific achievement goals is investigated at once (Elliot & Hulleman, 2017). In this dissertation, only achievement goals from the trichotomous model were investigated (i.e., mastery, performance-approach, and performance-avoidance goals). These are the most studied goals in the literature and the most relevant goals to students of most age groups (Elliot, 2005).

According to achievement goal theory, mastery goals are posited to be the most beneficial and performance-avoidance goals are posited to be the most detrimental of the three goals for students' achievement-relevant processes and outcomes. Performance-approach goals are posited to fall between mastery and performance-avoidance goals and to show some positive and some negative outcomes (Elliot & Hulleman, 2010). Concerning the motivational constructs that were investigated in the present research, achievement goals are posited to be subsequent to perceived competence (e.g., academic self-concept and self-efficacy)[4] and antecedent to interest and intrinsic motivation (Elliot & Hulleman, 2010). A high level of perceived competence is posited to predict high mastery and performance approach goals and a low level of perceived competence is posited to predict high performance-avoidance goals (e.g., Baranik et al., 2010). In turn, high mastery and performance-approach goals are posited to predict a high level of interest and high intrinsic motivation, whereas high performance-avoidance goals are assumed to predict a low level of interest and low intrinsic motivation (Elliot & Hulleman, 2017, Hulleman et al., 2010; Korn & Elliot, 2016). In addition, implicit assumptions about the changeability of intelligence (i.e., incremental theory versus entity

[4] An earlier hypothesis that perceived competence is a moderator of achievement goals was remained mostly unsupported by empirical research (Hong et al., 1999; Kaplan & Midgley, 1997)

theory; Cury, Da Fonseca, Rufo, & Sarrazin, 2002) and achievement motives (e.g., need for achievement, fear of failure; Elliot & Church, 1997) are posited to predict achievement goals. Moreover, basic human needs for competence, autonomy, and relatedness, as proposed by self-determination theory (Deci & Ryan, 1985), have also shown to predict (mastery) achievement goals in a growing number of studies (e.g., Lazarides & Raufelder, 2017; Ruzek & Schenke, 2018).

2.1.3. Self-Esteem.

Self-esteem, defined as the positive evaluations of oneself as a totality, is a fundamental and core human motive (Rosenberg et al., 1995; Rosenberg, Schooler, & Schoenbach, 1989). High self-esteem is a favorable global evaluation of one's own competence and accomplished achievements in subjectively important areas. Low self-esteem implies a dissatisfaction with oneself as a person and with previous achievements (Harter, 1990a). Different principles such as reflected appraisal, social comparison, and self-attributions contribute to the formation of self-esteem (Rosenberg, Schooler, & Schoenbach, 1989). In the hierarchical Shavelson et al. model of self-concept, self-esteem is defined as the most general level of self-concept (Shavelson, Hubner, & Stanton, 1976). Self-esteem shows positive correlations with job satisfaction and job performance (Judge & Bono, 2001), social relationships (Hendrick, Hendrick, & Adler, 1988; Neyer & Asendorpf, 2001), and subjective well-being (DeNeve & Cooper, 1998) and negative correlations with depression (Sowislo, & Orth, 2013) and health problems (O'Connor & Vallerand, 1998; Vingilis, Wade, & Adlaf, 1998). Furthermore, small positive relations between self-esteem and school achievement have been reported in the literature (e.g., Bowles, 1999; Davies & Brember, 1999; Hansford & Hattie, 1982).

2.1.4. Expectancy-Value Theory.

According to expectancy-value theory (Eccles et al., 1983; Wigfield & Eccles, 2002), student performance, choice, and persistence are most directly predicted by their expectancy of success and the subjective value they attach to success. In turn, task-specific beliefs such as perceptions of difficulty and assumptions of cost are assumed to influence personal expectancies and values. Expectancies of success are individuals' beliefs about how well they will master an upcoming task. These expectancies focus on personal competence beliefs and encompass individual competence in different areas (Wigfield & Eccles, 2002). The individual value of a task comprises four components: intrinsic value, attainment value, utility value, and cost. Intrinsic value is the enjoyment gained from performing the task for its own sake. Attainment value is the personal importance of being good at the task. Utility value is the usefulness of performing the task for further goals. Costs are all negative aspects that are associated with doing the task. Individuals' expectancies are more closely related to performance variables, whereas individuals' task values are stronger predictors of outcomes such as choice, effort, and persistence (Trautwein & Möller, 2016; Wigfield & Eccles, 2002). Furthermore, research suggests a positive interaction of expectancy and value (Nagengast et al. 2011; Trautwein et al., 2012). That is, the effect of expectancy on performance is additionally enhanced by a positive value.

In the literature, the constructs of expectancy and value have been operationalized in various ways. Hulleman, Barron, Kosovich, and Lazowski (2016) suggested a framework to allocate motivational constructs as either expectancy or value beliefs. In the present research, four of the investigated motivational constructs can be allocated to either expectancy or value beliefs in accordance with Hulleman or colleagues (2016). According to Hulleman and colleagues (2016), academic self-concept and academic self-efficacy are expectancy-related constructs, whereas intrinsic motivation and academic interest are examples of value-related constructs (i.e., intrinsic value).

2.1.5. Self-Determination Theory.

Self-determination theory is an approach that examines the conditions that elicit and sustain intrinsic and extrinsic motivation (Deci & Ryan, 1985; Ryan & Deci, 2000). Thus, in contrast to expectancy-value theory, self-determination theory focuses on the appearance of motivation rather than its outcomes. According to self-determination theory, the general and initial goal of individuals is self-determination, which can be derived from psychological needs for competence, autonomy, and relatedness. The need for competence is defined as the desire to be effective when dealing with the environment. Positive feedback tends to satisfy and negative feedback tends to undermine people's need for competence. The need for autonomy is the desire to control one's own life. People tend to interpret extrinsic rewards, directives, and deadlines as an external control of their behavior that undermine their need for autonomy, while making one's own decisions and opportunities for self-direction are in line with this need (Deci & Ryan, 2010). The need for relatedness is the desire to be connected with and caring for others. This need is undermined when one or one's work is ignored by others.

According to self-determination theory, there are two basic types of motivation: intrinsic and extrinsic motivation[5]. In contrast to intrinsic motivation, extrinsic motivation refers to the drive of activities in order to attain some separable and thus extrinsic outcome (Ryan & Deci, 2000). According to self-determination theory, intrinsic motivation for doing a task is enhanced when the psychological needs for competence, autonomy, and relatedness are satisfied. Thus, a challenging and novel task that can be solved efficiently, freedom of task selection and during the work process, and a connective social environment, where the individuals care for each other, will promote optimal intrinsic motivation for a task. Regarding the enhancement of extrinsic motivation, it is obvious that extrinsic rewards can stand in conflict with the need for autonomy. This has been impressively demonstrated by the so-called undermining effect of extrinsic rewards. The undermining effect means that extrinsic rewards for an intrinsic driven activity will foster extrinsic motivation, but undermine the enjoyment of doing the task (Deci, Koestner, Ryan, 1999). Thus, in contrast to intrinsic motivation, extrinsic motivation is not so much determined by the need for autonomy (Deci and Ryan, 2000b)[6]. On the other hand, satisfaction of the needs for competence and autonomy is proposed to foster extrinsic motivation (Deci and Ryan, 2000b). For example, those students who feel themselves respected and cared for by a teacher are more likely to accept extrinsic demands stated by this teacher. Similarly, adopting an extrinsic goal requires that one assumes to be efficacious with respect to it (Deci and Ryan, 2000b).

2.1.6. Volition Model of Action Phases.

According to the volition model of action phases (Heckhausen, & Heckhausen, 2006; 2018), individuals' motivated actions are directed by factors of the *person* and the *situation*, by the interaction of the *person* and the *situation*, and by the anticipated *outcome* of the *action* and its *consequences*. The basic model of action phases is depicted in Figure 1. All motivational influences that reside within the person are defined as *person factors*. *Person factors* can consist of universal behavior tendencies and basic needs as well as dispositions that vary between persons such as implicit motives, explicit motives, personal goals, and personality. In empirical research, individual dispositions are typically used to explain individual differences in behavior. However, according to Heckhausen and Heckhausen (2018), by only considering the personal factors, one would divide the world in thieves and

---

[5] Extrinsic motivation is additionally split into extrinsic motivation based on external, introjected, identified, and integrated regulation. Further, *amotivation* can be suggested as a third category of motivation (Ryan & Deci, 2000a; 2000b).

[6] Of note, integrated extrinsic motivation is more strongly dependent on the need for autonomy than externally regulated extrinsic motivation.

non-thieves, whereas, in reality, it is also the opportunity that makes a thief. *Situation factors* are defined as opportunities and possible incentives. *Situations* differ regarding the level of *situation-outcome* expectancies, *action-outcome* expectancies, and *outcome-consequences* expectancies. High *situation-outcome* expectancy means that a *situation* automatically leads to an *outcome*. For example, in a class where each student gets an A+ regardless of student achievement, the students would assume a high *situation-outcome* expectancy. Thus, this *situation* provides no incentive to act. In contrast, in classes where the grades are primarily determined by students' effort, students would assume high *action-outcome* expectancies. This situation provides incentives to act, particularly in cases where good grades are favorable for aimed *outcomes* such as praise from parents. *Actions* and *outcomes* can contain intrinsic incentives such as enjoyment of doing a task that can additionally motivate individuals. *Consequences* can contain extrinsic incentives that can be derived from the result of the *action* and the *outcome* (e.g., a well-paid job as a consequence of student effort and good grades in school). Finally, the interaction of the *person* and the *situation* means that based on *person factors* that vary between persons, the same *situations* are interpreted differently. Therefore, different persons will anticipate different incentives and show a different level motivation for actions, even in similar situations (Heckhausen, & Heckhausen, 2018).

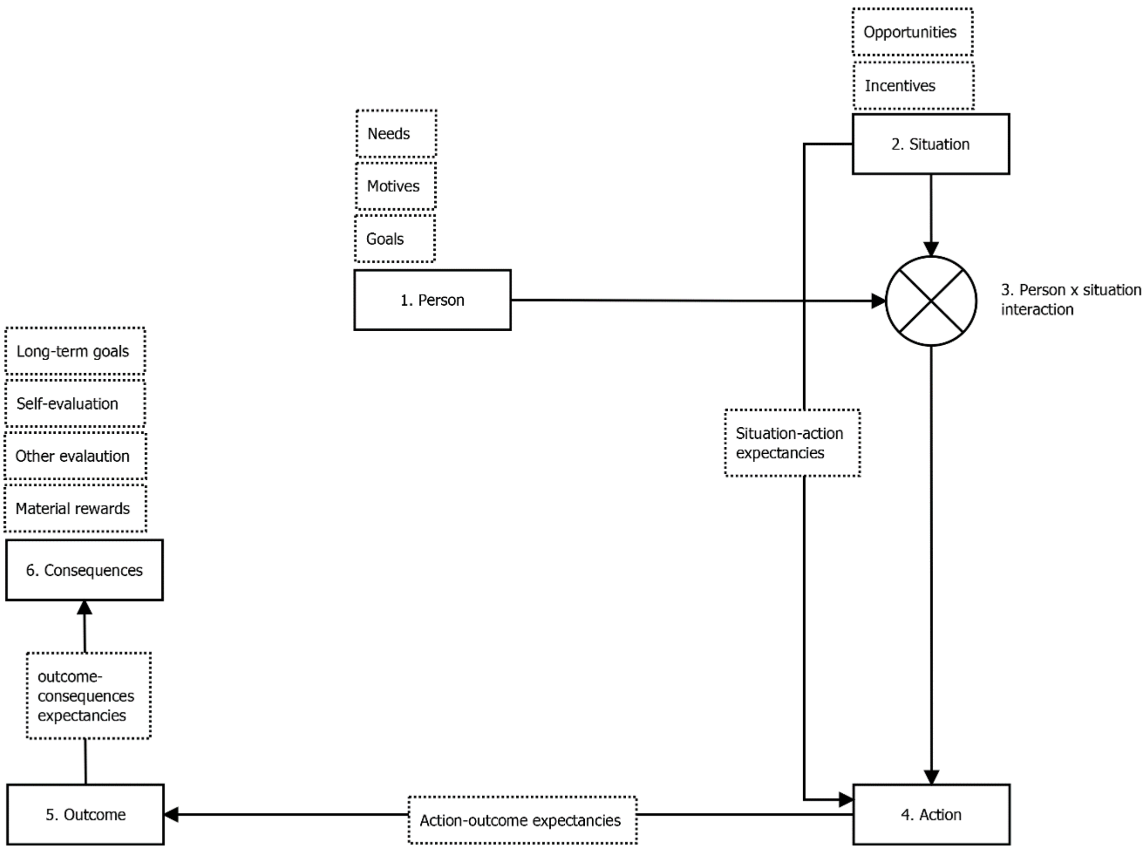


Figure 1. The volition model of action phases (Heckhausen, & Heckhausen, 2006; 2018).

2.1.7. An Integral Framework Based on the Volition Model of Action Phases.

Based on the volition model of action phases, Urhahne (2008) developed a framework to integrate several existing theories of achievement motivation (i.e., expectancy-value theory: Eccles et al., 1983; self-determination theory: Deci & Ryan, 1985; achievement goal theory: Dweck, 1986; achievement motive theory: McClelland, Atkinson, Clark, & Lowell, 1953; attribution theory; Kelley, 1967; social motivation theory: Baumeister & Leary, 1995) and motivational constructs considered by these theories. In the following, Urhahne's framework and his considerations of expectancy-value, self-determination, and achievement goals theories are discussed. His considerations about achievement motive, attribution, and social motivation theories are not presented in the following, because motivational constructs from these theories were not investigated in this dissertation. Finally, a suggestion is made on how self-esteem and academic self-concept could additionally be integrated into this framework and critiques from Spinath (2008) are presented.

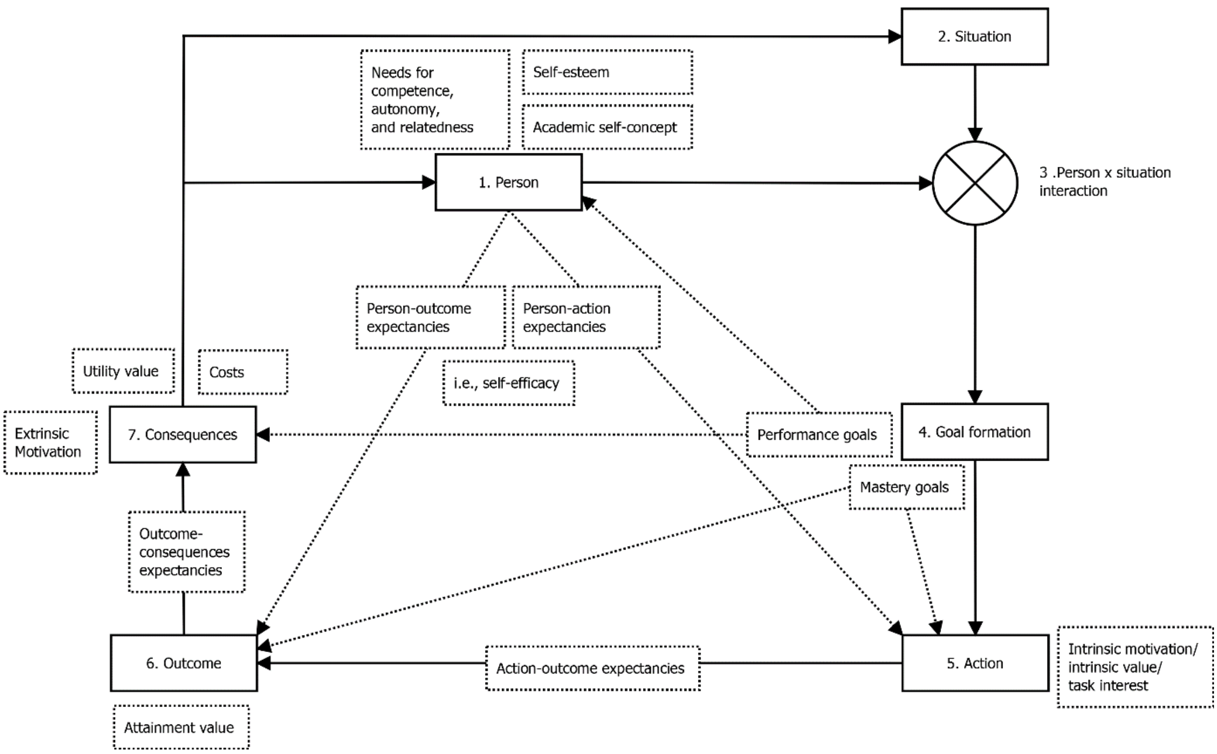


Figure 2. The volition model of action phases as an integrative framework for motivation aspects from expectancy-value, self-determination, achievement goals, and self-esteem theories. Performance goals comprise performance-approach and performance-avoidance goals.

In accordance to the volition model of action phases, Urhahne (2008) assumed that motivated actions are directed by *person factors*, *situation factors*, *the interaction of the* *person and the situation*, and by anticipated *outcomes* of the *actions* and their *consequences*. In addition, Urhahne suggested *goals formation* as a further phase that comes after the *interaction of the person and the situation* and before the *action*. In contrast to Heckhausen and Heckhausen (2018), Urhahne ordered the phases as a repetitive cycle to demonstrate that *consequences* of *actions* can, in turn, affect *person* and *situation factors*. The modified volition model of action phases is depicted in Figure 2.

Concerning the expectancies from expectancy-value theory and in accordance with the original model of action phases, Urhahne theorized that individuals assume *action-outcome* and *outcome-consequences* expectancies. In addition, Urhahne suggested that individuals assume, whether they can perform an action by themselves (e.g., "During the next week I am able to learn vocabulary one hour a day."). Finally, individuals assume whether they are able to achieve an outcome (e.g., "I am able to achieve a good grade on the next English test."). These *person-action* and *person-outcome* expectancies show great similarities to the construct of self-efficacy that is investigated in this dissertation. Concerning the values from expectancy-value theory, Urhahne theorized that the *action* phases can involve an intrinsic value, the *outcome* can involve attainment value and the *consequences* can involve utility value and costs for individuals. Regarding self-determination theory, Urhahne postulated that general psychological needs for competence, autonomy, and relatedness are examples of *person factors*. Opportunities to satisfy these needs in the *situation* will elicit and sustain intrinsic motivation during the *action*. Furthermore, Urhahne allocated extrinsic motivation exclusively to the *consequences*. Concerning achievement goal theory, Urhahne suggested that all achievement goals are created during the phase of *goal formation*, but also refered to further subsequent phases. Mastery goals have the purpose of develop competence. As competencies can be increased during *action* and verified by the *outcomes*, mastery goals focus on these two phases. Performance goals focus on the demonstration of competence and the avoidance of incompetence in comparison to others. The comparison to others is apparent in the *consequences* phase. Further, others are compared to the self, and the self-representation can, in turn, affect *person factors*. Therefore, performance goals (approach and avoidance) focus on the *consequences phase* and *person factors*.

Although Urhahne (2008) did not consider self-esteem and academic self-concept, both constructs could be additionally integrated into his framework and allocated to the *person factors*. Both constructs are relatively stable and trait-like evaluations of the self and refer to the sub-sum of previous experiences rather than to a single, concrete situation. Thus, analogous to other *person factors* mentioned by Heckhausen and Heckhausen (2018) such as motives and personality, self-esteem and self-concept can be defined as personal dispositions that vary between persons.

Note that the Urhahne framework to integrate different motivational theories was critized and rejected by Spinath (2008). Spinath's main criticism was that in contrast to Urhahne's theoretical model, in reality, *actions* are always part of the *situation* and motivation is always part of the *person*. In addition, Spinath argued that motivational constructs were sometimes seemingly arbitrarily assigned to the phases and that alternative allocations are equally (or even more) plausible. For example, she questioned why only performance goals and not mastery goals should affect *person factors*. Some of Spinath's criticism could be used to improve Urhahne's framework, while some critiques question the basic idea of the framework in itself. It is not the aim of this dissertation to create a new theoretical framework for achievement motivation. Thus, in this dissertation, Urhahne's model serves as a working framework to define and allocate the relevant motivational constructs, whereas it is noted that that this is not a perfect or unquestioned theoretical model.

2.2 Theoretical Views Regarding Motivational Declines in School-Aged Students

"Development is less gradual and more saltatory, suggestive of some ancient period of storm and stress" (Hall, 1904, p. xiii).

Nearly 100 years after Hall (1904) proposed his view of adolescence as a time of storm and stress, the assumption that adolescence is necessarily a turbulent time is still popular among parents, teachers, and adolescents (Buchanan, 2003). However, to a large extent, psychological research rejects the view that storm and stress are universal and inevitable in adolescence (for a review, see Arnett, 1999). However, when considering the motivational development of adolescents, some degree of storm and stress cannot be completely rejected, as longitudinal research indeed indicates a decline in motivation throughout students' school careers (e.g., Eccles et al., 1993; Gottfried et al., 2001; Wigfield & Wagner, 2005). In the following, several theoretical views explaining why student motivation often declines over the school career are presented. Then, conclusions based on the theoretical overview are made.

2.2.1. Cognitive-Developmental Approach.

One classic theoretical explanation for motivational decline in (primarily young) children and adolescents is based on Piaget's cognitive-developmental approach. Piaget

(1928a, 1928b) allocated children under seven to the *preoperational stage* and proposed that these children lack operational capabilities. It is difficult for children in the preoperational stage to distinguish between events and their own mental construction and representation of the events. Furthermore, they do not distinguish between abilities, luck, and effort to explain achievement. Therefore, preoperational children are unable to form coherent concepts and they perceive their abilities as unrealistically high (Butler, 2005). According to Piaget, older children develop logical reasoning, are more aware of external events, and become less egocentric. Thus, with advancing cognitive development, children perceive their abilities more realistically and, consequently, as lower. Moreover, research suggests that at later stages, children and adolescents increasingly base their self-evaluations on external feedback and social comparison (Harter, 1999). This additionally decreases their self-evaluations and leads to a better integration of positive and negative self-perceptions and a more balanced self-view.

The cognitive-developmental approach also offers an explanation for higher persistence and effort in young children. Researchers have argued that failure has a smaller negative impact on younger children than older children, as younger children do not attribute failure to deficiencies in their performance. Furthermore, young children believe that more effort will inevitably lead them to achieve the desired performance (Butler, 2005).

2.2.2. Environmental Influences.

A further explanation of motivational decline is students' experiences in school after the transition from elementary to middle school. The learning environment plays a crucial role in student motivation (Eccles et. al., 1993; Thapa et al., 2013; Wigfield, Eccles, & Rodriguez, 1998). For example, school climate is related to students' self-esteem (Hoge, Smit, & Hanson, 1990), self-concept (Heal, 1978), incentive to learn (Goodenow & Crady, 1993; Goudas & Biddle, 1994), and school satisfaction (Zullig, Huebner, & Patton, 2010).

Researchers have argued that the quality of the school environment usually decreases after the transition to middle school (Anderman & Maehr, 1994; Eccles et al., 1993; Wigfield & Wagner, 2005). As a consequence, negative changes to the school environment might affect student motivation. Changes to the school environment include the disruption of students' friendship networks, teachers teaching a larger number of students and not getting to know their students as well as in elementary school, and a decline of family involvement during the middle school years (Wigfield & Wagner, 2005). After transition to middle school, students have fewer opportunities to participate in decision-making regarding their own learning as whole class tasks are applied more frequently than in elementary school (Midgley

& Feldlaufer, 1987). Teachers and students in middle school perceive the school culture to be more performance-focused and less task-focused in comparison to teachers and students in elementary school (Midgley, Anderman, & Hicks, 1995). Furthermore, compared to elementary school teachers, middle school teachers exert more control and disciplinary behaviors (e.g., Mendez & Knoff, 2003; Midgley & Feldlaufer, 1987; Midgley, Feldlaufer, & Eccles, 1988) and feel less effective (Midgley, Feldlaufer, & Eccles, 1989; Midgley, Anderman, & Hicks, 1995). Eccles and colleagues (1993) suggest the more bureaucratic nature and the larger size of middle schools compared to elementary schools may explain these changes.

2.2.3 Stage-Environment Fit Theory.

In addition to negative environmental changes as an explanation of student motivation decline, Eccles and colleagues (1989, 1993) suggested that, to some degree, the motivational decline is a result of a growing mismatch between students' needs and the opportunities offered to them by their school (and home) environment. According to Eccles and colleagues (1989, 1993), developing psychological and social needs that are notably different from those of children and adults can be expected for adolescents. For example, with increasing age and pubertal development, needs for autonomy, individual identity, safe and intellectually challenging environments, self-focus, and peer orientation rise (Eccles et al., 1993; Simmons & Blyth, 1987). In addition, adolescents often attempt to create a recognizable distance between themselves and acquainted adults. In contrast, the opportunities to satisfy these needs offered by the environment decrease with rising grade level, especially after the transition to middle school (see the section above on environmental influences). This combination of developmental changes in both the individual and the environment results in a poor stage-environment fit, leading to a motivational decline in many students.

To test the implications of stage-environment fit theory, Midgley and colleagues (1989; for an overview, see also Eccles et al., 1993) conducted a large-scale, two-year, four-wave longitudinal study (i.e., The Michigan Study of Adolescent Life Transition MSALT), which followed a sample of 1500 adolescents moving from the sixth grade of an elementary school into the seventh grade of a middle school. Individual and environmental factors in the two schools were examined with regard to stage-environment fit theory. Students who moved from high-efficacy math teachers in elementary school to low-efficacy math teachers in middle school reported lower motivation in math than adolescents who experienced no change of teacher efficacy or moved from low-efficacy to high-efficacy teachers. Most middle school math teachers reported low self-efficacy. For student-teacher relationships, Midgley and colleagues (1989) reported similar findings. Furthermore, both the adolescents and their teachers reported fewer opportunities for participation in classroom decision-making in the seventh grade than in the sixth grade, whereas adolescents' desire for participation in classroom decision-making increased after the school transition. Interestingly, the female adolescents who were physically more mature expressed a greater desire for classroom decision-making, but reported fewer opportunities for participation, compared to the less physically mature female adolescents in the same classrooms. That is, students with faster developing needs perceived the new school environment as particularly inappropriate. These findings are in support of stage-environment fit as an explanation for the decline of adolescent achievement motivation.

2.2.4. Self-Determination Theory.

In the literature, self-determination theory has often been applied to explain student motivation (e.g., Lazarides & Rubach, 2017; Otis, Grouzet, & Pelletier, 2005) and self-esteem declines (e.g., Birkeland, Melkevik, Holsen, & Wold, 2012). According to self-determination theory, decreases of student motivation are the result of dissatisfaction of three fundamental human needs (i.e., the needs for achievement, autonomy and relatedness). For a more detailed description of self-determination theory, see section 2.1.4.

2.2.5. Conclusions from Theoretical Views Regarding Student Motivational Decline.

Motivational decline in school-aged students can be explained by various theoretical views. In principle, the theoretical views do not contradict each other and can be used complementarily. Motivational decline in elementary children is certainly partly based on students' cognitive development, leading to more realistic, and, consequently, lower self-expectancies (i.e., cognitive-development approach). It is up for discussion whether more realistic, but lower self-beliefs are favorable or undesirable. Motivational decline after the transition from elementary to middle school and during middle school can be reasonably explained by (school) environmental changes and a growing mismatch between student needs and opportunities (i.e., environmental influences and stage-environment fit theory). These truly undesirable effects could be avoided by better (middle) school practice (Eccles & Roser, 2009). In addition, self-determination theory is not limited to a specific developmental stage. According to this view, human needs for achievement, autonomy, and relatedness are fundamental and whenever these needs are not satisfied, a motivational drop can be expected.

For the most part, this dissertation refers to stage-environmental fit theory as a framework to explain motivational decline. First, this is because the empirical studies reported in Articles 2 and 3 were conducted in middle schools (i.e., the school stage that is the focus of stage-environment fit theory). Second, all motivational constructs that were investigated in the three articles were associated with stage-environment fit theory in previous research (i.e., self-esteem: Booth & Gerard, 2012; Gutman & Eccles, 2007; academic self-concept: Fenzel, 2000; Parker, 2010); academic self-efficacy: Booth & Gerard, 2012; Bong, 2005); intrinsic motivation: Lazarides & Raufelder, 2017; Lepper, Corpus, & Iyengar, 2005; achievement goals: Roeser, Midgley, & Urdan, 1996; Bong, 2005).

2.3 The Present Research

After consideration of the theoretical background, several things become clear. In general, all psychological process that initiate, organize, sustain, control, and evaluate purposeful actions of students in academic settings can be summarized under the broad term *achievement motivation* (see Grassinger, Dickhäuser, & Dresel, 2019). Depending on the theory, various reasons for why students learn can be differentiated and interactions between these reasons can be expected. Most reasons why students learn can be subsumed under the concepts of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, interest, or achievement goals (mastery, performance-approach, and performance-avoidance goals). These constructs can be allocated to one or several action phases of the modified volition model of action phases (Heckhausen, & Heckhausen, 2006; 2018; Urhahne, 2008). In the literature, student achievement motivation decline over the course of students' school career have been reported in a growing number of studies (e.g., Anderman, & Midgley, 1997; Gottfried et al., 2001; Wigfield & Wagner, 2005) and explained by several theories. In the following, three research questions regarding achievement motivation development are raised. Further, information is provided on how these research questions were investigated in the three articles that are presented in this dissertation.

2.3.1 Research Question 1: How Does Student Achievement Motivation Develop During the School Career?

According to stage environment-fit theory, students' mean-levels of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, interest, and mastery achievement goals decline as a function of the growing gap between student needs and opportunities afforded by their environment (primarily in middle schools). However, it is unclear whether these constructs, stemming from varying theories and allocated to different phases of the volition model of action phases, are affected by stage-environment mismatches in the same way. Theoretically, it is plausible that the quantity and direction of mean-level changes differs between the varying motivational constructs.

Given that self-esteem, academic self-concept, and academic self-efficacy refer to different levels of self-evaluation, the effects of stage-environment fit in school on the development of these constructs should display a different (quantitative) proportion. As self-esteem refers to a self-evaluation of oneself as a whole (Rosenberg et al., 1995), it is questionable whether mismatches between one's needs and opportunities in school should play a major role for this general construct. In contrast, a strong impact of the stage-environment mismatch could be expected for academic self-concept, as this construct refers to the self-evaluation of one's competences in school or a specific school domain (Marsh & Craven, 1997; Marsh & Shavelson, 1985; Shavelson et al., 1976). As self-efficacy refers to one's capability to solve a particular specific situation (Bandura, 1993; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Pajares, 1996), more variation of self-efficacy development (based on situational factors) can be expected. Therefore, it is reasonable to assume only a limited effect of the (general) stage-environment mismatch on students' specific self-efficacy. Accordingly, in comparison to the less specific self-esteem and the more specific self-efficacy, larger declines could be expected for students' academic self-concepts.

Regarding students' intrinsic motivation and interest, declines can be expected during the school career (Eccles et al., 1993). However, the magnitude of the declines is uncertain as no meta-analysis has investigated the average amount of decline on these constructs. Regarding students' achievement goals, stage-environment fit theorists have classically predicted a shift from mastery to performance goals (Anderman & Midgley, 1997; Eccles et al., 1993). However, these predictions were made before performance goals were split into performance-approach and performance-avoidance goals and before some adaptive effects of performance-approach goals were identified. Taking into account these updates of achievement goal theory (Elliot & Harackiewicz, 1996), two alterative sets of predictions for the development of achievement goals may be derived in accordance with the basic assumptions of stage-environment fit theory. One alternative is that the stage-environment mismatch results in students' decreased commitment to competence per se, including all types of goals (Paulick, Waterman, & Nückles, 2013). According to self-determination theory, this psychological condition could be labeled as *amotivation* (Ryan & Deci, 2000a; 2000b). Another alternative is a decreased commitment to mastery and performance approach goals and an increased commitment to performance-avoidance goals as a mechanism of self-protection (Alicke & Sedikides, 2011; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001).

Empirically, for nearly all of the considered motivational constructs, varying development trajectories (i.e., increases, decreases, or no change) have been reported in the literature (e.g., Anderman & Midgley, 1997; Bong, 2005; Bronstein, Chubb, Fertman, & Ross, 1997; Ginsburg, & Herrera, 2005; Gottfried, Fleming, & Gottfried, 2001)[7]. However, the shape and magnitude of development during the school career have not been investigated by meta-analytic methods for any of these constructs, except self-esteem (Orth, Erol, & Luciano, 2018).

To sum up, the literature shows empirical and theoretical inconsistencies regarding students' mean levels of development on central motivational constructs during their school careers. The main aim of this dissertation was to depict a clear empirical picture of student motivational development during the school career. Therefore, all three presented articles investigated the mean-level development on (some) motivational constructs. Article 1 aimed to dispel the inconsistencies resulting from previous empirical research and quantify the average magnitude of motivational development of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and mastery, performance-approach, and performance-avoidance goals. Therefore, a meta-analysis of longitudinal studies that investigated student development on these constructs was conducted. Students from first to 13th grade were included in this meta-analysis. In Article 2, two new longitudinal studies examining German middle school students (Study 1: n = 745, 5th grade to 7th grade; Study 1 n = 1420, 5th grade to 8th grade) from the highest school track (i.e., Gymnasium) are presented. This article focused on student development of interest and mastery, performance-approach, and performance-avoidance goals. In Article 3, a longitudinal study of gifted students (n = 296, 5th grade to 7th grade) in regular and special classes is reported. In this study, the development of academic self-concept in math was investigated.

2.3.2 Research Question 2: What Are the Correlates of Achievement Motivation Development?

Considering that an overall negative development of student achievement motivation during the school career can be expected based on previous research, two additional questions arise. First, (a) which conditions are antecedent to declines on motivational constructs? Second, (b) which changes are associated with changes in achievement motivation?

Research question 2a. The first question investigates causes of motivational decline. According to stage-environment fit theory, motivational decline is explained by a growing gap between student needs and afforded opportunities. It is assumed that this gap primarily occurs after the transition from elementary to middle school. Indeed, declines on motivational

[7] In article 1, previous findings regarding the development in self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals are described in more detail.

constructs after the transition to middle school have been reported by various longitudinal studies (e.g., Midgley et al., 1989; Paulick, Waterman, & Nückles, 2013). However, in addition, declines have also been found before (e.g., Weidinger, Steinmayr, & Spinath, 2017) and after school transition (e.g., Neel & Fuligni, 2012). Therefore, it remains unclear, whether motivational decline during the transition are greater than motivational decline during other school stages. Further, stage-environment fit theory assumptions are based on empirical evidence that dates back more than three decades (Eccles et al., 1989; 1993). Administrative changes in the educational system (e.g., as a result of the PISA survey; Beycioglu, & Kondakci, 2014; Grek, 2009) could have since made the assumptions regarding middle schools obsolete.

In this dissertation, various factors were investigated as possible predictors of motivational development in Articles 1 and 2. In the meta-analysis presented in Article 1, school transition, school stage (elementary, middle, and high school), grade level, particular academic domain, date of data collection, year of publication, and geographic location were tested as predictors of development of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and mastery, performance-approach, and performance-avoidance goals. In the longitudinal studies reported in Article 2, cognitive ability, gender, parents' highest educational degree, and language background were investigated as predictors of development of interest and mastery, performance-approach, and performance-avoidance goals. Note, predictors of motivational development were not the focus of Article 2 and were simply considered as control variables.

Research question 2b. The second question investigates whether development of motivational constructs is related to development of further constructs. Different correlates were derived for different motivational constructs depending on the underlying theory. For example, in achievement goal research, interest and achievement are "gold standard" outcomes that are assumed to be subsequent to achievement goals (Korn & Elliot, 2016, p. 4). From this perspective, one could also assume a associated development of achievement goals and their typical outcomes. However, only a few studies have investigated the correlation between achievement goals change and interest or achievement change (e.g., Lazarides & Raufelder, 2017; Shim et al., 2008). These studies indicated positive (or non-significant) relations between the development of mastery and performance-approach goals and the development of interest and achievement. Note that performance-avoidance goals were not investigated in these studies.

In this dissertation, the joint development of all three achievement goals as well as interest and achievement was investigated in the two longitudinal studies presented in Article 2. In addition, in the longitudinal study reported in Article 3, the joint development of academic self-concept and achievement was depicted. Although this relation was not the focus of Article 3, an associated development of academic self-concept could be expected, as reciprocal effects between these variables have been shown in previous research (Huang, 2011).

2.3.3 Research Question 3: How Does Ability Grouping Affect Student Motivational Development?

Stage environment fit theory assumes motivational decline as a consequence of a person-environment mismatch. Eccles and Roser (2009) suggested adjusting the school environment to student needs that can be expected in particular developmental stages (e.g., classroom practices that support student autonomy in the middle school stage) as one solution to overcome this mismatch. School reform programs that have been conducted in accordance to stage-environment fit theory support this suggestion (e.g., Maehr & Midgley, 1996; Midgley, Anderman, & Hicks, 1995). However, it is important to note that even students who are the same age differ with regard to their personality, ability level, and needs. Thus, general environment adjustments are more effective for some students than for others, when differences between students are not considered. This is especially true for students who strongly deviate from the average, such as high-ability or gifted students.

In educational practice, ability grouping is a common approach to ensure a better fit between students and their school environment by taking interpersonal differences into account. The effects of ability grouping on student achievement motivation and achievement have been investigated in numerous studies (for reviews, see Seaton, Marsh, & Craven, 2009; Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016). However, this research reveals a rather inconsistent picture – especially for high-ability students. Regarding student achievement, minimal benefits were reported in meta-analyses (e.g., Kulik, 1992). Regarding student self-concepts, some research showed that high-ability grouping was associated with "costs" (e.g., Craven, Marsh, & Print, 2000; Preckel, Zeidner, Goetz, & Schleyer, 2008; Nagengast & Marsh, 2012), whereas other research indicated no effects (e.g., Herrmann et al., 2016; Preckel & Brüll, 2010). Inconsistent results can be explained by the "big-fish-little-pond-effect" that is produced by contrast and assimilation effects (Marsh & Parker, 1984). Negative effects of ability grouping on self-concept can be explained by the contrast effects. Note that student's most salient social reference group is their own class (Wouters, Colpin, van Damme, de Laet, & Verschueren, 2013). Therefore, external comparison processes can lower high-ability students' self-concepts when they are grouped with other high-ability students and compare themselves to this high achieving reference group (i.e., their class mates; Marsh & Parker, 1984). On the other hand, some researchers have noted contrast effects are short term and not universal (Dai, 2004). In addition, being a part of a high-ability class can also make students proud of themselves and improve their self-concepts by means of assimilation effects (Seaton et al., 2008). Overall, only few studies have investigated the effects of grouping on the development of academic self-concept over time.

To conclude, the research is inconsistent regarding the benefits and costs of ability grouping. Therefore, in the longitudinal study reported in Article 3, effects of grouping on the development of academic self-concept and achievement were investigated in a sample of high-ability students. High-ability students visited either regular classes of the highest educational track (i.e., “Gymnasium”) or special classes for gifted students. Notice that selection bias are a common problem when investigating the ability grouping of high-ability students (Dai, Swanson, & Cheng, 2011). Therefore, differences between students in special and regular classes on various variables, such as cognitive ability, sex, or age, were controlled by using propensity score matching.

Chapter 3. Article 1

Development of Motivational Variables and Self-Esteem During the School Career: A Meta-Analysis of Longitudinal Studies

Vsevolod Scherrer and Franzis Preckel

3.1 Abstract

Theoretical approaches and empirical research suggest a decline in the levels of motivational variables and self-esteem among students during the school career. However, precise statements about the magnitude of the change remain elusive. Conducting a meta-analysis of 107 independent longitudinal studies with 912 effect sizes, we found an overall decrease of Glass's $\Delta = -.108$ over an average duration of 1.654 years. Change significantly differed by construct with the largest decreases in intrinsic motivation, math and language academic self-concepts, mastery achievement goals, and performance-approach achievement goals. There were no significant mean-level changes in self-esteem, general academic self-concept, academic self-efficacy, and performance-avoidance achievement goals. School stage and transition to middle school or high school were not significantly associated with the change. Findings generalized over academic domain and questionnaire used for all constructs except for academic self-concept. The decline was larger in Europe than in North America or Asia.

Key words: motivation, self-esteem, longitudinal research, meta-analysis, stage-environment fit

3.2 Introduction

Most researchers share the view of declining motivation during the school career (e.g., Eccles et al., 1993; Gottfried et al., 2011; Wigfield & Wagner, 2005). Reviewing the literature of motivational development, Eccles and colleagues (1993) stated: "… developmental declines have been documented for such motivational constructs as interest in school; intrinsic motivation; self-concept and self-perception; and confidence in one's intellectual abilities, especially following failure…" (p. 90-91). A vast amount of recent empirical findings also documents this motivational decline throughout the school career (e.g., Reddy, Rhodes, & Mulhall, 2003; Shapka & Keating, 2005; Wang & Pomerantz, 2009). For adolescents, the decline in levels of motivational variables is often explained by stage-environment fit (SEF) theory (Eccles et al., 1993). According to SEF theory, the motivational decline in part results from a mismatch between adolescents' developing needs and the opportunities afforded by the school environment. So far no quantitative review or meta-analysis has examined the motivational change during the school career. Therefore, precise statements about the magnitude of the change are missing. It is further unknown if this change is more pronounced at certain grade levels (e.g., in adolescence after middle-school transition) or if it differs by motivational construct.

With the present study, we want to close this research gap. Using meta-analytic methods we investigated the development of psychological constructs that were addressed by Eccles and colleagues (1993), assessed the magnitude of change in these constructs during the school career (i.e., first grade to thirteenth grade), and investigated whether the change was related to certain events like the transition to middle-school. Our theoretical expectations were primarily grounded on SEF theory as this theory has exerted a major influence on motivational research in the last three decades (e.g., Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Zimmer-Gembeck, Chipuer, Hanisch, Creed, & McGregor, 2006). Eccles and colleagues (1993) applied SEF theory to a wide range of theoretically distinguishable but sometimes overlapping psychological constructs such as intrinsic motivation, self-concept, self-perceptions, confidence in one's intellectual abilities, test anxiety, learned helplessness responses to failure, task mastery, or self-esteem. In the present meta-analysis, we investigate the development of a portion of these constructs, that is, self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals. For these constructs, a sufficient amount of longitudinal investigations are available and their development has been frequently explained by SEF theory (e.g., self-esteem: Booth & Gerard, 2012; Gutman & Eccles, 2007; academic self-concept: Fenzel, 2000; Parker, 2010; academic self-efficacy:

Booth & Gerard, 2012; Bong, 2005; intrinsic motivation: Lazarides & Raufelder, 2017; Lepper, Corpus, & Iyengar, 2005; achievement goals: Roeser, Midgley, & Urdan, 1996, Bong, 2005). We subsume these constructs under the broader categories of motivational variables and self-esteem, and frame the motivational variables in terms of competence motivation, defined as the way in which individuals energize and direct their behavior to reach effectiveness, ability, sufficiency, or success (Elliot, Dweck, & Yeager, 2017). By using the term *competence motivation* we do not reject the utility of the frequently used term *achievement motivation* but follow the arguments of Elliot, Dweck, and Yeager (2017, p. 3) according to which the term competence offers a more precise, clear, and broadly applicable definition than the term achievement.

The definition of competence motivation as the way in which individuals energize and direct their behavior can be directly applied to intrinsic motivation and achievement goals. While intrinsic motivation refers to pleasure and interest in learning for its own sake (Gottfried, 1990), achievement goals guide behavior to a future-focused competence-related end state (Elliot & Thrash, 2002). We further allocate academic self-efficacy and academic self-concept to the broader category of competence motivation. Both constructs partly explain why individuals choose to engage or disengage in different activities (Eccles & Wigfield, 2002; Ferla, Valcke, & Schuyten, 2010, Schunk, 1991). Academic self-efficacy refers to beliefs about one's future performance on a specific task and therefore influences students' persistence and task engagement (Pajares, 1996). Academic self-concept refers to beliefs about one's current abilities in broader academic domains and is featured prominently in major accounts of human motivation including expectancy-value theory (Eccles et al., 1983) and self-determination theory (Deci & Ryan, 2000; for an overview see Trautwein & Möller, 2016). In line with our framing, Murphy and Alexander (2000) defined self-competence, self-efficacy, intrinsic motivation, and goal-orientation as fundamental terms within the motivation literature associated with the study of academic achievement and academic development.

We assign self-esteem to a single category. SEF theory has been frequently applied to self-esteem. However, self-esteem comprises the overall emotional evaluation of one's self (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995), which is related to motivation but not a measure of competence motivation itself. In sum, in our meta-analysis we included constructs whose development was frequently explained by the SEF theory in the last decades, and organized them into the categories of competence motivation and self-esteem.

3.2.1. Stage-Environment Fit (SEF) Theory

SEF theory (Eccles et al., 1989, 1993) provides a prominent approach to explain students' decline in motivational variables and self-esteem. Eccles and colleagues (1989, 1993) advanced the hypothesis that some of the negative psychological changes associated with adolescent development result from a mismatch between the developing needs of the students and the opportunities offered to them by their social environment. A case in point is adolescents' experiences after the transition from elementary to secondary school (Thapa et al., 2013; Wigfield, Eccles, & Rodriguez, 1998). Researchers have argued that the quality of the school environment usually decreases after the transition to middle school (Anderman & Maehr, 1994; Eccles et al., 1993; Wigfield & Wagner, 2005). Changes in the school environment include the disruption of students' friendship networks, a decline in family involvement during the middle-school years; further, teachers lecture a larger number of students and do not get to know their students as well as in elementary school (Wigfield & Wagner, 2005). After transition to middle school, students have fewer opportunities to participate in decision-making regarding their own learning as whole class tasks are applied more frequently than in elementary school (Midgley & Feldlaufer, 1987). Teachers and students in middle school perceive the school culture as more performance-focused and less task-focused in comparison to teachers and students in elementary school (Midgley, Anderman, & Hicks, 1995).

On the other hand, students' developing needs include a reasonably safe and intellectually challenging environment, greater autonomy from parents' control, self-determination, peer orientation (including sexual relationships), self-focus and self-consciousness, and individual identity (Eccles et al., 1989, 1993; Simmons & Blyth, 1987). These needs increase with the age of the students and their pubertal development. According to SEF theory (Eccles et al., 1989, 1993), the gap between students' needs and opportunities offered in the school environment widens primarily after the transition to middle school as the opportunities to satisfy these needs offered by the environment often decrease. Therefore, the combination of developmental changes in both the individual and the environment leads to a bigger motivational decline than each would predict alone.

In 1993, Eccles and colleagues carried out a comprehensive review of the motivational literature available at that time and brought forward evidence for SEF theory. However, some open questions remain. First, SEF theory was introduced as a framework that can be applied to a wide range of psychological constructs, but it is unclear whether different constructs change in the same way. That is, it is unclear whether constructs like academic self-concept or self-esteem decline a comparable amount and during the same developmental phase. Second, it is unclear whether the decrease happens primarily after the transition to middle school or if it is a rather steady process during students' school career. Third, the empirical evidence present at the conceptualization of SEF theory dates back more than three decades. Since then, administrative changes in the educational system have significantly changed school environments (e.g., Au, 2011; Borman, Hewes, Overman, & Brown, 2003; Ertmer & Ottenbreit-Leftwich, 2010). One can assume that these changes could affect the motivational development of students. Thus, it is unclear if the conclusions of the review by Eccles and colleagues (1993) still apply today. The current meta-analysis of recent longitudinal research in motivational variables and self-esteem aims to shed light on these open questions by providing a high-resolution picture of the development of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals during the school career.

3.2.2. Self-Esteem

Self-esteem comprises an individual's positive or negative attitude towards the self as a totality (Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995). It is the overall emotional evaluation of one's self-worth or a value judgement of oneself (Burrus & Brenneman, 2016). High self-esteem refers to a highly favorable global evaluation of one's own competence in subjectively important domains, whereas low self-esteem is the outcome of negative assessments and subjective weaknesses (Harter, 1990a). In the hierarchical Shavelson model of self-concept (Shavelson, Hubner, & Stanton, 1976), self-esteem presents the most general level of self-concept, with underlying specific domains such as academic self-concept and physical self-concept.

A high level of self-esteem is positively related to job satisfaction and job performance (Judge & Bono, 2001), social relationships (Hendrick, Hendrick, & Adler, 1988; Neyer & Asendorpf, 2001), and subjective well-being (DeNeve & Cooper, 1998). A low level of self-esteem is linked to depression (Sowislo, & Orth, 2013) and health problems (O'Connor & Vallerand, 1998; Vingilis, Wade, & Adlaf, 1998). Many studies that investigated the relation of self-esteem and school performance reported small but significant positive correlations (e.g., Bowles, 1999; Davies & Brember, 1999; Hansford & Hattie, 1982). However, the results of longitudinal investigations do not support strong causal effects of self-esteem on school achievement (Baumeister et al., 2003; Bachman & O'Malley, 1986).

With regard to the development of self-esteem, researchers assume that it is relatively stable (Harter, 1990b). In a meta-analysis of retest correlations, Trzesniewski, Donnellan, and

Robins (2003) observed increasing rank-order stability throughout adolescence and early adulthood and declining rank-order stabilities in mid or late adulthood.

The evidence regarding whether the mean level of self-esteem decreases, increases, or does not change during the school career is mixed. Using data from 326,641 individuals from age 9 to 90, Robins, Trzesniewski, Gosling, and Potter (2002) found high self-esteem in the youngest age group, followed by a decline over childhood and adolescence. The implications of this study are limited by its cross-sectional design and selective bias when recruiting the sample. Some longitudinal studies have also observed a decline in self-esteem through adolescence (e.g., Greene, Way, & Pahl, 2006; Reddy, Rhodes, & Mulhall, 2003). Other studies found an increase in self-esteem (e.g., Harris et al., 2015; Way & Robinson, 2003), or no change over time (Chubb, Fertman, & Ross, 1997; Granleese & Joseph, 1994). Therefore, the developmental trend of self-esteem through the school career remains elusive.

3.2.3. Academic Self-Concept

Academic self-concept refers to the self-evaluation of one's ability in general (e.g., "I am good at school") or in a specific academic domain (e.g., "I am good at Math"; Marsh & Craven, 1997; Marsh & Shavelson, 1985; Shavelson, Hubner, & Stanton, 1976). Its structural organization is hierarchical, with a general academic self-concept at the apex of the hierarchy and domain-specific factors like mathematical academic self-concept or verbal academic self-concept at a subordinate level (Brunner et al., 2010).

Academic self-concept is positively related to students' effort (e.g., Skaalvik & Rankin, 1995), cognitive strategy use and self-regulation (Pintrich & De Groot, 1990), mastery achievement goals (Bong, 2001; Roeser, Midgley, & Urdan, 1996), and intrinsic motivation, interest, or academic value (Gottfried, 1990; Meece, Blumenfeld, & Hoyle, 1988). Academic self-concept and academic achievement show positive reciprocal relations over time (Huang, 2011). Longitudinal research suggests that correlations between domain-specific academic self-concept and domain-specific performance increase with students' age (Denissen, Zarrett, & Eccles, 2007), which indicates that the domain specificity of academic self-concept increases through the school career.

Regarding its development, academic self-concept shows high retest correlations that sometimes are even higher than that of the corresponding achievement (Marsh & Yeung, 1998; Niepel, Brunner, & Preckel, 2014a). Note that young children often generally view their abilities as unrealistically high and undifferentiated and this optimism persists even in the face of failure (Stipek & Daniels, 1988). During elementary school, however, self-

perceptions of ability become not only lower but also more realistic, more differentiated, and more responsive to underlying information (Harter, 1990b, 1999; Schmidt et al., 2017).
A mean level decrease in general academic self-concept (e.g., Shapka, & Keating, 2005; Spinath & Spinath, 2005), math academic self-concept (e.g., Dickhäuser et al., 2017; Wigfield et al., 1997), and reading academic self-concept (e.g., Wigfield et al., 1997) through the school career was observed in multiple longitudinal studies. Other longitudinal studies indicated no significant change (Bronstein, Ginsburg, & Herrera, 2005) or even an increase in general academic self-concept (Steiger, Allemand, Robins, & Fend, 2014). Therefore, evidence concerning the magnitude of change in general and in domain-specific academic self-concepts remains inconclusive.

3.2.4. Academic Self-Efficacy

Academic self-efficacy refers to students' specific beliefs about their efficacy to regulate their own learning activities and to master difficult subject matters (Bandura, 1993; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Wigfield & Eccles, 2002). According to Pajares (1996), academic self-efficacy influences students' choice behavior, how much effort they will expend on an activity, and their thought patterns and emotional reactions. Students with low academic self-efficacy may believe a task is tougher than it really is, and thus do not engage in the task or expend much effort on it. On the other hand, high academic self-efficacy students create feelings of serenity in approaching even difficult tasks, choose to engage in tasks, and put a lot of effort in accomplishing those (Pajares, 1996).

The academic self-efficacy construct shows an obvious and crucial similarity to the academic self-concept construct as the individual's sense of competence is a key part of both constructs (Wigfield & Eccles, 2002). In the literature, academic self-concept and academic self-efficacy are often used interchangeably as high correlations of these constructs have been obtained in empirical research (Eccles & Wigfield, 2002; Trautwein & Möller, 2016). However, academic self-concept and academic self-efficacy can be distinguished at both the theoretical level (Bong & Skaalvik, 2003) and the empirical level (Ferla, Valcke, & Cai, 2010; Scherer, 2013). For instance, academic self-efficacy is usually measured quite specifically and emphasizes the individual's own sense of whether he or she can accomplish a particular task, whereas academic self-concept comprises the beliefs about one's competence at a domain-specific level and includes comparative processes regarding competence in other domains and the competence of other individuals (Wigfield & Eccles, 2002).

Academic self-efficacy is positively related to course selection (Hackett & Betz, 1989; Pajares, & Miller, 1995), intrinsic motivation (Zimmerman & Kitsantas, 1999), academic aspirations (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001), and career choices (Bandura, Barbaranelli, Caprara, & Pastorelli, 2001). Several meta-analyses indicated a robust positive relationship between academic self-efficacy and academic achievement (Multon, Brown, & Lent, 1991; Richardson, Abraham & Bond, 2012; Robbins, Lauver, Langley, & Carlstrom, 2004).

Concerning its development, the rank-order stability of academic self-efficacy is moderate and slightly smaller than the rank-order stability of academic self-concept (e.g., Bong & Skaalvik, 2003; Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011; Pajares & Graham, 1999). Although most longitudinal investigations indicated a decrease in mean level academic self-efficacy through the school career (e.g., Anderman, Maehr, & Midgley, 1999; Pajares & Graham, 1999; Urdan & Midgley, 2003), other studies showed an increase (Shell, Colvin, & Bruning, 1995). The mixed results regarding the mean level development might be explained by the different grade levels of the samples and different scales of self-efficacy used, but an empirical test of these assumptions is still necessary.

3.2.5. Intrinsic Motivation

Intrinsic motivation can be understood as a non-drive-based motivational propensity that accounts for play, exploration, and other behaviors that do not require reinforcement for their maintenance (Deci & Ryan, 1985a; Hunt, 1965; Lepper & Henderlong, 2000). Academic intrinsic motivation refers to pleasure and interest in learning for its own sake (Gottfried, 1990). As intrinsic motivation shows strong conceptual similarities to the constructs of intrinsic value (Wigfield & Eccles, 2002), interest, and flow (Csikszentmihalyi, 1988; Renninger et al., 1992), we subsume these constructs under the term intrinsic motivation in the following overview and analyses.

Intrinsic motivation is positively related to academic achievement (Jacobs et al., 2002; Krapp, Schiefele, Schreyer, & 1993), and some longitudinal investigations have proposed bidirectional and reciprocal relations of these constructs (e.g., Marsh et al., 2005). Intrinsic motivation and academic self-concept are also positively related across time (Marsh et al., 2005; Spinath & Steinmayr, 2008).

Regarding the development of intrinsic motivation, empirical research shows high rank-order stability in intrinsic motivation throughout adolescence and an increase of this stability with students' age (Gottfried, Fleming, & Gottfried, 2001). In agreement with Eccles and colleagues (1989, 1993), a mean level decline in intrinsic motivation in adolescence can be observed in most longitudinal studies (e.g., Spinath & Spinath, 2005; Wigfield et al., 1997). Of note, Wang and Pomerantz (2009) found a decrease in intrinsic motivation in an

American sample but not among Chinese students. Furthermore, the magnitude of the decline in intrinsic motivation differed by study. No meta-analysis has yet identified the average amount of decline and potential moderators of it.

3.2.6. Achievement Goals

In general, achievement goals comprise future-focused cognitive representations that guide behavior to a competence-related end state (Dweck, 1986; Elliot & Thrash, 2002; Hulleman, Schrager, Bodmann, & Harackiewicz, 2010). The end state may include either the development of competence (i.e., mastery) or the demonstration of it (i.e., performance). Furthermore, achievement goals can be classified in approaching positive outcomes or avoiding negative ones (Elliot & McGregor, 2001; Pintrich, 2000).

This conceptualization results in a framework of four possible kinds of goals (i.e., mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance). Mastery-approach oriented individuals want to develop competence, whereas the mastery-avoidance oriented individuals try to avoid incompetence. Performance-approach achievement goals comprise the goal to perform better than others do, whereas performance-avoidance achievement goals comprise the goal to avoid doing worse than others do. Of note, the differentiation of approach and avoidance of the mastery achievement goals is a rather new field of research (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010) that has rarely been applied in longitudinal research. Therefore, in the present meta-analysis achievement goals are classified according to three facets: mastery achievement goals (mastery), performance-approach achievement goals (PAP), and performance-avoidance achievement goals (PAV). In meta-analytic research, PAP and PAV showed medium-sized positive correlations, whereas mastery was only weakly positively related to PAP and weakly negatively related to PAV (Payne, Youngcourt, & Beaubien, 2007). In a meta-analysis including school-aged children and ungraduated students, Hulleman, Schrager, Bodmann, and Harackiewicz (2010) found different relationships between mastery, PAV, and PAP with achievement and intrinsic motivation (positive for mastery approach, mixed for PAP, and negative for PAV). That is, mastery, PAP, and PAV showed different relations to the relevant school-related constructs.

In most longitudinal studies, achievement goals have shown moderate to high rank-order stabilities (e.g., Anderman & Anderman, 1999; Anderman & Midgely, 1997; Bong, 2005; Meece & Miller, 2001; Niepel, Brunner, & Preckel, 2014b). Some longitudinal research indicated a decrease in mastery but no change in PAP, and an increase in PAV through the school career (e.g., Anderman & Midgley, 1997; Bouffard, Boileau, & Vezeau, 2001; Urdan

& Midgley, 2003). Other researchers also observed a decrease in PAP and PAV (e.g., Shim, Ryan, & Anderson, 2008). Thus, overall the results of achievement goals mean-level changes are heterogeneous.

3.3. The Present Study

Based on an extensive literature review, we identified the need for a quantitative meta-analysis of the longitudinal research findings on the development of motivational variables and self-esteem during the school career. Considering that Eccles and colleagues (1993) carried out a comprehensive review of the literature available at that time, we constrained the meta-analysis to articles published after 1993. Our research questions were fourfold: (1) What is the exact magnitude of change in motivational variables and self-esteem through students' school career? (2) Is the change comparable over different constructs (i.e., self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals)? (3) Is the change more pronounced at certain stages, i.e., after school transitions, in higher grade levels during adolescence, or in certain school stages like middle school as compared to elementary school or high school? (4) Finally, what are possible further moderators of the change in the motivational variables and self-esteem? In this vein, we investigated whether the development of motivational variables differed by academic domain (i.e., general, math domain, and verbal domain) as subject-specificity of motivation is an important topic in recent research (Gogol, Brunner, Martin, Preckel, & Goetz, 2017). In addition, we carried out moderator-analyses for typical moderator variables in meta-analyses investigating motivation (e.g., Hulleman et al., 2010), including the questionnaire used to assess a construct, duration between measurement points, year of data collection, year of publication, and geographic region in which the study was conducted.

Using meta-analytic methods, we estimated the magnitude of the overall change for the motivational variables and self-esteem. In accordance with Eccles and colleagues (1993), we expected a general decrease through the school career. We further compared the magnitude of the change for the single constructs, that is, self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals. In previous research, mixed findings were reported regarding the motivational development of most of these constructs. Due to the inconsistent findings for self-esteem, we explored the magnitude of self-esteem change as an open research question. For academic self-concept, academic self-efficacy, and intrinsic motivation we expected a decline through students' school career in accordance with Eccles and colleagues (1993) and most recent research findings (Anderman, Maehr, & Midgley, 1999; Dickhäuser et al., 2017; Pajares & Graham, 1999; Shapka, & Keating, 2005; Spinath & Spinath, 2005; Wigfield et al., 1997). Also in accordance with Eccles and colleagues (1993) and based on recent empirical findings (e.g., Anderman & Midgley, 1997; Bouffard, Boileau, & Vezeau, 2001; Urdan & Midgley, 2003), we expected achievement goals to decline in the mastery facet, whereas we expected no change or even an increase in PAP and PAV over time.

We further investigated if the transition to middle school or high school, as well as the grade level and school stage (i.e., elementary school, middle school or high school) moderated the change in motivational variables and self-esteem. Based on SEF theory, we expected that the middle-school transition and the middle-school stage would be associated with a particularly strong decline in motivational variables and self-esteem. With respect to the grade level of students, no predictions were made. Finally, we investigated whether the magnitude of change varied by academic domain (i.e., general, math domain, verbal domain), questionnaire, duration between measurement points, year of data collection, year of publication, or geographic location (i.e., Asia, Europe, and North America).

3.4. Method

3.4.1. Identification of Relevant Articles and Classification of Scales

Figure 3 summarizes the detailed search process for articles containing longitudinal studies as well as the classification process of the scales used in these studies. Overall, 529 different articles were located by four search strategies (i.e., PsycINFO; articles that cited Eccles et al., 1993; reference lists of relevant reviews and meta-analyses; reference lists of articles found by the other search strategies and exploratory search). The first author scanned these articles and checked the studies for inclusion and exclusion criteria. The inclusion and exclusion criteria were: (a) Scales of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, or achievement goals were used in a longitudinal study (i.e., in at least two points of measurement). (b) There was no experimental manipulation or intervention relating to the constructs. (c) The participants were school students. (d) Measurement points could be assigned to a particular grade level. (e) The article was published after 1993. (f) The article was published in English in a peer-reviewed journal. In total, 425 articles were excluded due to the set criteria (see Table A1 in Appendix A for examples of excluded articles) leaving 104 articles (see Figure 3 for additional information). A second rater validated these ratings in a randomly assigned subsample of 50 articles. Both raters excluded the same 45 studies of this subsample.

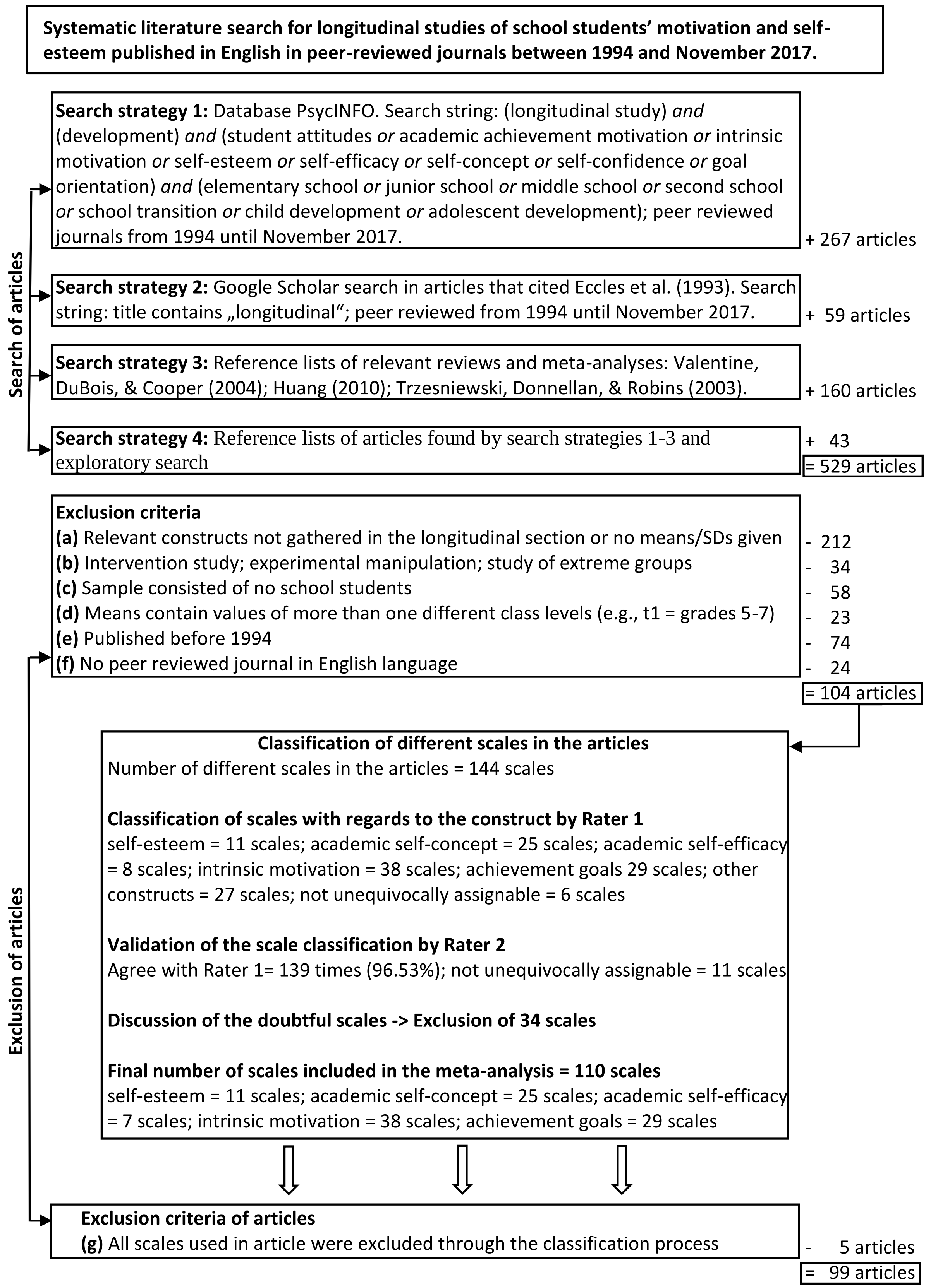


Figure 3. Chronological search process of the articles and classification process of the scales.

In the 104 articles, 144 different scales were applied to measure motivational constructs. Any single article could contain several different scales and a number of articles used identical scales (e.g., the self-esteem scale of Rosenberg was utilized in 19 different articles). Scale name, original author(s) of the scale, and example items and/or scale descriptions of the 144 scales were transferred into a table. The first author went through this table scale by scale and assigned each scale to the constructs self-esteem (11 scales), academic self-concept (25 scales), academic self-efficacy (8 scales), intrinsic motivation (37 scales), achievement goals (29 scales), “other construct” (27 scales) or “unknown” (6 scales). His ratings were added into the table. The second author checked this table scale by scale, agreed with 139 ratings (96.53%), and was unsure about 5 further scales. Both authors discussed the 11 questionable scales (i.e., unknown and unsure) and decided together to exclude 8 of these scales from further analyses and to assign 3 scales to the motivational construct categories. 26 scales that were assigned to “other constructs” by both raters were excluded from further analyses. Five articles included only scales that were discarded through this classification process. Therefore, these 5 articles were excluded from further analyses (exclusion criteria g). For examples of the assigned scales for each construct, see Table A2 in Appendix A.

At the conclusion of the classification process, 99 different articles with 110 scales remained for the meta-analysis. These articles contained 107 independent studies as six articles each reported several different studies with independent samples. The 110 scales were assorted by questionnaire (e.g., questionnaire of Harter, 1985). Overall, 52 different questionnaires were identified, and each questionnaire could contain several scales. For academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals, we classified scales for academic domain (i.e., general domain, math domain, or language domain). We classified the achievement goals scales into mastery, PAP, or PAV categories.

Publication bias. Meta-analytic methods have been criticized for the so-called publication bias because of their inclusion of a non-representative proportion of significant studies indicating positive directions (Duval & Tweedie, 2000; Egger, Smith, Schneider, & Minder, 1997). Publication bias occurs when the likelihood to publish an article depends on the outcome of an experiment or research study (i.e., significant results in the predicted direction). This is particularly problematic in meta-analyses that only consider studies from the published scientific literature (Duval & Tweedie, 2000).

On the other hand, unpublished literature is hard to identify by common search strategies, difficult to obtain, and especially heterogeneous, which makes it less amenable to traditional forms of archiving, analysis, synthesis, data extraction, and integration (Adams, Smart, & Huff, 2016). Furthermore, authors of meta-analyses who include unpublished studies tend to obtain a non-representative set of unpublished studies and to over-represent their own research (Ferguson & Brannick, 2012)

In our meta-analysis, most of the longitudinal studies we considered did not aim to investigate mean level change in motivation but rather cross-lagged relations of motivational variables with other variables (e.g., Heo & Kim, 2015; Lazarides & Rubach, 2017; Madjar & Chohat, 2016). Therefore, in most cases the likelihood of publishing an article should not depend on a particular or significant direction of the mean level change. For this reason, we focused our search strategy on published studies. To account for publication bias in these studies, we conducted a funnel plot analysis of the 107 independent studies with 912 effect sizes by using the metafor package in r (Viechtbauer, 2010). The funnel plot indicated no publication bias, as we observed a symmetric distribution of the effect sizes (see Figure A1 Appendix A). This result and the aforementioned limitations of unpublished studies suggest that our decision to only include published studies did not lower the quality of the present meta-analysis.

3.4.2. Study Variables and Coding Procedure

The 99 articles were coded by year of data collection (i.e., the year in which the data collection started), year of publication, country, grade level, grade level of transition to middle school, grade level of transition to high school, number of participants, and means and standard deviations (*SD*) in motivational constructs per measurement point. The complete dataset including all moderator variables, can be obtained via email as PDF file from the first author. Frequencies of study variables are depicted in Table 2. Variables were coded and/or computed as follows:

Effect sizes. The effect size of change was calculated by subtracting the mean score of a construct at the earlier measurement point (i.e., T1) from the mean score of the construct at the later measurement point (i.e., T2); the difference was divided by the SD of the earlier measure. This effect size is known as the pretest-posttest raw score effect size (Morris & DeShon, 2002) or as Glass's Δ. A negative effect size represents a decrease and a positive effect size represents an increase in a construct over time.

$$\text{Glass's } \Delta = \left(\frac{M\text{t}2 - M\text{t}1}{SD\text{t}1}\right)$$

In the cases in which a study provided more than two measurement points, the effect sizes were computed for all possible pairs of measurement points. That is, in a study with three measurement points, three effect sizes were calculated: one for the change between measurement point 1 and 2, a second one for the change between measurement point 2 and 3, and a third one for the change between measurement point 1 and 3. Overall, we computed 912 different effect sizes.

Effect size variance was estimated according to the formula proposed by Borenstein, Higgins, and Rothstein (2009, p. 29) using the effect size (*d or* Glass's Δ), the sample size (*n*), and the autocorrelations between the considered points of measurement.

$$v_d = \left(\frac{1}{n} + \frac{d^2}{2n}\right)2(1 - r)$$

Note that the correlations among measurements at different points in time were only available in a subsample of 65 studies. Therefore, we had to estimate plausible *r* coefficients for the studies without information of *r* based on the subsample with reported autocorrelations. Thus, we calculated the average correlations separated by construct and time interval (a < 1 year, b = 1 year, and c > 1 years) and used these average values as estimations for the missing correlations (see Table A3 in the Appendix A for the average correlations).

Percent of maximum possible (POMP) scores. Mean scores of each measurement point were transformed into POMP scores (Cohen et al., 1999). POMP scores represent a percent of the maximum possible score available on the scale and are commonly used to compare scales with different ranges (Cohen et al., 1999). For example, a mean of 3 on a scale from 1 to 4 was transformed into a POMP score of 66.67, as 66.67 percent of the maximum possible score has been reached. In the present meta-analysis, POMP sores were used descriptively to illustrate developmental trajectories of means in Figure 4. However, all statistical analyses (i.e., meta-regressions) were carried out using the standardized effect sizes Glass's Δ.

Construct. Constructs were classified into self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, mastery achievement goals, PAP, and PAV and coded as dichotomous dummy variables (e.g., academic self-concept: 1 = construct is academic self-concept, 0 = construct is not academic self-concept). These variables were used as moderators in meta-regressions of all available effect sizes in the complete dataset. The complete dataset was divided in seven sub-datasets based on the motivational construct. These datasets were used to compute separate meta-regressions for each construct (e.g., meta-regression of self-esteem) and further moderator analyses (see Table 2 for an overview of the separate datasets).

Table 2. *Frequency of Study Variables Separated by Construct*

Variable	Dataset							
	Overall	Self-esteem	Self-concept	Self-efficacy	Int. Mot.	Mastery	PAP	PAV
Effect sizes (*h*), studies (*k*) and participants (*n*)								
h/k	912/107	208/39	278/41	41/8	222/34	71/19	58/16	34/9
n in studies	121692	50357	35955	4179	32166	19281	18081	14858
Frequency Δ < 0	596	70	200	17	178	60	49	22
Frequency Δ >= 0	316	138	78	24	44	11	9	12
Minimal value of Δ	-1.26	-1.24	-1.26	-.83	-1.10	-.88	-.96	-1.22
Maximal value of Δ	1.16	.61	1.10	1.16	.24	.26	.26	.25
Domains (*h* effect sizes / *k* studies)								
general	329/46	0	121/21	9/4	83/19	56/15	42/11	18/6
math	253/38	0	118/22	14/4	89/16	11/6	12/7	9/4
language	122/24	0	39/13	18/2	50/12	4/4	4/4	7/2
Questionnaires (*h* effect sizes / *k* studies)								
Eccles et al.	68/12		34/7		34/7			
Harter	181/17	89/11	78/11	1/1		7/2	6/1	
Midgley et al.	75/10			9/4		28/9	25/8	13/3
Marsh	58/8	2/2	50/8		6/1			
Rosenberg	99/19	99/19						
Spinath	144/5		67/3		67/3	4/2	3/1	3/1
Other	287/36	18/7	49/12	31/3	115/23	32/6	24/9	18/5
School stage (*h* effect sizes / *k* studies)								
Elementary	279/27	24/8	120/16	2/2	97/11	17/3	18/4	1/1
Middle-school	367/58	87/20	127/21	9/5	36/12	46/15	32/11	30/7
High-School	199/34	58/17	24/7	18/1	84/15	6/2	6/2	3/1
Missing Information	67/20	39/14	7/5	12/1	5/4	2/1	2/1	0
School transition (*h* effect sizes / *k* studies)								
Transition	260/75	58/19	65/15	29/6	59/12	22/11	21/10	6/4
No transition	552/52	90/24	206/29	4/1	154/25	43/12	31/10	24/8
Missing information	100/25	60/16	7/6	8/1	9/5	6/2	6/2	4/1
Geographic location (*h* effect sizes / *k* studies)								
Asia	65/6	6/1	3/1	18/1	23/2	10/3	4/2	1/1
Europe	344/38	46/10	147/17	0	118/19	7/3	7/3	19/4
North America	481/59	154/26	109/19	23/7	80/12	54/12	47/11	14/4
Other	22/5	2/2	19/4	0	1/1	0	0	0
Start year of data collection								
Min	1966	1966	1979	1994	1979	1993	1993	2005
Max	2014	2006	2009	1995	2014	2011	2007	2010
Year of publication								
Min	1994	1994	1994	1995	2001	1997	1997	2005
Max	2017	2017	2017	2010	2017	2015	2015	2010
Average	2008	2005	2008	2003	2010	2007	2006	2011

Note. Overall = complete dataset. Self-concept = Academic self-concept. Self-efficacy = Academic self-efficacy. Int. mot. = Intrinsic motivation. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals

Transition to middle school or high school. Transition to middle school or high school was coded as one dichotomous dummy variable (1 = transition to middle school or high school took place between the two measurement points, 0 = both measurement points took place at the same school stage). Overall, there were 260 effect sizes for the change between two measurement points with a transition. For 100 effect sizes, an unambiguous assignment to the transition variable was not possible as no information about school transition was available.

Grade level. Grade level of students was gathered from the articles and coded at 3-month intervals. For example, a scale completed at the beginning of grade 3 was coded as 3.00 and a scale completed halfway through grade 7 was coded as 7.5. Although information from two grade levels were available for each computed effect size (i.e., grade level at measurement point 1 and the grade level at measurement point 2), we only considered the grade level at measurement point 2 in further analyses. That is, when means were collected at Grade 6 (measurement point 1) and Grade 8 (measurement point 2), the computed effect size was coded as Grade 8. Measurement point 1 was not included in further analyses because the duration of intervals between two points of measurement was computed for each effect size in number of years as one further variable. Thus, no information was lost.

School stage. The 912 effect sizes were coded as either the elementary school stage, the middle school stage, or the high school stage. We computed three dichotomous dummy variables representing these stages based on the school stage of the later measurement point (e.g., elementary school stage: 1 = T2 took place in elementary school, 2 = T2 took place not in elementary school). An unambiguous assignment to one of the school stages was not possible for 67 effect sizes as no information about the school type was available.

Year of data collection and year of publication. In 52 studies, the year in which data collection started was available. This information was coded as a continuous variable. Year of publication was coded as a further continuous variable.

Domain. Several domains (i.e., math domain, language domain, or general domain) could be distinguished in all constructs besides self-esteem. These domains were obtained from the articles and coded as separate dichotomous dummy variables (e.g., math domain: 1 = domain is math, 0 = domain is not math).

Questionnaire. Overall, 51 different questionnaires were utilized in the considered articles. Each questionnaire could contain several scales of different constructs and in some studies several different questionnaires were used. Only the questionnaires of Eccles and colleagues (1983), Harter (1985), Midgley and colleagues (2000), Marsh and colleagues (1983), Spinath and Spinath (2005), and Rosenberg (1989) were used in four or more independent studies (for frequencies see Table 2). We coded each of these six questionnaires as separate dichotomous dummy variables (e.g., Rosenberg, 1989: 1 = questionnaire is

Rosenberg, 1989; 0 = questionnaire is not Rosenberg, 1989). All other questionnaires were subsumed into one further dichotomous variable “other questionnaire”.

Geographic location. Studies were carried out in 17 different countries. Due to their geographic location, we subsumed these countries into the four separate dichotomous variables: North America (United States and Canada), Europe (Belgium, Croatia, Estonia, England, Finland, Germany, Italy, Netherlands, and Norway), Asia (China, Korea, and Thailand), and other location (Australia, Israel, and New Zealand).

3.4.3. Data Analysis

Our main research aim was to summarize findings from longitudinal studies reporting changes in motivational variables and self-esteem by a meta-analysis and to quantify the magnitude of change. We selected 107 independent studies for this analysis, but many contained several different constructs and domains or scales. Furthermore, multiple studies collected data at more than two measurement points and hence contained more than one effect size based on the same participants. This led to a complex data structure with 912 partly dependent effect sizes within 107 independent studies. Common strategies to deal with dependency (e.g., averaging dependent effect sizes within studies or selecting one effect size per study) are easy to implement. However, they reduce the statistical power of the analysis (Hedges & Pigott, 2001, 2004), lead to a loss of information, and obscure differences within studies (Cheung, 2014; Hedges, Tipon, & Johnson, 2010; see Bornstein et al., 2009, p. 217-223 for a review of common methods). Therefore, we used the recently introduced Robust Variance Estimation (RVE) method that controls for the dependence of effect sizes within studies without the loss of power and information that results from the common approaches (Hedges, Tipton, & Johnson, 2010; Tipton, 2013). To conduct the RVE meta-analysis, we followed the guidance of Tanner-Smith, Tipton, and Polanin (2016) and used the *robumeta* package in r (Fisher & Tipton, 2016).

Magnitude of change. To estimate the magnitude of change in motivational variables and self-esteem, we calculated a random effect RVE meta-regression for the full dataset of the 912 effect sizes. The calculated overall effect size Glass’s Δ represents the average change between all possible pairs of measurement points in the selected studies. When determining the average time duration of the overall effect size we took into account the weights of effect sizes used in the RVE regression. That is, the duration of each particular effect size was included in the mean duration calculation proportional to its weight. In RVE meta-regression weights are calculated as follows:

$$w_{ij} = \frac{1}{k_j(v_{\cdot j} + \tau^2)}$$

Where $v_{.j}$ is the average sampling variance in study j, τ^2 is the estimate of the between-study variance, and kj is the number of effect sizes within each study j (Hedges, Tipton, & Johnson, 2010; Tanner-Smith, Tipton, & Polanin, 2016). For exact weights of all effect sizes in the considered datasets see Figure A2 in the Appendix A.

Moderator Analyses. We conducted several moderator analyses to examine whether different moderator variables were associated with different changes. Each moderator analysis was conducted independently of the other moderator analyses. If we found a significant categorical moderator, we calculated separate RVE regressions for the separate categories to estimate the particular magnitude of change in each category. Of note, separate meta-regressions were calculated only for those categories that included at least four independent studies as a minimum of four *df* is recommended for RVE meta-regressions (Fisher & Tipton, 2016).

Change in different constructs. To test whether the change was comparable over different constructs we included the dichotomous dummy variables academic self-concept, academic self-efficacy, intrinsic motivation, mastery, PAP, and PAV into the meta-regression of the complete dataset. The self-esteem category was set as the reference category. All further moderator analyses were conducted for the complete dataset as well for the separate datasets of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, mastery, PAP, and PAV.

Change after transition and at different grades or school stages. To test whether transition to middle school and transition to high school were associated with motivational development the dichotomous dummy variable “school transition” with the reference category “no transition” was included in the analyses. The dataset of academic self-efficacy was excluded from these analyses because of *df* values lower than 4.

To examine whether the change of motivational variables and self-esteem was more pronounced at certain developmental stages and whether it was influenced by environmental changes we included the continuous variable grade level (T2) into the meta-regressions. The separate datasets of academic self-efficacy and PAP were excluded from these analyses because of *df* values lower than 4.

Next, we tested whether the change varied by school stage. Therefore, we included the dichotomous dummy variables elementary school and high school into the meta-regressions, with middle school set as the reference category. Of note, effect sizes without unambiguous assignment to one of the school stages were excluded from these analyses. Furthermore, we excluded the separate datasets of academic self-efficacy, mastery, PAP, and PAV due to *df* values lower than 4.

Further moderators: Domain, questionnaire, duration of intervals, year of data collection/publication, and geographic location. Next, we examined whether specific domains were associated with a different magnitude of change. The dichotomous dummy variables "math domain" and "language domain" were included in the meta-regression, whereas the general domain was set as reference category. The dataset of self-esteem was excluded from these analyses, as articles in this dataset did not differentiate between specific domains. We did not consider the RVE regressions in the separate dataset of academic self-efficacy because the *df* values were less than 4.

Likewise, we tested whether the degree of change differed by questionnaire. Therefore, we included several dichotomous dummy variables representing individual questionnaires in the meta-regressions. The most frequently used questionnaire was set as reference category. Of note, these analyses were not applied in the separate dataset of academic self-efficacy because of a *df* value lower than 4. Because most questionnaires assess a particular construct, this moderator analysis was not conducted for the complete dataset.

We tested whether the duration of the interval between measurement points showed a relation to the effect sizes by including this continuous variable into the meta-regressions of the full dataset and of the separate datasets. Because of *df* values lower than 4, the separate datasets of academic self-efficacy, PAP, and PAV were not considered in these analyses.

Finally, in three separate analyses we tested whether the continuous variables year of data collection and year of publication were significant moderators of the change and whether the change differed by geographic location. In the analyses of year of data collection and the analyses of year of publication, we did not consider the separate datasets of academic self-efficacy and PAV due to *df* values lower than 4. In the analyses of geographic location, Europe was set as the reference category. We did not consider the separate datasets of academic self-efficacy, PAP and PAV in these analyses because of *df* values lower than 4.

3.5. Results

3.5.1 Magnitude of Change.

The RVE meta-regression analysis of the complete dataset including $k = 107$ independent studies with $h = 912$ effect sizes of change indicated a small but significant overall decrease in the considered constructs ($\Delta = -.108$, $p < .001$, 95% CI [-.154, -.063]). The average interval duration was 1.654 years. That is, on average the mean level of the constructs decreased by .108 *SD* over 1.654 school years in the 107 studies.

3.5.2. Change in Different Motivational Constructs.

Table 3 provides the detailed results of the moderator analyses of each construct. In the complete dataset, the change differed significantly by construct ($p < .01$). More specifically, the dichotomous moderator variables of academic self-concept ($b = -.152$, $df = 62.41$, $p < .05$), intrinsic motivation ($b = -.214$, $df = 58.24$, $p < .001$), and mastery achievement goals ($b = -.206$, $df = 20.90$, $p < .01$) indicated a significantly stronger decrease than the reference category self-esteem.

RVE meta-regression analyses in datasets separated by construct revealed a significant decrease in academic self-concept ($\Delta = -.122$, $p < .001$, 95% CI [-.187, -.058]), intrinsic motivation ($\Delta = -.190$, $p < .001$, 95% CI [-.253, -.127]), mastery ($\Delta = -.196$, $p < .001$, 95% CI [-.280, -.111]), and PAP ($\Delta = -.144$, $p < .05$, 95% CI [-.250, -.039]). The average durations of intervals in academic self-concept, intrinsic motivation, mastery, and PAP were 1.718, 1.656, 1.103, and 1.094 years. However, the RVE meta-regressions revealed no significant change ($p > .05$) in the datasets of self-esteem, academic self-efficacy, and PAV. For more information on the overall change in the separate datasets, see Table 4.

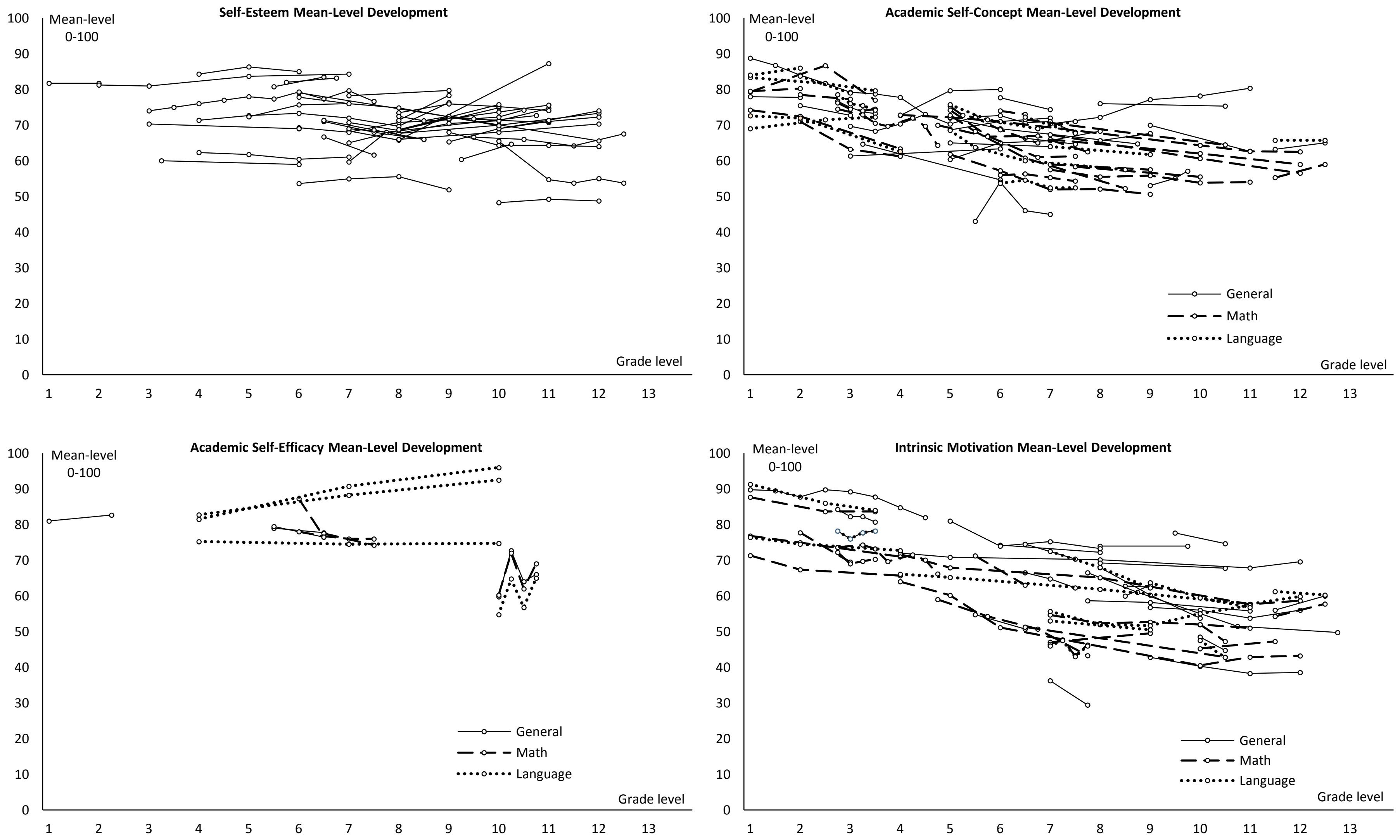
To visualize the observed differences in the development of the motivational variables and self-esteem, all means were transformed into POMP scores and depicted in Figure 4. Descriptively, the mean levels of academic self-concept, intrinsic motivation, mastery, and PAP seemed to decrease, whereas the mean levels of self-esteem remained rather stable. For academic self-efficacy and PAV, contradictory developmental trends could be noted in different studies. This picture was in agreement with the results of the changes that were observed for the different motivational constructs in the meta-regressions. For a detailed overview of all effect sizes and studies used in the meta-regression analysis, see also the forest plots shown in Figure A2 in the Appendix A.

Table 3. Moderator Analyses of Motivational Development in Different Datasets

Moderator analysis	Dataset	Predictor	Reference category	*b*	*df*	*p*	95% Confidence interval	I^2	τ^2
Construct									
	Overall	Self-concept	Self-esteem	-.15	62.41	.01	-.27, -.03	98.34	.06
		Self-efficacy	Self-esteem	-.02	7.00	.90	-.36, .32		
		Int. mot.	Self-esteem	-.21	58.24	<.01	-.34, -.09		
		Mastery	Self-esteem	-.21	20.90	<.01	-.33, -.08		
		PAP	Self-esteem	-.15	18.09	.07	-.31, .02		
		PAV	Self-esteem	-.19	8.84	.15	-.46, .09		
School transition									
	Overall	Transition	No transit.	-.02	78.60	.68	-.14, .09	98.70	.07
	Self-esteem	Transition	No transit.	-.02	29.60	.86	-.24, .20	98.80	.08
	Self-concept	Transition	No transit.	.02	23.50	.82	-.14, .17	97.62	.05
	Int. mot.	Transition	No transit.	-.14	18.90	.09	-.31, .03	97.90	.06
	Mastery	Transition	No transit.	-.08	16.36	.29	-.23, .07	96.74	.04
	PAP	Transition	No transit.	.09	13.42	.33	-.10, .28	97.82	.06
	PAV	Transition	No transit.	-.11	4.24	.75	-.1.01, .78	98.81	.10
Grade level									
	Overall	Grade level		.01	45.00	.47	-.01, .03	98.78	.07
	Self-esteem	Grade level		.00	14.70	.90	-.04, .04	98.53	.06
	Self-concept	Grade level		.00	19.70	.97	-.03, .02	97.40	.05
	Int. mot.	Grade level		.00	15.70	.70	-.02, .03	97.72	.05
	Mastery	Grade level		.06	5.13	.02	.02, .10	95.24	.03
	PAP	Grade level		.03	4.22	.20	-.02, .07	97.60	.06
School type									
	Overall	Elementary	Mid-school	.03	37.20	.56	-.07, .13	98.14	.06
		High-school	Mid-school	.02	63.10	.71	-.10, .14		
	Self-esteem	Elementary	Mid-school	.08	8.94	.19	-.05, .21	97.81	.05
		High-school	Mid-school	.02	28.65	.88	-.20, .23		
	Self-concept	Elementary	Mid-school	-.03	25.93	.70	-.18, .12	97.57	.05
		High-school	Mid-school	-.08	9.19	.47	-.34, .17		
	Int. mot.	Elementary	Mid-school	.05	17.4	.58	-.12, .21	97.84	.06
		High-school	Mid-school	.02	18.0	.80	-.16, .20		
Academic domain									
	Overall	Math	General	-.06	54.80	.16	-.14, .02	97.79	.05
		Language	General	.00	33.10	.95	-.14, .11		
	Self-concept	Math	General	-.19	33.70	<.01	-.31, -.06	96.95	.04
		Language	General	-.15	18.20	.10	-.33, .30		
	Int. mot.	Math	General	.02	24.80	.76	-.11, .15	97.69	.05
		Language	General	-.03	15.90	.74	-.18, .13		
	Mastery	Math	General	.11	6.18	.16	-.06, .27	97.00	.04
		Language	General		<4.00				
	Pap	Math	General	.02	9.93	.84	-.22, .27	97.92	.06
		Language	General		<4.00				
	Pav	Math	General	.05	4.88	.82	-.52, .63	99.27	.17
		Language	General		<4.00				
Questionnaire									
	Self-esteem	Harter	Rosenberg	-.04	23.13	.47	-.16, .08	98.52	.06
		Marsh			<4.00				
		Other quest.	Rosenberg	-.35	10.98	.08	-.75, .06		
	Self-concept	Eccles	Harter	-.21	10.94	.04	-.42, -.01	97.30	.05
		Marsh	Harter	-.07	15.53	.48	-.28, .14		
		Other quest.	Harter	-.14	20.55	.14	-.32, .05		
		Spinath	Harter		<4.00				
	Int. mot.	Marsh	Eccles		<4.00			97.75	.06
		Other quest.	Eccles	-.07	9.20	.99	-.17, .17		
		Spinath	Eccles		<4.00				

Moderat- or analysis	Dataset	Predictor	Reference category	*b*	*df*	*p*	95% Confidence interval	I^2	τ^2
	Mastery	Harter	Midgley		<4.00			92.21	.04
		Other quest.	Midgley	.11	10.86	.30	-.11, .33		
		Spinath	Midgley		<4.00				
	PAP	Harter	Midgley		<4.00			96.09	.07
		Other quest.	Midgley	-.18	10.90	.13	-.41, .06		
		Spinath	Midgley		<4.00				
	PAV	Other quest.	Midgley	.25	4.33	.26	-.27, .76	98.87	.19
		Spinath	Midgley		<4.00				
Duration of interval									
	Overall	Duration		-.05	23.4	.12	-.12, .02	98.62	.07
	Self-esteem	Duration		-.07	8.0	.44	-.29, .14	98.58	.06
	Self-concept	Duration		-.08	10.7	<.01	-.12, -.04	96.82	.04
	Int. mot.	Duration		-.07	7.41	.01	-.12, -.03	97.28	.04
	Mastery	Duration		-.16	6.95	.04	-.31, -.01	96.88	.04
Year of collection									
	Overall	Year of coll.		.00	17.7	.76	-.01, .01	98.98	.07
	Self-esteem	Year of coll.		.00	6.65	.69	-.02, .02	98.99	.07
	Self-concept	Year of coll.		.00	5.18	.93	-.01, .02	96.19	.03
	Int. mot.	Year of coll.		.01	6.33	.07	.00, .02	98.24	.06
	Mastery	Year of coll.		.01	4.17	.50	-.02, .03	98.49	.05
	PAP	Year of coll.		-.04	4.36	.02	-.05, -.01	98.35	.05
Year of publication									
	Overall	Year of pub.		.00	55.30	.94	-.01, .01	98.63	.07
	Self-esteem	Year of pub.		.01	22.20	.08	.00, .02	98.45	.06
	Self-concept	Year of pub.		.00	21.80	.57	-.01, .01	97.39	.04
	Int. mot.	Year of pub.		.01	20.70	.02	.00, .02	97.68	.05
	Mastery	Year of pub.		.01	8.50	.33	-.01, .02	96.82	.04
	PAP	Year of pub.		-.02	8.60	.01	-.03, -.01	97.69	.06
Geographic location									
	Overall	Asia	Europe	.21	6.80	.03	.03, .39	98.65	.07
		N. America	Europe	.11	78.60	.05	.00, .22		
		Other	Europe	.16	5.00	.07	-.02, .34		
	Self-esteem	Asia	Europe	-.03	8.90	.83	-.35, .28		
		N. America	Europe	.05	16.00	.74	-.26, .35		
		Other	Europe		<4.00				
	Self-concept	Asia	Europe		<4.00			96.97	.04
		N. America	Europe	.08	34.20	.24	-.05, .21		
		Other	Europe		<4.00				
	Int. mot.	Asia	Europe		<4.00			97.73	.05
		N. America	Europe	.00	22.60	.95	-.13, .13		
		Other	Europe		<4.00				
	Mastery	Asia	Europe	.19	4.00	.37	-.32, .71	96.55	.05
		N. America	Europe		<4.00				
		Other	Europe		<4.00				

Note. Overall = complete dataset. Self-concept = Academic self-concept. Self-efficacy = Academic self-efficacy. Int. mot. = Intrinsic motivation. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals. Analyses were conducted in the complete dataset and in datasets separated by construct if possible. Results are not reported if *df* < 4.00. A reference category is reported if the moderator variable was coded as dichotomous dummy variable.


Self-Esteem Mean-Level Development
Mean-level 0-100
Grade level
Academic Self-Concept Mean-Level Development
Mean-level 0-100
Grade level
General
Math
Language
Academic Self-Efficacy Mean-Level Development
Mean-level 0-100
Grade level
General
Math
Language
Intrinsic Motivation Mean-Level Development
Mean-level 0-100
Grade level
General
Math
Language

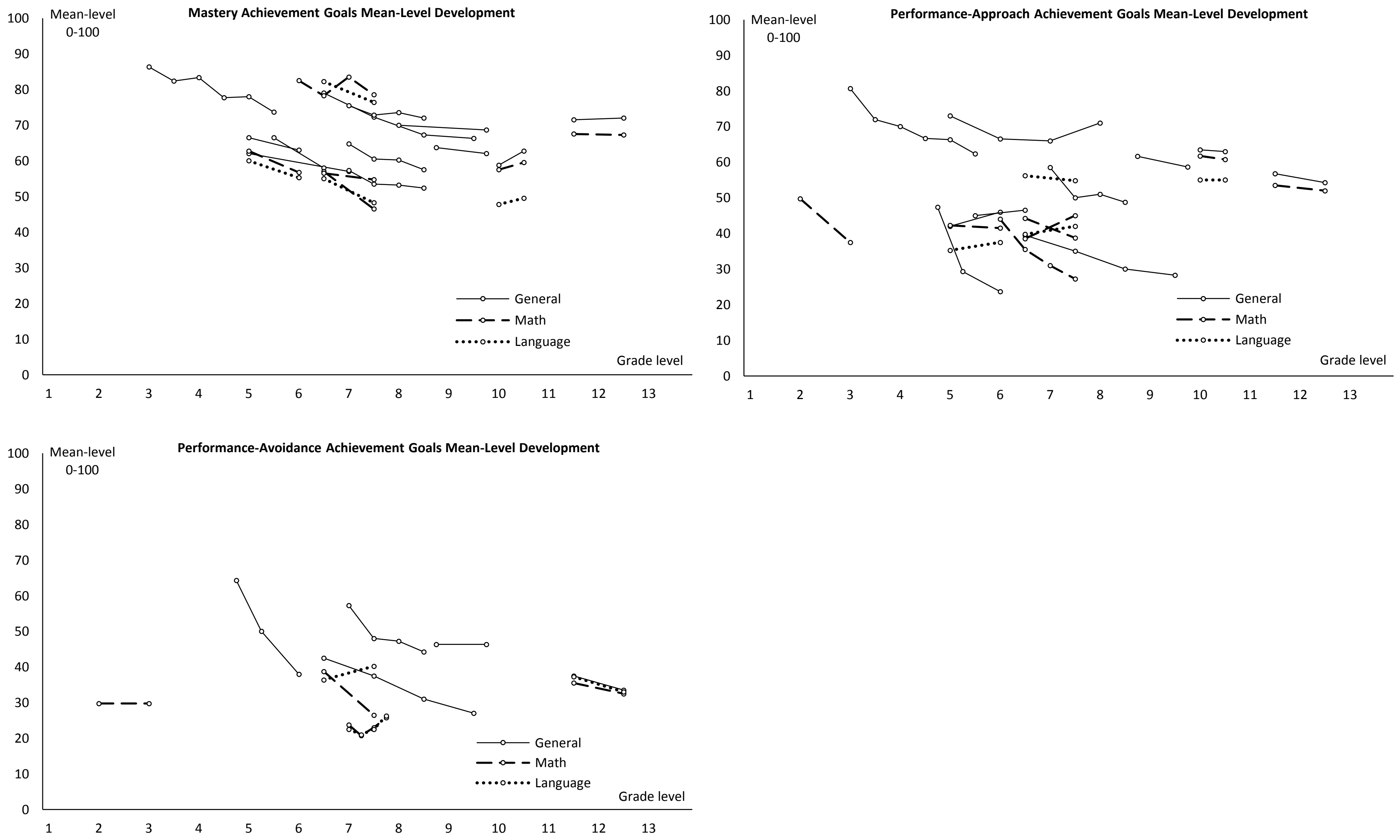


Figure 4. All reported means of the considered studies separated by motivational construct and transformed into POMP scores (0-100). Points represent collected measurement points in longitudinal studies. Horizontal line depict time of collection in school grades.

3.5.3. Change in Different Stages.

The transition to middle school or high school was not related to a distinct change as in all datasets the dichotomous variable transition was not a significant moderator of the development ($p > .05$; for exact numbers see Table 4). In the dataset examining mastery, grade level of students was a significant moderator ($b = .056$, $df = 5.13$, $p < .05$) of the developmental change. That is, with rising grade level the decrease in mastery became smaller. In the datasets of all further constructs and in the complete dataset, the change was not significantly related to grade level ($p > .05$; for exact numbers see Table 3). The particular school stage was not significantly associated with change; the dichotomous variables elementary school and high school were not significant predictors of the meta-regressions in the complete dataset or in the datasets separated by construct ($p > .05$; for exact numbers see Table 3).

Further moderators. Moderator analyses of the specific domain (i.e., general, math, verbal) in the complete dataset (without self-esteem) and in the separate datasets of intrinsic motivation, mastery, PAP, and PAV indicated no significant differences by domain. That is, for these constructs the two dichotomous moderator variables *math domain* and *language domain* did not differ ($p > .05$; for exact numbers see Table 3) from the reference category *general domain*. In the dataset of academic self-concept, however, we found a significantly larger decrease in the math domain ($b = -.173$, $df = 32.7$, $p < .01$) and in the language domain ($b = -.178$, $df = 20.3$, $p < .01$) in comparison to the general domain. Conducting separate RVE meta-regression for the academic self-concept domains, we found a significant decrease in the academic self-concept math domain ($\Delta = -.181$, $p < .001$, 95% CI [-.249, -.113]; average interval duration = 1.651 years) and the academic self-concept language domain ($\Delta = -.149$, $p < .05$, 95% CI [-.244, -.055]; average interval duration = 1.200 years). General academic self-concept showed no significant change ($\Delta = -.022$, $p = .659$, 95% CI [-.125, .081]; average interval duration = 1.457 years).

Moderator analyses with specific questionnaires as dichotomous predictors revealed no significant difference in the datasets of self-esteem, intrinsic motivation, mastery, PAP, and PAV ($p > .05$; for exact numbers see Table 3). In the dataset of academic self-concept, we observed a significantly bigger decrease in studies that used the scales provided by Eccles and colleagues (1983) compared to studies that used the questionnaire of Harter (1985) ($\Delta = -.213$, $p < .05$, 95% CI [-.416, -.009]).

The duration of interval was negatively related to the change in the datasets examining academic self-concept ($b = -.080$, $df = 10.7$, $p < .001$), intrinsic motivation ($b = -.073$, $df =$

7.41, $p < .01$), and mastery achievement goals ($b = -.158$, $df = 6.95$, $p < .05$). Thus, effect sizes based on longer intervals between measurement points indicated a bigger decrease than effect sizes based on shorter intervals. In the complete dataset and the dataset examining self-esteem, duration of interval was not a significant moderator ($p > .05$; for exact numbers, see Table 3).

The year of data collection was not a significant moderator of the change in all datasets ($p > .05$; for exact numbers see Table 3), except for the dataset of PAP where earlier publications reported a smaller decrease than later studies ($b = -.037$, $df = 4.36$, $p = .019$). The year of publication was a significant moderator only in the dataset examining intrinsic motivation ($b = .012$, $df = 20.7$, $p < .05$) and PAP ($b = -.020$, $df = 8.6$, $p < .05$). That is, older publications reported a larger decrease in intrinsic motivation, whereas a larger decrease of PAP was found in newer publications.

Table 4. *Overall Change Separated by Motivational Construct, Domain, and Geographic Location*

Dataset	*k* studies	*h* effect sizes	Δ	*df*	*p*	95% confidence interval	I^2	τ^2	Duration of intervals
Overall	107	912	-.11	106.00	< .01	-.15, -.06	98.62	.07	1.65
Self-esteem	39	208	.01	37.80	.74	-.07, .10	98.61	.06	1.83
Self-concept	41	278	-.12	39.40	< .01	-.19, -.06	97.36	.04	1.72
Self-efficacy	8	41	-.09	7.00	.48	-.38, .20	98.03	.11	1.22
Int. motivation	34	222	-.19	32.80	< .01	-.25, -.13	97.68	.05	1.66
Mastery	19	71	-.20	17.90	<. 01	-.28, -.11	96.84	.04	1.10
PAP	16	58	-.14	15.00	.01	-.25, -.04	97.77	.05	1.09
PAV	9	34	-.19	8.00	.11	-.43, .05	99.15	.13	.97
Self-concept general	20	121	-.02	18.80	.66	-.13, .08	96.79	.07	1.46
Self-concept math	22	118	-.18	10.90	< .01	-.25, -.11	97.50	.03	1.65
Self-concept math	13	39	-.15	11.60	< .01	-.24, -.06	91.60	.02	1.20
Europe	38	344	-.19	36.90	< .01	-.28, -.09	99.21	.12	1.75
North America	58	481	-.08	56.80	< .01	-.13, -.03	97.82	.05	1.65
Asia	6	65	.03	4.99	.69	-.13, 18	98.57	.03	1.10

Note. Self-concept = Academic self-concept. Self-efficacy = Academic self-efficacy. Int. motivation = Intrinsic motivation. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals. The duration of intervals is reported in years.

In the complete dataset the development differed by geographic location, as studies in Asia ($b = .212$, $df = 6.8$, $p < .05$) and North America ($b = .108$, $df = 78.6$, $p < .001$) reported a smaller decrease than studies in the reference category Europe. For the datasets that included all motivational constructs but that were separated by the geographic location, RVE meta-regression analyses revealed a significant decrease in motivation in Europe (Δ = -.189, $p <$ .001, 95% CI [-.283, -.094], average interval duration = 1.750 years) and North America (Δ =

-.079, $p < .01$, 95% CI [-.133, -0.025], average interval duration = 1.645 years), whereas studies from Asia indicated no significant change of motivation ($\Delta = .022$, $p = .713$, 95% CL [-.126, .176], average interval duration = 1.100 years) (for more information, see Table 4).

3.6. Discussion

In the current meta-analysis, we investigated the development of motivational variables and self-esteem in school-aged children to answer the following main research questions: What is the exact magnitude of the change through students' school career? Is the change comparable over different constructs? Is the change more pronounced at certain stages, i.e., after school transitions, in higher grades or in middle school, respectively? In addition, further moderator analyses investigated a possible impact of academic domain, questionnaire used to assess a construct, duration of interval between measurement points, year of data collection, year of publication, or geographic location. Our meta-analysis included 107 independent longitudinal studies with 912 effect sizes. Results indicated an overall decrease in the considered constructs of .108 *SD* after an average duration of 1.654 years. However, decreases significantly differed by construct. While findings revealed significant decreases for academic self-concept, intrinsic motivation, mastery, and PAP, we found no significant changes in self-esteem, academic self-efficacy, and PAV. In academic self-concept, we only found decreases in the math domain and the language domain but not for the general domain. The decrease in mastery became smaller with increasing grade level. In all constructs, particular school stage (i.e., elementary, middle or high school) or school transitions were not significantly associated with the change. Moreover, the decrease of PAP was greater in more recent studies than in older studies. Finally, the decrease was bigger in Europe than in North America or Asia.

3.6.1. Limitations

Before discussing these findings, we provide some limitations of our meta-analysis. The strengths of the analysis are the large number of 107 independent studies with 912 effect sizes and nine moderator variables that were gathered from the articles. These moderator analyses were conducted for the complete dataset and for the datasets separated by construct as we observed different development trends in self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, mastery, PAP, and PAV. However, there were relatively few independent studies available after categorizing the separate datasets of the different constructs into further categories (i.e., domain, questionnaire, school-stage, school transition, and geographic location). That is, the statistical power of the moderator analyses in the datasets separated by construct was restricted and some moderator analyses in the datasets of academic self-efficacy, mastery, PAP, and PAV goals could not be computed due to the small number of independent studies per moderator category.

The differentiated evaluation of grade level as a continuous moderator of the development is a further strength of our analysis. However, using meta-analytic methods we could not investigate quadratic development trends of motivation that were observed in some investigations (e.g., Greene, Way, & Pahl, 2006; Van de Gaer, Pustjens, Van Damme, & De Munter, 2009).

Although we computed our meta-analyses in separate datasets of several different motivational variables and self-esteem, we did not capture the full variety of possible psychological constructs that were discussed in the context of SEF theory (e.g., utility value, perceived costs, or test anxiety; Cassady & Johnson, 2002; Eccles et al., 1993; Wigfield & Eccles, 2002). However, according to our systematic literature search based on four search strategies (see Figure 3), very few studies investigated these constructs in a longitudinal setting. Therefore, we decided to limit the present analysis to self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, mastery, PAP, and PAV. Further motivational constructs could be added to future meta-analyses.

All studies included in our meta-analysis used self-report questionnaires to assess student motivation and self-esteem. Self-report questionnaires are the most prevalent method for assessing constructs like academic self-concept. In fact, when a person's experiential states and thinking are of interest, asking the person may even be the only choice of assessment method (Baumeister, Vohs, & Funder, 2007). Self-report questionnaires have been criticized for a variety of reasons including common method variance and socially desirable or consistent responding (e.g., Podsakoff & Organ, 1986) and for not being able to depict person-environment interactions or behavior (Ortner & van de Vijver, 2015). Therefore, limitations of self-reports could bias the estimation of the effect sizes of change presented here.

Finally, meta-analytic methods do not allow one to test the theoretical explanations for motivational development. Instead, we quantified the magnitude of the change by conducting meta-regressions and discussed this magnitude afterwards by applying SEF theory. Research designs other than meta-analysis are needed to test specific predictions and implications of available theories.

3.6.2. Magnitude of Change

Considering all constructs together, we found a significant decrease throughout students' school career. This finding is in line with SEF theory (Eccles et al., 1993) which describes and explains a motivational decline in school-aged children. On average, students lost around .108 *SD* in the considered constructs during a period 1.654 years.

3.6.3. Change in Different Motivational Constructs

The magnitude of change differed by construct. While self-esteem, academic self-efficacy, and PAV remained stable over time, academic self-concept decreased by .122 *SD* over 1.718 years, intrinsic motivation by .190 *SD* over 1.656 years, mastery by .196 *SD* over 1.103 years, and PAP by .144 *SD* over 1.041 years. According to SEF theory, the mismatch between the developing needs of students and the opportunities offered to them by their social environment affects the decrease. Interestingly, this mismatch does not seem to affect all constructs in the same way but rather seems to affect intrinsic motivation, mastery, PAP, and domain-specific academic self-concepts (i.e., math and verbal academic self-concept) but not the more general competence beliefs (i.e., general self-concept) or self-esteem.

The observed decreases in academic self-concept, intrinsic motivation, and mastery were in accordance with our expectations based on the previous literature. Self-esteem, however, stayed stable through the whole school career. Furthermore, we expected no change or even an increase in PAP and PAV but found a significant decrease in PAP and no significant change in PAV. This finding does not support the assumptions that the mismatch of students' needs and opportunities results in a shift from mastery to performance orientation. Instead, both mastery and PAP decreased. In PAV, we found no significant change over time although descriptively PAV showed a bigger decrease than PAP (.186 *SD* over a period of .971 years vs. .144 *SD* over a period of 1.041 years). It is important to note that only nine independent longitudinal studies investigated PAV. Thus, the statistical power of the PAV meta-regression was restricted due to the small number of studies.

Surprisingly, we also found no significant general change in academic self-efficacy. One plausible explanation for this finding is the high variability between studies. Different academic self-efficacy scales refer to different specific tasks and efficacy scores are strongly related to specific events such as achievement feedback. Furthermore, only eight independent longitudinal studies were available for the academic self-efficacy construct, resulting in low statistical power for the academic self-efficacy meta-regression.

Although SEF theory has been applied to a variety of psychological constructs, our findings contradict the assumption that the decline is a general phenomenon that can be found in all constructs. As one implication for further research, we suggest adapting general motivational development theories such as SEF theory to specific motivational constructs. For example, based on the findings of this meta-analysis, it is plausible to assume that the mismatch between students' needs and opportunities in school does not affect general self-esteem or general competence beliefs but rather more domain-specific constructs. Further research could investigate which type of mismatch undermines which specific construct.

3.6.4. Change at Different Stages

Of all considered constructs, only mastery development was positively related to grade level. That is, with increasing grade level of students the decrease in mastery became smaller. However, specific school stage (i.e., elementary, middle, and high school) or school transition were not related to the development of mastery and of the further constructs under study. This is surprising as the stage of middle school and the school transition are often discussed as critical events in the literature (e.g., Eccles et al., 1993). Note that our findings do not necessarily indicate that there is no growing mismatch between students' needs and opportunities leading to stronger motivational declines after the school transition as alternative explanations are also plausible. For example, as students become older and more autonomous, less motivation might be associated with less participation in studies or even school dropout. That is, a relation between motivation and study participation (i.e., self-selection bias) might mask motivational declines. Moreover, the number of independent studies per school stage and separate construct was very small in our analyses. Furthermore, we only considered studies that were published after 1993. Therefore, positive administrative changes in middle schools might explain why we did not observe a stronger motivational decrease in middle schools or after the school transition. For example, the results of PISA exercised a world-wide influence on national education systems and played an important role for development efforts in individual schools and broader school systems (Beycioglu, & Kondakci, 2014; Grek, 2009).

Yet our findings clearly contradict the assumption of adolescence as a critical stage for students' motivation as the observed decrease in the motivational variables and self-esteem was not specific to this or another stage. In accordance with our findings, most researchers that overviewed the motivational development in adolescence in the last decades did not consider adolescence as a necessarily turbulent time (Wigfield & Wagner, 2005, p. 223) and did not find adolescents to be globally unmotivated (Yeager, Lee, & Dahl, 2017, p. 431). Likewise, we found no support for the assumption of a stronger decline in students' motivation or self-esteem after the transition to middle school (or high school) because of a stronger mismatch of students' needs and opportunities. Further research could test whether transition to middle school impacts specific facets of students' motivation and what kind of school environment might foster students' motivation after the school transition.

3.6.5. Further Moderators

Domain. Change was not significantly moderated by specific subject domains (i.e., math, language, and general domain) in any constructs except for academic self-concept. In academic self-concept, we observed significant decreases in the math and language domain but not in the general domain. One explanation for the declines in domain-specific academic self-concepts is the age-related differentiation of academic self-concepts that is indicated by decreasing correlations between the different academic self-concept domains with age (Denissen, Zarrett, & Eccles, 2007; Marsh, Craven, & Debus, 1998; Schmidt et al., 2017). Younger students tend to perceive their competence in different academic domains as similarly high, whereas older adolescents tend to perceive themselves as more competent in some domains but less competent in other domains. Note that Marsh's (1986) internal/external frame of reference model explains this finding by dimensional comparison processes across academic domains that "lower the self-concept in the worse off domain while raising it in the better off domain" (Möller & Marsh, 2013, p. 546). This increasing domain specificity of academic self-concepts with age could be one explanation for the average decreases in the specific academic self-concepts. On the other hand, correlations between the different academic self-concepts and the general academic self-concept typically do not decrease with age (Marsh, Craven, & Debus, 1998; Schmidt et al., 2017). That is, the general academic self-concept seems to be unaffected by the dimensional comparison processes across academic domains. This could explain why we found a decrease only in math and language academic self-concepts but not in the general academic self-concept. Of note, these assumptions are quite speculative and need more investigation.

Questionnaire. The particular questionnaire used was not significantly associated with the development in any construct except for academic self-concept. In academic self-concept, we found a larger decrease in studies that used the scales provided by Eccles and colleagues (1983) in comparison to studies that used the questionnaire of Harter (1985). All studies that gathered academic self-concept by the questionnaire of Eccles and colleagues (1983) investigated either the math domain or the language domain of academic self-concept, whereas most studies that used Harter's questionnaire (1985) investigated the general domain of academic self-concept. This confound of questionnaire and academic domain is a plausible explanation for the observed differences between these questionnaires as we found significant decreases in domain-specific academic self-concepts but not in students' general academic self-concept.

Duration of interval. In academic self-concept, intrinsic motivation, and mastery goal orientation, longer intervals between measurement points were associated with a larger decrease. As the decreases generalized over age groups in academic self-concept and intrinsic motivation, possible cumulative long-term implications of the decrease in these constructs should not be underestimated. We suggest the metaphor of a long-lasting drizzle to describe the development of the motivational variables through students' school career. While over a short period the drop in motivational variables appears to be rather small ("drizzle"), changes over a longer period might be substantial as a huge quantity of water falls on earth from several years of drizzle.

Year of data collection and year of publication. Year of data collection was only significantly associated with the development of PAP and year of publication was only significantly associated with the development of intrinsic motivation and PAP. The decrease in PAP was smaller in older studies than in more recently published ones, whereas a bigger decrease in intrinsic motivation was observed in older publications.

The fact that year of data collection and year of publication were not associated with the development of the other constructs does not support the assumption that administrative changes in the educational system during the last decades (e.g., Au, 2011; Borman, Hewes, Overman, & Brown, 2003; Ertmer & Ottenbreit-Leftwich, 2010) play a crucial role in the development of students' motivation and self-esteem. In accordance with this conclusion, meta-analytic research on academic achievement suggest that distal variables such as state policies are more weakly related to academic achievement than proximal variables such as differences between teachers (Hattie, 2008; Schneider & Preckel, 2017).

Geographic location. Studies from Europe reported a significantly larger motivational decrease than studies from North America or Asia. These differences might be explained by distinct learning ideologies, school politics, and culture-specific differences in learning-related mindsets (i.e., growth mindset). For example, in comparison with students from western countries students in China perceive learning as a priority and put much effort to achieve mastery of the material, and to perfect themselves morally and socially (Li, 2005; Tweed & Lehman, 2002). In addition, in the United States school tracking is implemented on the within-school level and on specific academic domains, whereas in many European countries it is carried out on the between-school level and on the overall achievement (Schnabel et al., 2002; Nagy et al., 2010). Recent research suggests that the influence of tracking on students' motivation differs according to the particular type of tracking (Chmielewski, Dumont, & Trautwein, 2013). In some European countries like Germany tracking occurs very early (i.e., in grade 5). Therefore, cultural context and school politics need to be taken into consideration when conducting research of motivational development. More cross-cultural research including measurement invariance tests between cultures and operationalization of cultural differences in school systems is needed to explain why the motivational development differs by geographic location.

3.6.6. Conclusions

Examining 107 independent longitudinal studies that investigated the development of students' motivation and self-esteem during their school career, we found decreases in intrinsic motivation, mastery, and PAP. We further found a decrease in subject-specific academic self-concepts (i.e., math academic self-concept and language academic self-concept), but not in students' general academic self-concept. Reliable statements about the development of academic self-efficacy and about the development of PAV, which we found to be stable over time, are hard to make because of the small number of independent studies that investigated these constructs.

The motivational declines were moderate in size and specific to some constructs. More research is needed to clarify how practically relevant such moderate declines are. Previous longitudinal investigations have demonstrated that the change in motivational variables over around one year predicts subsequent school achievement controlling for its previous level (e.g., Gehlbach, 2006; Wang & Pomerantz., 2009). It is therefore plausible to assume that even moderate motivational declines are relevant for students' academic development which calls for interventions to counteract the decline in students' motivation. Based on SEF theory, Eccles and Roser (2009) proposed that the fit between students' needs and opportunities offered by the school environment could be improved on several ecological levels (i.e., classrooms, schools, school districts, and communities). For example, the school start time is one possible point of intervention at a higher ecological level. Students develop a preference towards evening-orientation during adolescence, whereas the start time of school typically does not change with students' grade level (Roennenberg et al., 2007). Note that students' orientation towards eveningness is considered a risk factor in the literature as it is negatively related to school grades (e.g., Preckel et al., 2011), and motivational variables such as academic self-concept, school value, and achievement goals (Preckel et al., 2013; Scherrer, Roberts, & Preckel, 2016). In terms of SEF theory, a later school start time might reduce the mismatch between students' rising needs (i.e., the developing need for evening activities) and their opportunities (i.e., no risk for sleep deficits because of a later school start). Alternatively, educational interventions that target particular motivational constructs may be recommended based on our results. As we observed decreases only in intrinsic motivation, mastery, PAP, and in subject-specific self-concepts, intervention programs should primary target these constructs (for a meta-analytic overview of intervention programs, see Hulleman et al., 2016 and Lazowski & Hulleman, 2016). For example, Craven, Marsh, and Raymound (1991) demonstrated that math and reading self-concepts could be enhanced by a researcher-administrated intervention that combined internally focused performance feedback and attributional feedback. Note that intervention programs in self-concept are often specific to particular domains (O'Mara, Green, & Marsh, 2006). That is, if an intervention targets enhancing math self-concept, the largest changes will be observed in math self-concept compared to other domains or the general self-concept. As we observed decreases only in math and language academic self-concepts, intervention programs that target these self-concept domains could be recommended based on our results. Note that this recommendation can be discussed against the background of Marsh's (1986) internal/external (I/E) frame of reference model. According to the I/E model, students compare their domain-specific academic self-concepts by internal and external comparison processes within and across academic domains. That is, students compare their achievement in one domain to the achievement of other students in the same domain (external comparison) and in relation to their own achievement in other domains (internal comparison). Fostering one particular academic self-concept domain of students might entail the devaluation of the other academic self-concept domains because of the internal comparison processes. However, meta-analytic findings regarding interventions in academic self-concepts do not suggest that intervention programs have a negative impact on the untargeted self-concept domains (O'Mara, Green, & Marsh, 2006).

For programs fostering students' intrinsic motivation, the intervention of Guthrie and colleagues (2006) is exemplary. To encourage long-term interest in reading, Guthrie and colleagues (2006) stimulated situational interest by so-called hands-on activities (i.e., physically interacting with a concrete object). Students who received stimulating tasks showed a higher intrinsic motivation to read compared to the control group. As we observed a decline in intrinsic motivation in our results, such intervention programs are needed to prevent motivational decreases. Finally, our findings suggest that intervention programs for fostering students' motivation should not be limited to one particular school stage. That is, intervention programs are necessary in elementary school, middle school, and high school as we found motivational decreases in all school stages.

To conclude, considering different motivational variables and self-esteem in our meta-analysis, we found that the motivational decrease throughout students' school career is not a general phenomenon affecting all constructs in the same vein. In addition, we found that the motivational decline is not restricted to a certain age group. Rather, our findings indicate that supporting a positive motivational development of students is a task that is relevant in all school years and not only at certain critical stages, as small decreases over time might add up to large decreases over the whole school career. Motivation and its development is affected by environmental variables like context, culture, and instruction, and therefore open to (psychological) interventions (e.g., Cohen, Garcia, & Goyer, 2017; Hulleman et al., 2016). We hope that the present meta-analysis will encourage further research that helps ensure positive motivational development in more students from the beginning of their school career on.

Chapter 4. Article 2

Development of Achievement Goals and Their Relation to Academic Interest and Achievement in Adolescence: A Review of the Literature and Two Longitudinal Studies

Vsevolod Scherrer, Franzis Preckel, Isabelle Schmidt, and Adrew J. Elliot

4.1. Abstract

Adolescence is important for the development of achievement motivation, including achievement goal pursuit. Longitudinal research is scarce on adolescents' goal development and its implications for academic outcomes. In our research, we first present a systematic review of findings on achievement goals in adolescence. Then, we report two longitudinal studies with German adolescents in which we investigated the separate as well as joint development of achievement goals, interest, and achievement in the domain of mathematics. Study 1 comprised 745 students assessed in four waves in grades 5-7 (43% female*;* age M_{T1} = 10.66). Study 2 comprised 1420 students assessed in four waves in grades 5-8 (47% female*;* age M_{T1} = 10.58). Students reported their mastery, performance-approach, and performance-avoidance goals and their interest in mathematics. Mathematics achievement was assessed by school grades (from student's records in Study 1, self-reported in Study 2) and standardized test scores (Study 1). Data were analyzed using latent growth curve and multiple processes correlated change structural equation modeling. Findings for the two studies evidenced a substantial degree of consistency. All goals decreased and the decrease became smaller over time in most instances. In both studies, findings for the joint development of goals with interest and achievement revealed that change in mastery and performance-approach goals was positively related or unrelated to change in interest and achievement, while change in performance-avoidance goals was negatively related or unrelated to change in interest and achievement. The findings are discussed in the context of social cognitive and stage-environment fit theories.

Keywords: achievement goals, adolescence, development, interest, achievement

4.2. Introduction

Achievement goals have been central to the study of achievement motivation in students for over three decades (Dweck, 1986; Elliot & Hulleman, 2017; Nicholls, 1984). An achievement goal is an individual's competence-relevant purpose for engaging in behavior, and this goal is thought to establish a framework for how students think, feel, and behave in school settings. The specific types of goals that students pursue are posited to influence their achievement-relevant outcomes.

A large literature on students' achievement goals and their implications (several hundreds of studies) has accumulated over the years. The vast majority of this research has been cross-sectional. Some longitudinal research is also available, but it is relatively scarce. In the present research, we focus on adolescents' achievement goals and their implications for academic interest and achievement. Adolescence is typically considered a critical stage for the development of motivational constructs (Gottfried, 2019; Nicholls, 1989; Wigfield & Eccles, 2002) and academic interest and achievement are two central ("gold standard") achievement outcomes (Korn & Elliot, 2016, p. 4). Herein we present systematic reviews of the relevant cross-sectional and longitudinal literatures on achievement goals and interest/achievement; these reviews are valuable in both summarizing the current state of empirical work in this area and establishing the context for the present research. We then report two new longitudinal studies investigating the mean-level development of achievement goals over time as well as the joint development of achievement goals, interest, and achievement in the domain of mathematics.

4.2.1. Achievement Goal Models

Several different models of achievement goals have been proffered over the years. The initial, dichotomous model distinguished between two types of goals: Mastery goals that focus on trying to develop competence through task mastery and performance goals that focus on trying to demonstrate competence relative to others (Dweck, 1986; Nicholls, 1984). Both of these were explicitly characterized as approach goals (i.e., goals focused on the possibility of success; Ames, 1992; Nicholls, Patashnick, Cheung, Thorkildsen, & Lauer, 1989). In the trichotomous model, mastery goals remained the same, but performance goals were bifurcated into performance-approach goals that focus on trying to demonstrate competence relative to others and performance-avoidance goals that focus on trying to avoid demonstrating incompetence relative to others (Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997). A subsequent 2 x 2 achievement goal model (Elliot, 1999; Pintrich, 2000a) extended the trichotomous model by additionally distinguishing between mastery-approach goals that focus on trying to approach intrapersonal or task-based competence, and mastery-avoidance goals that focus on trying to avoid intrapersonal or task-based incompetence. Finally, a 3 x 2 model (Elliot, Murayama, & Pekrun, 2011) split both mastery-approach and mastery-avoidance goals according to the intrapersonal/task-based distinction, yielding two approach goals (i.e., trying to approach intrapersonal competence and trying to approach task-based competence) and two avoidance goals (i.e., trying to avoid intrapersonal incompetence and trying to avoid task-based incompetence).

In the present research, we focus on the trichotomous model of achievement goals. The trichotomous model was the predominant model at the time that the longitudinal studies of the present research were started. Furthermore, the three goals of this model have been the most studied in the achievement goal literature and are thought to be the most relevant to students of most age groups (Elliot, 2005).

4.2.2. Achievement Goals, Academic Interest and Achievement, and Development

In the achievement goal literature in general, mastery(-approach) goals are posited to be the most beneficial of the three goals for students' achievement-relevant processes and outcomes. The appetitive and intrapersonal/task-based focus of these goals is thought to facilitate processes such as challenge appraisals, task absorption, deep-processing, persistence, and autonomous engagement, which should have positive implications for both interest and achievement over time (Dweck, 1999; Nicholls, 1989). Performance-avoidance goals are assumed to be the most detrimental of the three goals. The aversive and demonstration/other focus of these goals presumably facilitates processes such as threat appraisals, self-concern, anxiety, inflexible processing, and self-protective (dis)engagement, with negative implications for both interest and achievement (Elliot & Hulleman, 2017). Performance-approach goals are posited to fall between mastery and performance-avoidance goals in their beneficial-detrimental implications. The appetitive focus of these goals is likely to promote some of the same positive processes as mastery goals, but the demonstration/other focus of the goals is likely to promote some of the same negative processes as performance-avoidance goals. The net influence of these goals for students' interest and achievement may depend on whether students focus on demonstrating ability, using others as a standard, or both; the demonstration focus of goals is viewed as particularly pernicious, and likely to undermine both interest and achievement (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010; Senko & Dawson, 2017).

Several theorists have posited that adolescence brings intrapsychic and environmental changes that are problematic for students' motivation, including their achievement goal pursuit. Social-cognitive theorists (Dweck, 1986; Nicholls, 1984) note that cognitive development during early adolescence facilitates the differentiation of ability from effort, and that this leads many students to focus on demonstrating ability with as little effort as possible. Stage-environment fit (SEF) theorists (Eccles et al., 1993) contend that adolescence brings a mismatch between students' emerging needs and the types of school environments that they encounter (especially following the middle school and high school transitions). Specifically, adolescent students encounter a disruption in friendship networks, fewer opportunities to develop close relationships with their teachers, increased normative evaluation, and a decrease in felt autonomy in their school environment (Eccles et al., 1989; Midgley & Feldlaufer, 1987; Wigfield & Wagner, 2005), all at a developmental period when they are particularly in need of strong relational connections, moderately challenging academic tasks that foster skill development, and a sense of autonomy and choice (Eccles et al., 1993; Simmons & Blyth, 1987).

Social-cognitive and SEF theorists have offered the same set of predictions for achievement goal pursuit—a decrease in mastery goal pursuit and an increase in performance goal pursuit during the adolescent period (Anderman & Midgley, 1997; Eccles et al., 1993; Nicholls, 1989). This set of predictions was offered prior to performance goals being bifurcated by approach-avoidance, and performance-approach goals being identified as at least somewhat adaptive. Two alterative sets of predictions that incorporate the approach-avoidance distinction may be derived from these theories in a way that retains their core premise of motivational difficulty during adolescence. One alternative is that the myriad changes and challenges faced by adolescents lead to a decreased commitment to competence per se, across standards of evaluation, such that not only mastery goal pursuit, but also performance-approach and performance-avoidance goal pursuit decreases (Paulick, Waterman, & Nückles, 2013). This would represent a complete decline in the *quantity* of motivation. A second alternative is that adolescents use self-protection as a mechanism for coping with the changes and new challenges they encounter (Alicke & Sedikides, 2011; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001), which is maximized by a decrease in approach goal pursuit (mastery and performance-approach) and an increase in performance-avoidance goal pursuit. This would represent a complete decline in the *quality* of motivation.

Both social-cognitive and SEF theory focus on mean-level changes in achievement goal pursuit during adolescence; neither theory focuses on nor offers predictions regarding the implications of this achievement goal change for adolescents' downstream outcomes such as academic interest and achievement. Two possibilities seem plausible. One possibility is that the predictive patterns for the three achievement goals during adolescence are similar to the predictive patterns articulated above for individuals in general—mastery goals are most beneficial, performance-avoidance goals are most detrimental, and performance-approach goals are in between (Elliot & Moller, 2003; Midgley, Middleton, Gheen, & Kumar, 2002). A second possibility emphasizes the importance of students adapting to their changing school climate. As the academic environment becomes more performance-focused and less mastery-focused following the middle school and high school transitions (Anderman & Midgley, 1997; Midgley, Anderman, & Hicks, 1995), performance-approach goals may become more adaptive and beneficial for students, and mastery goals may become less adaptive and beneficial (Harackiewicz, Barron, Carter, Lehto, & Elliot, 1997; Shim, Ryan, & Anderson, 2008). The avoidance component of performance-avoidance goals would render these goals maladaptive for students across school climates (Murayama & Elliot, 2009).

4.2.3. Narrative Reviews of the Literature

In the following, we review the cross-sectional and longitudinal research that has been conducted examining relations between the goals of the trichotomous achievement goal model and both academic interest and academic performance. In addition, within the longitudinal research, we focus on the development of achievement goals over time (per se). We include work on mastery goals from the dichotomous achievement goal model and mastery-approach, performance-approach, and performance avoidance goals from the 2 x 2 achievement goal model, given that they are conceptually the same as the goals from the trichotomous model (Elliot & Hulleman, 2017). To the extent possible, we focus on adolescent students in these reviews.

Academic interest may be defined as interest in or enjoyment of an educational activity for its own sake (Deci & Ryan, 1985). This variable is usually operationalized using self-reports of the extent to which the respondent finds a school-based activity, a class, or schoolwork to be enjoyable, fun, or interesting. Academic achievement may be defined as performance relative to a standard of competence in an educational environment (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014). This variable is usually operationalized using objective records or self-reports of exam scores, class or school grades, or standardized tests.

Cross-sectional research. The vast majority of research on achievement goals and academic interest/achievement has used cross-sectional methodologies, most commonly a single session correlational design or a prospective design in which the goals are assessed at one time point and the outcomes are assessed at a second time point. In our review, we focus on published meta-analyses of this cross-sectional literature, targeting the goals of the trichotomous model (or equivalent) and interest and achievement specific to academic environments (when specified). We target meta-analyses that included data from secondary and/or middle school students; we do not include meta-analyses that focused on undergraduates or adults only.

We conducted a systematic literature search for articles containing relevant meta-analyses using the following PsychINFO keywords: (*goal(s) orientation* **OR** *achievement goal(s)* **OR** *performance goal(s)* **OR** *mastery goal(s)* **OR** *performance-avoidance* **OR** *performance-approach* **AND** *(meta-analysis)* **in** *(peer reviewed journals in the English language)*. We also did a second search within relevant journals (*Psychological Bulletin* **OR** *Educational Research Review* **OR** *Educational Psychology Review* **OR** *PLOS ONE* **OR** *Review of Educational Research* **OR** *Personality and Social Psychology Review*). We used the same keywords as in the first search strategy and searched within the titles and abstracts of articles published in these journals. The search process, as well as the inclusion and exclusion criteria, are reported in Figure B1 in Appendix B.

We found 9 meta-analyses through our search process. The main findings from these meta-analyses may be summarized in terms of the median relation for each achievement goal and outcome variable (see Table 5 for additional details); when there was an even number of coefficients, we reported the mean of the two middle numbers. Mastery goals showed a positive relation with academic interest (median r = .44, median number of studies in the meta-analyses = 13) and academic achievement (median r = .13, median number of studies in the meta-analyses = 74). Performance-approach goals showed a positive relation with both academic interest (median r = .15, median number of studies in the meta-analyses = 13) and academic achievement (median r = .10, median number of studies in the meta-analyses = 76). Performance-avoidance goals showed a negative relation with both academic interest (median r = -.07, median number of studies in the meta-analyses = 13) and academic achievement (median r = -.13, median number of studies in the meta-analyses = 59).[8]

[8] In summarizing the data from these meta-analyses, we used the goals from the trichotomous model when available, but used the equivalent goals from the dichotomous or 2 x 2 models when necessary. Likewise, we used data from measures of interest when available, but used measures of enjoyment or intrinsic motivation when necessary. Data from 3 meta-analyses that focused only on university students and adults (Cellar et al., 2010; Payne, Youngcourt, & Beaubien, 2007; Richardson, Abraham, & Bond, 2012) were not included in the summary. All 9 meta-analyses that were included in the summary considered data from secondary students, as well as data from university students and adults. Six of these meta-analyses tested age, grade level, or school type as a moderator of the relation between the achievement goals and the academic outcomes (Huang, 2011; Huang, 2012; Hulleman et al., 2010; Senko & Dawson, 2017; Van Yperen, Blaga, & Postmes, 2014; Wirthein, 2013). Age, grade level, and school type did not significantly moderate the relations of achievement goals and academic outcomes, excepting two significant effects. Hulleman et al. (2010) found that the positive relation between PAP goals and interest decreases with grade level, and Wirthwein et al. (2013) reported a stronger positive relation between PAP goals and academic achievement in university students compared to middle school students. Two additional types of research related to the work covered in the text may be briefly mentioned. First, research has been conducted on manipulated achievement goals and interest/achievement, but the majority of this research focused on lab-based games and exercises rather than academic activities (for meta-analyses, see Rawsthorne & Elliot, 1999; Van Yperen, Blaga, & Postmes, 2015). In addition, some research has focused on linking achievement goals to performance in sport contexts (for a meta-analysis, see Lochbaum & Gottardy, 2015).

Table 5. *Meta-analyses of Cross-sectional Studies Investigating Relations of Achievement Goals with Academic Interest and Academic Achievement in School-aged Students and Adults*

	Considered achievement goals	Relations to interest-related variables	Relations to academic achievement
Baranik et al.		*Interest*	*Grades, exam performance*
(2010)	M:	*r*=.61*/ *k*=12	*r*=.10*/ *k*=17
	PAP:	*r*=.17*/ *k*=12	*r*=.13*/ *k*=17
	PAV:	*r*=.09*/ *k*=12	*r*=-.18*/ *k*=17
Burnette et al.			*Achievement*
(2013)	MAP:		*r*=.14/k=4
	PAP:		*r*=.16/k=7
	PAV:		*r*=-.22/k=7
Huang (2011)		*Interest*	
	M (from dich):	*r*=.42*/ *k*=16	
		Enjoyment	
	M (from dich):	*r*=.43*/ *k*=9	
		Interest	
	M (from trich):	*r*=.54*/ *k*=13	
	PAP (from trich):	*r*=.21*/ *k*=13	
	PAV (from trich):	*r*=-.08/ *k*=13	
		Enjoyment	
	M (from trich):	*r*=.42*/ *k*=6	
	PAP (from trich):	*r*=.04/ *k*=6	
	PAV (from trich):	*r*=-.08/ *k*=6	
Huang (2012)			*Grades, test scores*
	M (from dich):		*r*=.13*/ *k*=84
			Grades, test scores
	M (from trich):		*r*=.13*/ *k*=76
	PAP (from trich):		*r*=.07*/ *k*=76
	PAV (from trich):		*r*=-.12*/ *k*=76
			Grades, test scores
	M (from 2 x 2):		*r=.10*/k=19*
	PAP (from 2 x 2):		*r*=.13*/ *k*=19
	PAV (from 2 x 2):		*r*=-.13*/ *k*=19
Hulleman et al. (2010)		*Interest*	*Performance outcomes*
	M:	*r*=.44*/ *k*=52	*r*=.11*/ *k*=95
	PAP:	*r*=.07*/ *k*=52	*r*=.06*/ *k*=98
	PAV:	*r*=-.07*/ *k*=34	*r*=-.13*/ *k*=63
Murayama and Elliot (2012)			*Performance*
	PAP:		*r*= .10*/ *k*=472
	PAV:		*r*= -.12*/ *k*=472
Senko and Dawson (2017)		*Enjoyment*	
	PAP:	*r*=.15*/ *k*=35	
Van Yperen, Blaga, and Postmes			*Educational achievement*
(2014)	M:		*r*=.13*/ *k*=72
	PAP:		*r*=.10*/ *k*=75
	PAV:		*r*=-.14*/ *k*=55
Wirthwein et al. (2013)			*Academic achievement*
	M:		*r*=.13*/k=209
	PAP:		*r*=.08*/k=142
	PAV:		*r*=-.12*/k=109

Note. $^*p < .05$. M = Mastery goals. PAP = performance-approach goals. PAV = performance-avoidance goals. Dich = dichotomous achievement goal model. Trich = trichotomous achievement goal model. All meta-analyses considered school-aged students and adults.

Longitudinal research. Longitudinal methodologies involve repeated measurement of an independent variable, a dependent variable, or both over at least two time periods (Caruana, Roman, Hernández-Sánchez, & Solli, 2015). Longitudinal studies on the relation between achievement goals and academic interest/achievement are sparse relative to cross-sectional studies. Meta-analyses investigating the longitudinal relations of achievement goals and academic outcomes have yet to be conducted, undoubtedly due to the limited number of studies available in the literature. Herein, we conduct a narrative review of existing longitudinal studies with adolescents, targeting the goals of the trichotomous model (or equivalent) and academic interest, academic achievement, or both. Within this set of studies, we also provide a narrative review of mean-level change over time for each of the three achievement goals. We target studies focusing on middle school and/or high school students only (i.e., adolescents) in this review.

We conducted a systematic literature search for articles containing relevant studies using the following PsycINFO keywords search: (*goal(s) orientation* **OR** *achievement goal(s)* **OR** *performance goal(s)* **OR** *mastery goal(s)* **OR** *performance-avoidance* **OR** *performance-approach*) **AND** (*longitudinal* **OR** *adolescent development* **OR** *childhood development* **OR** *secondary school* **OR** *middle school* **OR** *high school*) **AND** (*academic achievement* **OR** *values* **OR** *motivation* **OR** *interest*) **in** (*peer reviewed journals in the English language*). In addition, we searched the reference section of a recent review on competence motivation development by Scherrer and Preckel (2019) for other relevant studies. The search process, as well as the inclusion and exclusion criteria, are reported in Figure B2 in Appendix B.

Our search process revealed 16 articles reporting 17 longitudinal studies. We categorized these articles according to three different analytic approaches, A to C, used to investigate the longitudinal relations of achievement goals and academic interest and/or achievement: A = previous goals predict subsequent interest/achievement, B = correlated change of goals and interest/achievement, and C = goal profiles predict subsequent interest/achievement (see Table 6 for additional details on these studies).[9] Note that the considered studies varied in terms of the duration between measurement points. Because the magnitudes of the coefficients' absolute values will shift as the time lags become shorter or

[9] Additional longitudinal research related to the work reviewed in the text may be noted. Lazarides et al. (2017), Martin and Elliot (2016), and Tuominen-Soini et al. (2011) investigated the longitudinal relations of achievement goals and general school value. General school value captures perceived usefulness and perceived importance, in addition to interest in school. In many instances, multiple analytic approaches were used within a single study.

longer, the observed magnitudes of the path coefficients are not directly comparable across different time spans (Oud & Delsing, 2010).

In Approach A, 10 articles reported cross-lagged effects of achievement goals on academic interest or achievement using structural equation modeling (e.g., Spinath & Steinmayr, 2012), multiple regression (e.g., Wentzel, 1996), or latent growth curves (Wang & Pomerantz, 2009). Cross-lagged investigations test how mean-levels of achievement goals at one measurement point predict a subsequent outcome controlling for the previous level of the outcome. Mastery goals were positive cross-lagged predictors of subsequent academic interest (Spinath & Steinmayr, 2012) and positive (Niepel, Brunner, & Preckel, 2014; Paulick, Watermann, & Nückles, 2013; Seaton et al., 2013; Wang & Pomerantz, 2009) or non-significant (Gutman, 2006; Patrick, Ryan, & Pintrich, 1999; Rubie-Davies & Peterson, 2016; Wentzel, 1996) predictors of subsequent academic achievement. Researchers have found no significant cross-lagged effects of performance-approach goals on academic interest (Spinath & Steinmayr, 2012). The findings for cross-lagged effects of performance-approach goals on academic achievement are mixed; some studies reported positive cross-lagged relations (Niepel et al., 2014; Rubie-Davies & Peterson, 2016, in a sample of Mäori; Seaton et al., 2013), whereas others found no significant cross-lagged relations (Rubie-Davies & Peterson, 2016, in a sample of Päkeha), and some have even found negative cross-lagged relations (Paulick et al., 2013). Performance-avoidance goals have been shown to be non-significant cross-lagged predictors of academic interest (Spinath & Steinmayr, 2012). Either negative (Niepel, Brunner, & Preckel, 2014) or non-significant (Paulick et al., 2013) cross-lagged effects of performance-avoidance goals on academic achievement have been found.

In Approach B, 4 articles investigated how interindividual differences in the development of achievement goals were related to interindividual differences in the development of academic interest or achievement. In these articles, a change in achievement goals was represented using manifest change scores (Gehlbach, 2006; Shim et al., 2008), latent change scores (e.g., Lazarides & Raufelder, 2017), or latent growth curves (e.g., Wang & Pomerantz, 2009). Change in mastery goals was positively related to change in academic interest (Gehlbach, 2006; Lazarides & Raufelder, 2017) and achievement (Gehlbach, 2006; Shim et al., 2008; Wang & Pomerantz; 2009). Change in performance-approach goals was not significantly related to change in interest (Gehlbach, 2006), but was positively related to change in achievement (Shim et al., 2008). Approach B has not been used to test the relations of change in performance-avoidance goals and change in academic interest or achievement.

In Approach C, 3 articles reported a person centered-approach using median split or latent profile analyses to estimate achievement goal profiles of students and to predict academic outcomes measured at a later time point by these profiles, controlling for the previous level of the outcome (Hornstra, Majoor, & Peetsma, 2017; Pintrich, 2000b; Schwinger & Wild, 2012). The results of these studies are hard to quantify or to compare because different achievement goal profiles were estimated in these studies. In the study by Hornstra et al. (2017), students having approach oriented profiles (high mastery, high performance-approach, and low performance-avoidance goals) and moderate profiles (medium mastery, performance-approach, and performance-avoidance goals) showed stronger increases in academic achievement than students having multiple goal profiles (medium to high mastery, performance-approach, and performance-avoidance goals). However, in the studies by Pintrich (2000b) and Schwinger and Wild (2012), achievement goal profiles at a previous measurement point had no consistent effects on subsequent academic interest or achievement, controlling for the previous level of these outcomes.

With regard to mean-level change of the achievement goals per se, few studies explicitly focused on this issue, but 9 studies provided descriptive statistics that allowed us to compute the effect size of change ($d = \frac{M\text{t}2 - M\text{t}1}{SD\text{t}1}$) for goals at adjacent time points. For ease of interpretation, we additionally report an adjusted average *d* value for each achievement goal that refers to an interval duration of one year (e.g., when a two-year interval was considered and *d* = .20 was conducted, this was scaled to *d* = .10 for the duration of one year). The findings may be summarized in terms of the median effect size of change for one year. Mastery goals (median *d* = -.14, number of studies = 9), performance-approach goals (median *d* = -.16, number of studies = 4), and performance-avoidance goals (median *d* = -.20, number of studies = 3) showed small mean-level declines over time.[10] With the exception of Paulick et al. (2013), the issue of the functional form of change over time (i.e., linear versus non-linear) has been ignored in this literature.

[10] Some of the studies reviewed herein provided multiple, non-independent effect sizes, so caution in interpreting the medians is warranted, accordingly.

Table 6. *Longitudinal Studies Investigating Relations of Achievement Goals with Academic Interest and Academic Achievement in Middle School and High School Students*

	Key findings on longitudinal relations to academic interest and achievement assorted by analysis approach A. to C.	Mean-level change in *d*	Duration (in years) between time points	Scaled up average *d* for 1 year
Gehlbach (2006) *n*=917 (from US) T1=grade9/10 (start) T2=grade9/10 (end)	B. The change score of mastery goals positively predicted subsequent achievement on history (history test: *b*=.08*; grades in history: *b* =.23*), and subsequent interest in social studies classes (*b*=.52*) at T2. The change score of PAP goals did not significantly predicted subsequent achievement on history and subsequent interest in social studies classes at T2. This was tested by multiple regression with T1 outcomes as further predictors.	M1-2: *d*=-.25 PAP1-2: *d*=-.14	T1-T2: .75	M: *d*=-.33 PAP: *d*=-.19
Gutman (2006) *n*=50 (from US) T1=grade8 T2=grade9	A. T1 mastery goals did not significantly predict T2 grades (controlled for T1 grades). This was tested by multiple regression. T2 mastery was also included as a predictor of grades in this model.	M1-2: *d*=.11	T1-T2: 1.00	M: *d*=.11
Hornstra, Majoor, and Peetsma (2017) *n*=722 (from the Netherlands) T1=grade5 (halfway) T2=grade6 (start) T3=grade6 (halfway)	C. Approach-oriented (high mastery, high PAP, and low PAV goals) and moderate profiles (medium mastery and moderate PAP and PAV goals) showed stronger increases on language achievement tests than multiple goals profile (medium to high mastery, PAP, and PAV goals). No associations between goal profiles and math achievement tests were found. This was tested by latent profile and multilevel analyses.	Could not be conducted.	T1-T2: .75 T2-T3: .25	Could not be conducted .
Lazarides and Raufelder (2017) *n*=1088 (from Germany) T1=grade8 (start) T2=grade9 (end)	B. Change in mastery goals was positively associated with change in intrinsic motivation (*β*=.13*). This was tested by relations of latent change scores in a structural equation model.	M1-2: *d*=-.04	T1-T2: 1.75	M: *d*=-.03
Niepel, Brunner, and Preckel (2014) *n*=769 (from Germany) T1=grade5 T2=grade6 T3=grade8	A. T1 mastery (*β*=.20*), T1 PAP (*β*=.20*), and T1 PAV goals (*β*=-.37*) were significant predictors of subsequent T2 grades. T2 mastery goals (*β*=.14*) positively predicted subsequent T3 grades, whereas T2 PAP and T2 PAV goals did not significantly predict subsequent T3 grades. This was tested by a cross-lagged structural equation model.	Could not be conducted.	T1-T2: 1.00 T2-T3: 2.00	Could not be conducted .
Patrick, Ryan, & Pintrich (1999) *n*=445 T1=grade7-8 (start) T2=grade7-8 (end)	A. T1 mastery goals in math, English, and social studies did not significantly predict T2 grades (controlled for T1 grades). This was tested by multiple regression.	Math M1-2: *d*=-.22 English M1-2: *d*=-17 Social studies M1-2: *d*=-.17	T1-T2: .75	Math M: *d*=-.29 English M: *d*=-.23 Social studies M: *d*=-.23

Paulick, Watermann, and Nückles (2013) *n*=1646 (from Germany) T1=grade 4 (end) T2=grade5 (halfway) T3=grade6 (start)	A. T1 mastery (β=.14*) and T1 PAP goals (β=-.10*) significantly predicted subsequent T2 grades, whereas T1 PAV goals (β=.02) did not significantly predicted subsequent T2 grades. T2 mastery goals (β=.12*) significantly predicted subsequent T3 grades, whereas PAP β= (β=-.02) and PAV goals (β=-.02) did not significantly predict T3 grades. This was tested by a cross-lagged structural equation model.	M1-2: *d*=-.24 M2-3: *d*=-.54 PAP1-2: *d*=-.62 PAP2-3: *d*=-.23 PAV1-2: *d*=-.66 PAV2-3: *d*=-.47	T1-T2: .75 T2-T3: .75	M: *d*=-.52 PAP: *d*=-.57 PAV: *d*=-.75
Pintrich (2000b) *n*=150 T1=grade8 (start) T2=grade8 (end) T3=grade9 (end)	C. Participants were assorted to one of four goals categories by median split of their T1 mastery and PAP goal ratings. These four groups (high M/high PAP; high M/low PAP; low M/ high PAP; low M/ low PAP) were not related to the change in school grades (T1, T2, and T3) This was tested by ANOVA with T1, T2, and T3 grades as repeated measures and T1 goals category as predictor.	Could not be conducted.	T1-T2: .75 T2-T3: 1.00	Could not be conducted.
Rubie-Davies, & Peterson (2016) *n*=78 (Mäori from New Zealand) *n*=319 (Päkeha from New Zealand) T1=grade7-8 (start) T2=grade7-8 (end)	A. T1 mastery and T1 PAP goals did not significantly predict T2 achievement test scores (controlled for T1 achievement) in Päkeha. In the Mäori sample, T1 mastery was not a significant predictor of T2 achievement test scores (controlled for T1 achievement), whereas T1 PAP (*b*=.17*) positively predicted the subsequent T2 achievement test scores. This was tested by a multilevel regression.	Could not be conducted.	T1-T2: .75	Could not be conducted.
Schwinger and Wild (2012) *n*=302 (from Germany) T1=grade3 T2=grade4 T3=grade5 T4=grade6 T5=grade7	C. No differences between the latent profiles were found on subsequent test performance. On interest, only a single effect was found: Class 2(moderate multiple goals) showed the lowest subsequent interest (β=-.11*). Furthermore, subsequent school grades were negatively predicted by class 2 (moderate multiple goals; β=-.15*) and positively predicted by class 3 (primarily mastery-oriented; β =.16*). This was tested by latent profile analyses of achievement goals and cross-lagged models based on these latent profiles.	Could not be conducted.	T1-T2: 1.00 T2-T3: 1.00 T3-T4: 1.00 T4-T5: 1.00	Could not be conducted.

Seaton et al. (2013) n=2786 (from Austria) T1=grade7/8/9/10 T2=T1+6 months T3=T2+6 months T4=T3+6 months	A. T1 Mastery goals did not significantly predict T2 achievement test scores and T3 mastery goals did not significantly predict T4 achievement test scores, whereas T2 mastery goals (β=.04*) positively predicted T3 achievement test scores. T1 PAP goals did not significantly predict T2 achievement test scores, whereas T2 PAP goals (.04*) positively predicted T3 achievement test scores and T3 PAP goals (.03*) positively predicted T4 achievement test scores. This was tested by a cross-lagged structural equation model.	Could not be conducted.	T1-T2: .50 T2-T3: .50	Could not be conducted .
Sebanc, Guimond, & Lutgen, (2016) *n*=146 (From US) T1=grade5 (end) T2=grade6 (start) T3=grade6 (end)	A. T1 mastery goals did not predict T2 grades (controlled for T1 grades). T2 mastery did not significantly predict T3 grades (controlled for T2 grades). This was tested by multiple group (boys vs. girls) structural equation modeling.	M1-2: *d*=.19 M2-3: *d*=-.32	T1-T2: .25 T2-T3: .75	M: *d*=.06
Shim, Ryan, and Anderson (2008) *n*=588 (from US) T1=grade6 (fall) T2=grade6 (spring) T3=grade7 (fall) T4=garde7 (spring)	B. Change in mastery goals over time (T1 to T4; *b*=.12*) and change in PAP goals (T1 to T4; *b*=.15*) over time both positively predicted the rates of change in GPA. This was tested by growth curves and hierarchical linear modeling over time.	M1-2: *d*=-.18 M2-3: *d*=-.01 M3-4: *d*=-.11 PAP1-2: *d*=-.35 PAP2-3: *d*=.03 PAP3-4: *d*=-.07 PAV1-2: *d*=-.31 PAV2-3: *d*=-.02 PAV3-4: *d*=-.10	T1-T2: .50 T2-T3: .50 T3-T4: .50	M: *d*=-.20 PAP: *d*=-.26 PAV: *d*=-.29

Spinath and Steinmayr (2012) *n*=348 (from Germany) T1=grade11 T2=grade12	A. T1 mastery goals significantly predicted T2 intrinsic motivation (general: β=.21*; math: β=.30*; language: β=.35*). T1 PAP (general: β=.10; math: β=.09; language: β=.10) and T1 PAV (general: β=.02; math: β=.06; language: β=.01) did not significantly predict T2 intrinsic motivation. This was tested by a cross-lagged structural equation model.	General M1-2: *d*=.04 Math M1-2: *d*=-.01 Language M1-2: *d*=-.06 General PAP1-2: *d*=-.13 Math PAP1-2: *d*=-.07 Language PAP1-2: *d*=-.10 General PAV1-2: *d*=-.20 Math PAV1-2: *d*=-.15 Language PAV1-2: *d*=-.20	T1-T2: 1.00	General M: *d*=.04 Math M: *d*=-.01 Language M: *d*=-.06 General PAP: *d*=-.13 Math PAP: *d*=-.07 Language PAP: *d*=-.10 General PAV: *d*=-.20 Math PAV: *d*=-.15 Language PAV: *d*=-.20
Wang and Pomerantz (2009) Sample 1: *n*=374 (from US T1=grade7 (fall) T2=grade7 (spring) T3=grade8 (fall) T4=grade8 (spring) Sample 2: *n*=451 (from China) T1=grade7 (fall) T2=grade7 (spring) T3=grade8 (fall) T4=grade8 (spring)	A. Sample1: T1 mastery goals (β=.10*) positively predicted T4 school grades while controlling for T1 school grades. Sample2: T1 mastery goals (β=.08*) positively predicted T4 school grades while controlling for T1 school grades. B. Sample1: Mastery slope (β=.13*) positively predicted T4 school grades while controlling for T1 school grades. Sample2: Mastery slope (β=.12*) positively predicted T4 school grades while controlling for T1 school grades. This was tested by structural equation (including growth curve) modeling of mastery and auto-regression of school grades.	Sample 1: M1-2: *d*=-.18 M2-3: *d*=-.02 M3-4: *d*=-.03 Sample 2: M1-2: *d*=-.16 M2-3: *d*=.03 M3-4: *d*=-.08	T1-T2: .50 T2-T3: .50 T3-T4: .50	Sample 1 M: *d*=-.15 Sample 2 M: *d*=-.14
Wentzel (1996) *n*=290 (from Finland) T1=grade6 (end) T2=grade8 (end)	A. T1 mastery goals did not predict T2 English grades (controlling for T1 English grades). This was tested by multiple regression.	Could not be conducted.	T1-T2: 2.00	Could not be conducted.

Note. *p < .05. M = Mastery. PAP = performance-approach goals. PAV = performance-avoidance goals. $d = \frac{Mt2-Mt1}{SDt1}$. Negative *d* values indicate a mean-level decrease. Analysis approach A. = Previous achievement goals predict subsequent interest/achievement. Analysis approach B. = Change in goals predictssubsequent interest/achievement. Analysis approach C. = Latent profiles of the achievement goals predict subsequent interest/achievement. The standardized β is reported when available. The unstandardized *b* is reported when the standardized β was not available.

Conclusions from the reviews. From the theoretical overview and empirical reviews of the literature, several things become clear. First, adolescence is a critically important time in which to study achievement goals and their implications. Second, a great deal of cross-sectional research has been conducted in adolescence on links between the goals of the trichotomous achievement goal model and two central achievement outcomes—academic interest and academic achievement. Third, very little longitudinal research has been conducted in adolescence on these goals and outcomes, and on change of achievement goals over time. Fourth, the longitudinal research that has been conducted has tended to support mastery goals as largely adaptive, performance-avoidance goals as maladaptive (or null), and performance-approach goals as in between (positive, mixed, or null), with all three goals showing small declines over time. In short, nearly all of the extant empirical work is cross-sectional, and the sparse longitudinal work that does exist has not yielded clear or consistent findings for performance-based goals. Clearly there is a pressing need for more longitudinal research in this important area.

4.2.4. The Present Research

In the present research, we conducted two longitudinal studies designed to address this need. Specifically, we carried out two longitudinal studies with relatively large samples of adolescent students in Germany that investigated the mean-level development of achievement goals in these samples as well as the functional form of this development (linear or non-linear), and that tested whether interindividual differences in the initial level (i.e., intercept) and development (i.e., slope) of achievement goals, interest, and achievement are related to each other. We conducted our studies in the domain of mathematics, a core, critical subject area in the school curriculum. Students' cognitive ability, sex, socioeconomic status, and language background are important predictors of their interest and achievement in mathematics (Deary, Strand, Smith, & Fernandes, 2007; Frenzel, Goetz, Pekrun, & Watt, 2010; Sirin, 2012; Strenze, 2007; Watt, 2004). We therefore controlled for these variables in our investigations.

In Study 1, 745 students participated in four measurement points from the beginning of Grade 5 (T1) to the middle of Grade 7 (T4). In Study 2, 1420 students participated in four measurement points from the beginning of Grade 5 (T1) to the middle of Grade 8 (T4). A unique strength of our work is that the two longitudinal studies are highly comparable in design, allowing a clear test of the robustness of our results.

As seen in our theoretical overview, several different sets of predictions may be derived for both the mean-level development of achievement goals and the joint development of achievement goals, academic interest, and academic achievement. Empirically, longitudinal data on achievement goals are sparse and inconsistent in linking change in performance-based goals to downstream interest and achievement. Accordingly, we did not generate *a priori* predictions in our studies. Nevertheless, on the basis of our empirical reviews one might tentatively expect the following patterns to emerge: a) a decrease in mastery, performance-approach, and performance-avoidance goals over time (either linear or nonlinear), and b) mastery goals being the most adaptive for academic interest and achievement over time, performance-avoidance goals being the least adaptive, and performance-approach goals being in between.

4.3. Methods Study 1

4.3.1. Participants and Procedure

Data from four waves of measurement were collected as part of the "PULSS" project (Schneider, Stumpf, Preckel, & Ziegler, 2012), which assessed students' development of motivation and achievement upon entering the highest track of the German three-track secondary school system -- Gymnasium. The "PULSS" project was approved by the State Ministry for Education and Culture of Bavaria ("Bayerisches Staatsminesterium für Unterricht und Kultur"; protocol number: lll.4-5 LO355 lll 40 - 1.18 555). The sample comprised 745 students, 425 boys and 320 girls (*M* age at T1 in years = 10.66, *SD* = 0.43) from 7 schools in 2 federal states of Germany (i.e., Baden-Wuerttemberg = 363; Bavaria = 382). Parents' highest educational degree (assessed as the highest degree of either the father or the mother) was distributed as follows within the sample: 0.40% no school degree, 4.94% certificate of secondary education ("Hauptschulabschluss"), 23.38% general certificate of secondary education ("Realschulabschluss"), 18.63% general qualification for university entrance ("Abitur"), 39.35% university degree, and 13.31% PhD. Most students had a German language background (87.17%); the language background of the other students was Russian (4.26%), English (1.65%), Turkish (2.34%), or other (< 1.00%: Chinese, Italian, French, and Spanish).

Data collection started with the 2008/2009 school year (i.e., Fall 2008). Students were tested in groups in their classrooms at the beginning of Grade 5 (T1), at the second term of Grade 5 (T2), at the second term of Grade 6 (T3), and at the second term of Grade 7 (T4). On average, the measurements took place 39 days (T1), 246 days (T2), 606 days (T3), and 917 days (T4) after entry into Gymnasium. Trained research assistants carried out the data collection. During the first few days in Grade 5, students were given questionnaires for their parents to complete at home and return within one week. Students participated voluntarily and written parental consent was obtained for all students. In this study, there were no experimental manipulations, all data exclusions and variables analyzed are reported, and data collection was completed before any analyses were conducted.

4.3.2. Materials

Achievement goals. Mastery goals, performance-approach (PAP) goals, and performance-avoidance (PAV) goals for math were assessed with three items each from the German translation of the Achievement Goal Questionnaire (Elliot & McGregor, 2001). Students responded to the items on a 1 (“not true”) to 5 (“true”) scale. Cronbach’s Alpha ranged from α = .87 to α = .90 across measurement points. All reliability coefficients as well as item wordings are provided in Table B1 in Appendix B.

Math interest. Interest in mathematics was assessed with three items (e.g., “I work on math tasks because I enjoy it.”) on a 1 (“not true”) to 5 (“true”) scale based on the Self-Regulation Questionnaire (SRQ-A; Ryan & Connell, 1989). Items were adapted to the math domain. Cronbach’s Alpha ranged from α = .87 to α = .90 across measurement points. All reliability coefficients as well as item wordings are provided in Table B1 in Appendix B.

Grades in math. Students’ mathematics grades were obtained from school midterm records in Grade 5 (for T2), Grade 6 (for T3), and Grade 7 (for T4). These grades represent cumulative attainment in mathematics from the ongoing academic year until the midterm. Math grades in Germany range from 1 to 6, with higher numbers indicating lower achievement; for ease of interpretation, grades were reflected such that higher numbers represented better math attainment.

Math achievement test scores. Math achievement was assessed with standardized achievement tests (Götz, Lingel, & Schneider, 2013a, 2013b) that measured competence in arithmetic, algebra, geometry, and word problems at each time point. Four separate test versions with partly overlapping material were used (T1 test version: 26 items; T2 test version: 31 items; T3 test version: 37 items; T4 test version: 39 items). The minimum and maximum scores for the test versions ranged from 0 to 14 in the T1 test version, from 0 to 16 in the T2 test version, from 0 to 25 in the T3 test version, and from 0 to 27 in the T4 test version. To compare math achievement over the different versions of the test, we used vertical linear equating under the common item equivalent groups design (mean/mean equating; Kolen & Brennan, 2004). All items were taken into account with the same weight using an IRT 1PL-Model (Embretson & Reise, 2000) for the linking procedure.

Cognitive ability. The German adaptation (Heller & Perleth, 2000) of the Cognitive Abilities Test developed by Thorndike and Hagen (1971; Kognitiver Fähigkeitstest für 4. bis 12. Klassen; KFT 4-12+R) was used to measure students' cognitive ability at the beginning of Grade 5. The KFT 4-12+R can be used to assess verbal, numerical, and figural reasoning, and a composite score can be created. The 90-min short version of the test was used in the present investigation with a paper and pencil format. In our analyses, only the composite score (number of correct responses on all items) was utilized; internal consistency was $\alpha = .95$. The composite score was transformed into an IQ score based on the norming sample of the KFT.

Demographic variables, socioeconomic status (SES), and language background. At each wave of measurement, students reported their age, gender, and language background. As the vast majority of students reported a German language background, background was coded dichotomously (German native speaker: yes = 1, no = 0). We used parents' highest educational degree as a proxy for students' SES. During the first few days in Grade 5, students brought home questionnaires for their parents to complete at home and return within one week. Parents reported the highest educational degree achieved by the mother and the father (1 = "no school degree"; 2 = "lowest school degree"; 3 = "middle school degree"; 4 = "high school degree"; 5 = "university degree"; 6 = "PhD degree"). Based on these reports, a variable was created that reflected the highest educational degree that was achieved by at least one of the parents.

4.3.3. Data Analysis

The data were analyzed with SPSS 23 (IBM Corp., 2015) and Mplus 8 (Muthén & Muthén, 1998-2017). In all analyses with Mplus, we used the maximum likelihood estimator with robust standard errors (MLR; Kaplan, 2009) and handled missing data with the full information maximum likelihood (FIML) approach. At any time of measurement, 31% to 100% of the students provided data to estimate variances or covariances (covariance coverage: mastery goal items: 56-89%; PAP goal items: 57-89%; PAV goal items: 56-89%; math interest items: 57-89%; grades in math: 31-94%: math achievement test scores: 58-80%; cognitive ability: 60-82%; sex: 62-100%; parents' highest educational degree: 52-71%; language background: 61-97%). To account for the nested data structure ("students within classes"), the "type is complex" option was applied.

Attrition. To test whether dropout affected the study variables, participants were divided into three groups by the structure of their missing values: 1 = students dropped out after T1, T2, or T3; 2 = students joined the study at T2, T3, or T4; 3 = students participated at T1 and T4). Students who joined later than T1 but dropped out before T4 were assigned to

Group 1. We calculated Cohen's *d* for mean differences between the three groups for all study variables.

Factor structure and invariance. Three confirmatory factor analyses (CFA) were conducted to test whether a model with mastery, PAP, and PAV goals as three correlated factors (*trichotomous model*) showed a better fit to the data than a two correlated-factor model (*dichotomous model*; Factor 1: all mastery goal items; Factor 2: all PAP and PAV goal items) or a one-factor model (all goal items loaded on a single factor). McDonald's Omega reliability coefficients (McNeish, 2017) for each goal factor at each time point were conducted based on the *trichotomous models* at the respective time points. For interest, McDonald's Omega coefficients were conducted based on CFAs that were additionally estimated for each measurement point. Next, we tested measurement invariance over the four measurement points for each scale assessing mastery, PAP, and PAV goals and for the math interest scale. Using one autocorrelative SEM per scale, we successively tested increasing measurement invariance levels against each other (i.e., configural, metric, and scalar factorial invariance). We used effect coding (Little, Slegers, & Card, 2006) to identify the conducted models: The average factor loadings were constrained to 1 and the average item intercepts were constrained to 0. Latent factor means and standard deviations calculated by effect coding can be interpreted as the average mean of the indicators weighted by their factor-loadings (Little, Preacher, Selig, & Card, 2007). Following Chen's (2007) recommendation, in addition to χ^2-difference tests, we used the difference between comparative fit indices (ΔCFI) to compare two models. ΔCFI values of .01 or less were interpreted as a tolerable deterioration in model fit.

Development of achievement goals, interest, and achievement. To test the development of mastery, PAP, and PAV goals, and math interest, latent growth curve models (LGCMs) were calculated for each construct separately. For this purpose, we used the latent factor scores of each construct from autocorrelative SEM models with (partial) scalar measurement invariance over time as manifest indicators in the LGCMs. For each construct, we conducted a linear LGCMs and an unspecified LGCM, which allows detection of a nonlinear shape of the growth trajectory in the data (Wang & Wang, 2012), and tested these models against each other based on the Chi-square difference tests and on differences in CFI values. In the linear LGCMs, the factor loadings of the latent intercept factor were fixed to 1 and the factor loadings of the linear slope factor were fixed to .00, 2.00, 5.50, and 8.50 for T1, T2, T3, and T4, respectively, to account for the different interval durations between measurement points. In the unspecified LGCMs, the factor loadings of the latent intercept factor were fixed to 1, the factor loadings of the slope factor for T1 and T4 were fixed to 0 and 8.5 as in the linear model, and the factor loadings of the slope factor for T2 and T3 were freely estimated. The freely estimated factor loadings of the slope factor for T2 and T3 represent the amount of change between T1 and these time points relative to the amount of change between T1 and T4 fixed to 8.50. Based on a linear model and the interval durations between the time points, values of 2.00 and 5.50 are expected for factor loadings of the slope factor for T2 and T3, respectively. If the freely estimated factor loading of the slope factor is larger than the fixed factor loading from the linear model, the change that occurred between T1 and the considered measurement point was larger than predicted by a linear LGCM. For example, an estimated value of 4.25 of the factor loading at T2 indicates that the change that occurred between T1 and T2 was larger for this time interval than predicted by the linear LGCM (i.e., 4.25 is bigger than 2.00). More specifically, an estimated value of 4.25 of the factor loading at T2 indicates that half of the overall change occurred between T1 and T2 (i.e., 4.25 is half of the factor loading of 8.50 for T4).

We used comparable LGCMs for the achievement test scores. However, here we did not use latent factor scores as indicators, but rather person parameters estimated by weighted maximum likelihood estimation (Warm, 1989) within the 1PL-Model of the linking procedure. Math grades were only available from T2 on. Furthermore, measurement invariance of grades over time could not be tested which is a prerequisite for conducting a LGCM. Therefore, we used a latent change score model to estimate the change in grades between the second (T2) and the last measurement point (T4).

Joint development of achievement goals, interest, and achievement. To investigate whether change in one achievement goal (i.e., mastery, PAP, or PAV) is associated with change in academic interest and achievement, we used multiple processes correlated change structural equation models (SEMs) (Allemand & Martin, 2016). For illustrative purposes, the SEM for mastery goals is depicted in Figure 5 (the SEMs for PAP and PAV goals are depicted in Figures S3 and S4 in Appendix B). Of note, in each SEM, LGCMs for one achievement goal, interest, and test achievement were included simultaneously. For each construct, we chose a linear or unspecified LGCM, depending on the results for the appropriate form found in the previous LGCM analyses. All intercepts and slopes were allowed to correlate. The difference score of math grades, taken from the previously conducted latent change score model, was additionally included in the SEM and allowed to correlate with all intercepts and slopes. Correlations between slopes or between a slope and the difference score of math grades represent the correlated change of the two considered constructs. A positive correlation of the slopes indicates that students who show positive development in one construct are also more likely to show positive development in another construct. Correlations between intercepts and slopes (or the difference score of math grades) represent the relation between the initial level (i.e., mean-level at T1) and change in the same or another construct.

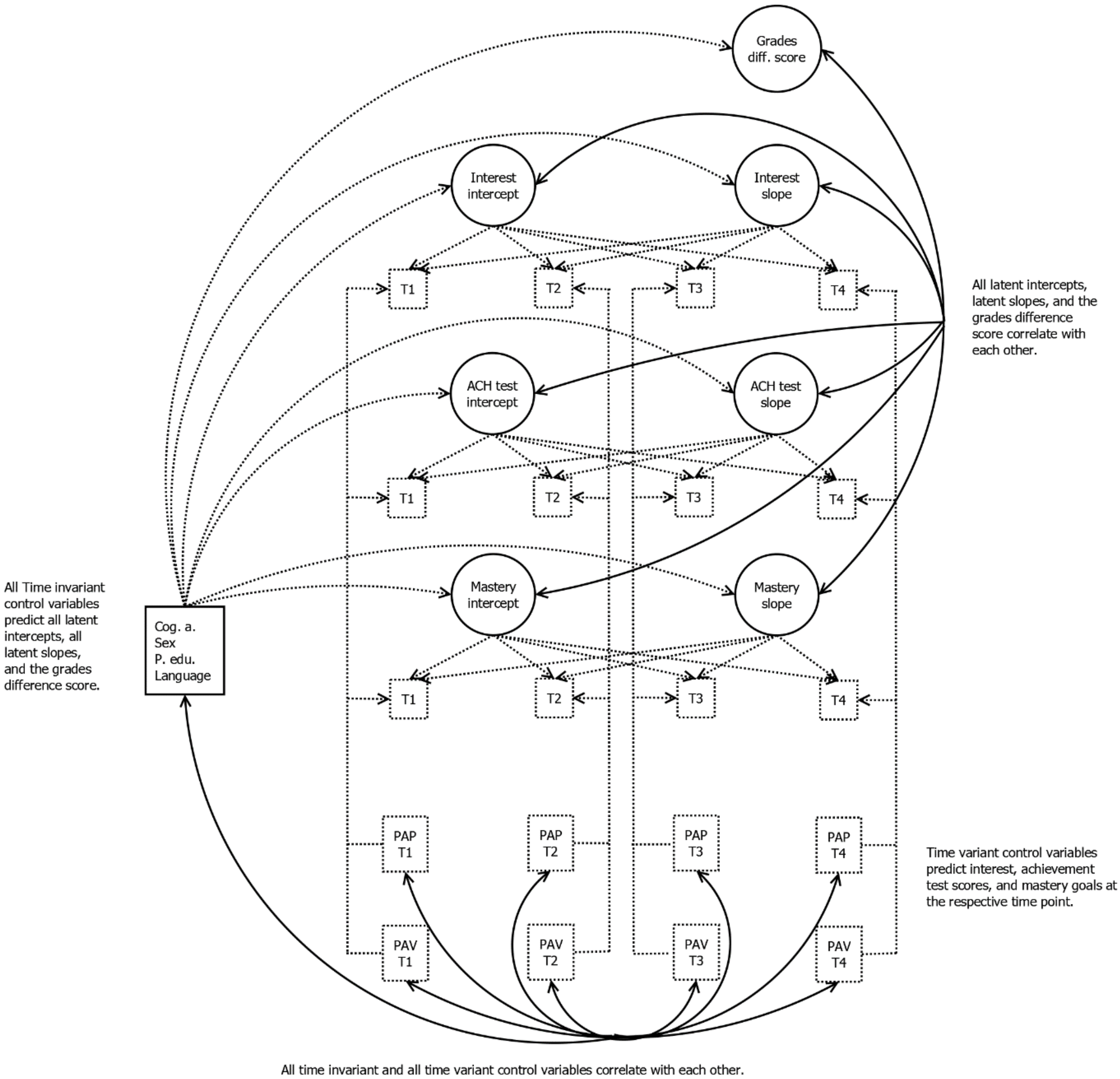


Figure 5. Correlated change of mastery goals, interest, achievement test scores, and grades controlled for students' cognitive ability, sex, parents' highest educational degree, and language background as time invariant predictors and for PAP and PAV goals as time variant predictors in Study 1. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. ACH = math achievement test score. Grades diff. score = latent change score for grades in math. Correlations are depicted as continuous lines. Regression paths are depicted as dotted lines.

Students' cognitive ability, sex, parents' highest educational degree, and language background were included as time invariant control variables (i.e., these variables were allowed to predict all intercepts, slopes, and the difference score of math grades). Furthermore, the T1, T2, T3, and T4 factor scores of the goals other than the focal goal were included as time variant control variables (e.g., in the SEM for mastery goals, we controlled for PAP and PAV goals). These variables were allowed to predict the T1, T2, T3, or T4 factor scores of the focal goal, the factor scores of interest, and the person parameters for the achievement test at the respective time points. By including these control variables, we investigated the incremental relations between the change in one achievement goal construct and interest or achievement over and above the remaining goals from the trichotomous framework (and over and above students' cognitive ability, sex, parents' highest educational degree, or language background).

As a complementary analysis, we tested the direction of the longitudinal relations between the achievement goals and each outcome variable within a reciprocal effects model (REM) over all four points of measurement. The REM included one outcome variable (i.e., interest, achievement test score, or math grades) and all achievement goals. In addition, students' cognitive ability, sex, parents' highest educational degree, and language background were included as time invariant control variables. For the purpose of illustration, the REM for interest as the outcome variable is depicted in Figure B5 in Appendix B.

4.4. Results Study 1

4.4.1. Preliminary Analyses

Correlations for the investigated variables are reported in Table 7 and descriptive statistics are reported in Table 8.

Attrition. Missing values at each time point are reported in Table B2 in Appendix B. Effect sizes for mean differences (Cohen's *d*) between attrition groups for all study variables are reported in Table B3 in Appendix B. There were a few significant differences ($d > .20$). Students who dropped out had lower cognitive ability, performed worse on the achievement test at T1, T2, and T3, and had worse grades at T2 and T3 compared to students who reported T1 and T4 values. Students who joined the study at a later time point had lower mastery goals at T4, but also better grades than students who participated at T1 and T4.

Table 7. *Correlations of the Investigated Variables in Study 1*

	M T1	M T2	M T3	M T4	PAP T1	PAP T2	PAP T3	PAP T4	PAV T1	PAV T2	PAV T3	PAV T4	INT T1	INT T2	INT T3	INT T4	ACH T1	ACH T2	ACH T3	ACH T4	Grades T2	Grades T3	Grades T4	Cog. a.	Sex
M T1	1																								
M T2	.44*	1																							
M T3	.24*	.44*	1																						
M T4	.29*	.36*	.46*	1																					
PAP T1	.29*	.20*	.19*	.20*	1																				
PAP T2	.17*	.37*	.20*	.20*	.54*	1																			
PAP T3	.06	.18*	.40*	.27*	.41*	.50*	1																		
PAP T4	.12*	.17*	.29*	.40*	.40*	.49*	.59*	1																	
PAV T1	.29*	.16*	.16*	.21*	.71*	.39*	.29*	.27*	1																
PAV T2	.20*	.39*	.26*	.22*	.47*	.75*	.45*	.39*	.44*	1															
PAV T3	.08	.23*	.43*	.27*	.39*	.43*	.78*	.50*	.36*	.48*	1														
PAV T4	.13*	.19*	.33*	.48*	.36*	.46*	.57*	.87*	.33*	.42*	.54*	1													
Int. T1	.54*	.28*	.19*	.23*	.28*	.23*	.14*	.20*	.23*	.16*	.17*	.18*	1												
Int. T2	.32*	.58*	.32*	.25*	.24*	.39*	.24*	.19*	.14*	.33*	.23*	.15*	.48*	1											
Int. T3	.19*	.31*	.58*	.32	.21	.21	.42*	.31	.15*	.21*	.38*	.32*	.37*	.50*	1										
Int. T4	.25*	.21*	.38*	.56*	.16*	.11*	.28*	.44*	.13*	.13*	.25*	.39*	.34*	.38*	.54*	1									
ACH T1	.03	.10*	.06	.10*	0	.03	.14*	.08	.01	.08	.11*	.07	.11*	.21*	.14*	.19*	1								
ACH T2	.06	.05	.04	.03	.01	.08	.08	.04	.01	.12*	.07	.03	.17*	.21*	.16*	.19*	.55*	1							
ACH T3	.10*	.09	.08	.12*	.12*	.02	.08	.14*	.03	.11*	.11*	.12*	.24*	.21*	.20*	.24*	.49*	.49*	1						
ACH T4	.04	.05	.08	.17*	.04	.07	.09	.17*	.03	.07	.04	.16*	.20*	.19*	.17*	.25*	.47*	.45*	.59*	1					
GradesT2	.02	.11*	.09	.10	.08	.10	.09	.17*	.07	.09	.09	.17*	.13*	.30*	.14*	.18*	.43*	.22*	.32*	.35*	1				
GradesT3	.04	.09	.10	.13	.07	.00	.14*	.15*	.03	.02	.07	.16*	.10	.22*	.17*	.19*	.40*	.25*	.41*	.37*	.64*	1			
GradesT4	.03	.06	.10*	.15*	.03	.04	.12*	.18*	-.02	.03	.08	.18*	.12*	.20*	.17*	.29*	.40*	.25*	.48*	.43*	.46*	.65*	1		
Cog. a.	.03	.05	.07	.08	-.01	-.01	.11*	.08	0	.07	.08	.08	.05	.15*	.11*	.17*	.49*	.44*	.44*	.44*	.37*	.38*	.36*	1	
Sex.	-.02	.00	.01	.07	-.13*	-.12*	-.09*	-.09*	-.13*	-.11*	-.12*	-.06	-.20*	-.13*	-.16	-.13*	-.19*	-.19*	-.11*	-.14*	-.13*	-.02	-.02	-.03	1
P. edu.	-.04	-.05	-.02	-.06	-.05	-.08	-.02	-.02	-.01	-.07	-.09	-.02	.00	.01	.06	.07	.03	.08	.09	.07	.05	.19*	.07	.11*	.01
Language	.07	.01	.04	.02	.15*	.09*	.07	.11*	.11*	.09*	.09*	.07	.05	.03	.04	-.02	-.12*	-.08	-.08	-.10*	-.02	-.13*	-.14*	-.13*	-.07

Note. **p* < .05. M = Mastery achievement goals. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals. INT = Interest in math. ACH = math achievement test score. Grades = grades in math. Cog. a. = cognitive ability score. P. edu. = Parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Correlations including grades or parents' highest educational degree were estimated by Spearman's rank order correlation. Sex: 0 = boys, 1 = girls.

Table 8. *Descriptive Statistics of the Investigated Variables in Study 1*

	n	Min	Max	*M*	*SD*	Skew	Kurtosis	ICC
M T1	605	1.00	5.00	4.55	.59	-1.71	3.78	.04
M T2	661	1.67	5.00	4.25	.76	-.89	-.03	.09
M T3	611	1.00	5.00	4.03	.84	-.70	.01	.07
M T4	531	1.00	5.00	4.07	.86	-.87	.30	.08
PAP T1	605	1.00	5.00	3.55	1.14	-.48	-.70	.07
PAP T2	662	1.00	5.00	3.31	1.12	-.19	-.78	.06
PAP T3	613	1.00	5.00	3.18	1.10	-.07	-.70	.05
PAP T4	532	1.00	5.00	3.30	1.13	-.27	-.67	.04
PAV T1	605	1.00	5.00	3.97	1.00	-.90	.13	.06
PAV T2	662	1.00	5.00	3.59	1.05	-.42	-.62	.07
PAV T3	612	1.00	5.00	3.41	1.03	-.30	-.52	.07
PAV T4	532	1.00	5.00	3.42	1.10	-.35	-.63	.06
Int. T1	605	1.00	5.00	4.05	.99	-.94	.22	.04
Int. T2	662	1.00	5.00	3.63	1.08	-.47	-.56	.08
Int. T3	613	1.00	5.00	3.40	1.08	-.25	-.66	.09
Int. T4	532	1.00	5.00	3.44	1.04	-.28	-.51	.03
ACH T1	591	-3.87	2.63	-.86	.85	.04	.43	.07
ACH T2	594	-3.65	3.36	-.55	.97	.26	.51	.22
ACH T3	508	-4.64	2.82	-.23	1.12	-.28	1.48	.14
ACH T4	459	-2.78	2.94	.07	.85	.10	.49	.12
Grades T2	364	2.00	6.00	4.69	.74	-.45	.28	.14
Grades T3	294	1.00	6.00	4.31	.92	-.39	.01	.09
Grades T4	565	1.50	6.00	4.06	.93	-.23	-.39	.05
Cog.a.	612	76.00	136.00	107.21	.90	-.23	.35	.05

Note. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Int. = Interest in math. ACH = math achievement test score. Grades = grades in math. Cog. a. = cognitive ability score. ICC = intraclass correlation between the school classes.

Factor structure and invariance. Results from the CFAs indicated that at all measurement points a three-factor model (mastery vs. PAP vs. PAV) fit the data better than a two-factor model (mastery vs. PAP and PAV) or a one-factor model (for fit indices, see Table B4 in Appendix B). McDonald’s Omega for each goal ranged from $\omega = .75$ to $\omega = .93$ across measurement points (see Table B1 in Appendix B). Results of measurement invariance tests for all scales are presented in Table B5 in Appendix B. We found scalar measurement invariance for mastery goals, PAV goals, and interest. For PAP goals, we had to set free the intercept of the first item at T1 to reach adequate fit (i.e., partial scalar measurement invariance).

4.4.2. Main Analyses

Development of achievement goals, interest, and achievement. For each achievement goal and interest, the unspecified LGCM fit the data better than the linear LGCM (ΔCFI < .01; see Table B6 in Appendix B for fit indices). In the unspecified LGCM for mastery and PAV goals, we observed a non-significant negative residual in T1 factor scores, which we fixed to zero (Chen, Bollen, Paxton, Curran, & Kirby, 2001). For math test scores, the unspecified LGCM did not show a better fit than the linear LGCM.

Model coefficients of the LGCMs are reported in Table 9 and the developmental trajectories for the achievement goals and interest based on factor means are depicted in Figure 6. The developmental trajectory for achievement test scores based on person parameters estimated by weighted maximum likelihood is depicted in Figure B6 in Appendix B. All three achievement goals and interest showed a significant decrease (mastery goal slope: $b = -.06$, $p < .001$; PAP goal slope: $b = -.02$, $p < .001$; PAV goal slope: $b = -.06$, $p < .001$; interest slope: $b = -.07$, $p < .001$). These decreases were not linear, as indicated by the better fit of unspecified LGCMs compared to linear LGCMs. For each achievement goal and interest, the estimated factor loadings of the slope factor for T2 and T3 were noticeably larger than expected according to a linear LGCM (T2: $b = 2.00$; T3: $b = 5.50$). In fact, about half of the decline occurred between T1 and T2 and the other half occurred between T2 and T3, whereas the decreases stopped and even reversed thereafter. For math test scores, the linear growth curve model indicated a continuous increase over time (slope: $b = .10$, $p < .001$). Finally, a latent change score model of math grades based on T2 and T4 assessment of grades indicated a significant decrease in math grades ($p = .002$). Frequencies of slope estimates are depicted in Figure B7 in Appendix B. Note that 83.67% of the mastery goal slopes, 65.99% of the PAP goal slopes, 78.81% of the PAV goal slopes, 84.89% of the interest slopes, and 2.06% of the achievement test score slopes were negative, indicating a decrease in students' achievement goals and interest and an increase in students' test achievement over time.

Table 9. *Model Parameters and Fit Indices of the Latent Growth Curve Models in Study 1*

Parameters	Mastery unspecified LGCM			PAP unspecified LGCM			PAV unspecified LGCM			Interest unspecified LGCM			ACH linear LGCM		
	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*
Means															
Intercept	4.53	.02	< .001	3.45	.06	< .001	3.96	.04	< .001	4.05	.04	< .001	-.84	.06	< .001
Slope	-.06	.00	< .001	-.02	.00	< .001	-.06	.00	< .001	-.07	.00	< .001	.10	.01	< .001
Variance															
Intercept	.17	.02	< .001	.88	.16	< .001	.60	.05	< .001	.69	.08	< .001	.49	.07	< .001
Slope	.00	.00	< .001	.01	.00	= .035	.01	.00	< .001	.01	.00	< .001	.01	.00	< .001
Factor loadings															
slope on T1	.00 fix	-	-	.00 fix	-	-	.00 fix	-	-	.00 fix	-	-	.00 fix	-	-
slope on T2	4.84*	.34	< .001	5.48*	.34	< .001	5.66*	.31	< .001	5.70*	.44	< .001	2.00 fix	-	-
slope on T3	9.27*	.43	< .001	9.97*	.43	< .001	8.93*	.26	< .001	8.89*	.35	< .001	5.50 fix	-	-
slope on T4	8.50 fix	-	-	8.50 fix	-	-	8.50 fix	-	-	8.50 fix	-	-	8.50 fix	-	-
Covariances															
I with S	-.01	.00	< .001	-.04	.02	= .049	-.03	.01	< .001	-.04	.01	< .001	-.02	.01	= .029
Correlations															
I with S	-.31	.06	< .001	-.58	.11	< .001	-.50	.05	< .001	-.57	.06	< .001	-.32	.11	= .004
Fit indices															
	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*
	.969	24.73 (4)	< .001	.969	43.52 (3)	< .001	.988	28.48 (4)	< .001	.934	58.089 (3)	< .001	.986	9.185 (5)	= .102

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. ACH = math achievement test score. LGCMs were conducted separately for mastery, PAP, and PAV goals, interest and math achievement test score. For mastery, PAP, and PAV goals and interest, unspecified LGCMs are presented. For the math achievement test score a linear LGCM is presented. In linear model, the factor loadings of the slope factor were fixed to 0, 2, 5.5, and 8.5 for T1, T2, T3, and T4, respectively. In the unspecified models, the factor loadings of the slope factor were fixed to 0 and 8.5 for T1 and T4 respectively, whereas factor loadings for T2* and T3* were freely estimated. In the mastery and the interest LGCMs we fixed the residual variance of the T1 factor score to zero as the T1 factor score showed a slightly negative, non-significant, residual variance in the unspecified models.

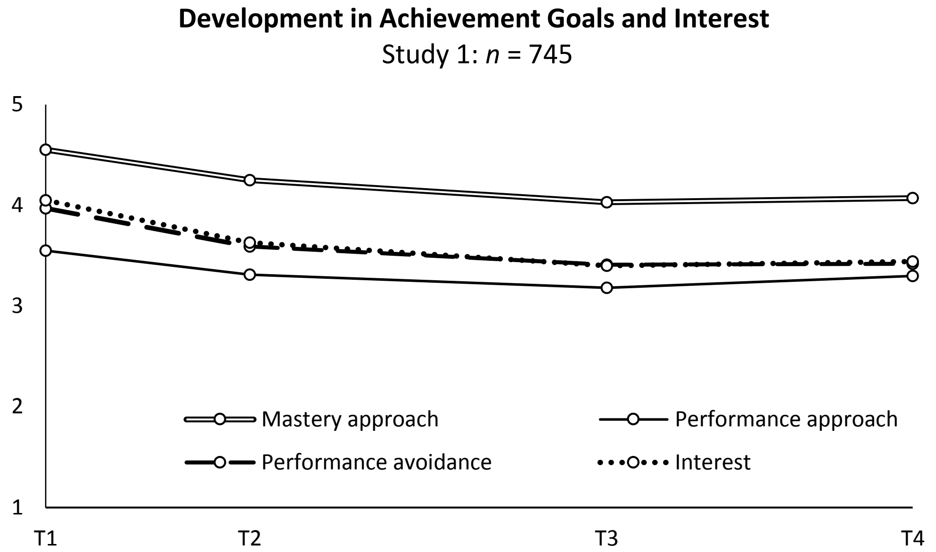


Figure 6. Developmental trajectories of achievement goals and interest based on factor means in Study 1.

Joint Development of Achievement Goals, Interest, and Achievement. Correlation coefficients between the initial levels (intercepts) and change scores (slopes and math grade difference score) of achievement goals, interest, achievement test scores, and math grades are presented in Tables 10a, 10b, and 10c for mastery, PAP, and PAV goals as focal goal, respectively. Regression coefficients of the time variant control variables are reported in Tables B7a, B7b, and B7c in Appendix B and correlation coefficients between the control variables are reported in Tables B8a, B8b, and B8c in Appendix B. Fit indices of the SEMs are reported in Table B9 in Appendix B.

Of central interest, the slope of mastery goals was positively related to the slope of interest (r = .51; p < .001). That is, students with larger declines in mastery goals showed larger declines in their math interest over time. Likewise, the slope of PAP goals was positively related to the slope of interest (r = .16; p = .044). That is, students with larger declines in PAP goals showed larger declines in their math interest over time. There was no significant relation between the slopes of these two goals and the slope for math test achievement and the change score for math grades. The slope of PAV goals was not significantly related to the slope of interest or test achievement and the change score for math grades.

Regarding the relations of initial level and change over time, the intercept of each construct was negatively related to the slope of the same construct. That is, students who reported higher achievement goals or interest also reported larger declines for these variables than students with lower initial levels. Similarly, the intercepts of mastery (r = -.30; p < .001) and PAP goals (*r* = -.12; *p* = .013) were negatively related to the slope of interest, and the intercept of interest was negatively related to the slope of mastery goals (*r* = -.15; *p* = .020). That is, students with initially higher mastery or PAP goals reported a larger decline in interest than students with initially lower mastery or PAP goals.

Table 10a. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of Mastery Goals, Interest, and Achievement in Study 1*

	Mastery intercept	Mastery slope	Interest intercept	Interest slope	ACH intercept	ACH slope	Grades diff.
Correlations with							
Mastery slope	-.39***						
Interest intercept	.51***	-.15*					
Interest slope	-.30***	.51***	-.61***				
ACH intercept	.07	.01	.16*	-.02			
ACH slope	.02	.07	.14*	-.08	-.34*		
Grade diff. score	.03	.05	.06	.03	.14*	.19*	
Regressions on							
Cog. a.	.04	.06	.06	.08	.60***	-.06	.11**
Sex	.03	.10	-.19***	.07	-.21**	.12	.06
P. edu.	-.03	.03	.02	.05	-.05	.17***	.02
Language b.	.00	.01	-.02	.02	.10**	-.03	.10

Note. **p* < .05. ***p* < .01. ****p* < .001. ACH = math achievement test score. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. The residual variances of T1 mastery goals factor score and the T1 interest factor score were fixed to zero as these factor scores showed slightly negative, non-significant, residual variances at T1.

Table 10b. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of PAP Goals, Interest, and Achievement in Study 1*

	PAP intercept	PAP slope	Interest intercept	Interest slope	ACH intercept	ACH slope	Grades diff.
Correlations with							
PAP slope	-.82***						
Interest intercept	.20***	-.08					
Interest slope	-.12*	.16*	-.35**				
ACH intercept	-.02	.07	.19**	-.02			
ACH slope	.03	.02	.12	-.05	-.34*		
Grade diff. score	.05	-.01	.05	.08	.14*	.18*	
Regressions on							
Cog. a.	.02	.01	.10*	.04	.60***	-.07	.11**
Sex	-.05	.00	-.26***	.03	-.22**	.10	.06
P. edu.	-.07	.11*	.05	.03	-.05	.18***	.02
Language b.	-.06	.01	-.03	.02	.10**	-.05	.10

Note. **p* < .05. ***p* < .01. ****p* < .001. PAP = Performance-approach goals. ACH = math achievement test score. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. The residual variance of the T1 PAP goals factor score was fixed to zero as the T1 factor score showed a slightly negative, non-significant, residual variance.

Table 10c. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of PAV Goals, Interest, and Achievement in Study 1*

	PAV intercept	PAV slope	Interest intercept	Interest slope	ACH intercept	ACH slope	Grades diff.
Correlations with							
PAV slope	-.49***						
Interest intercept	-.08	.06					
Interest slope	.04	-.07	-.33***				
ACH intercept	.01	-.10	.19**	-.02			
ACH slope	-.09	.10	.11	-.06	-.34*		
Grade diff. score	-.09	.07	.05	.07*	.14*	.18*	
Regressions on							
Cog. a.	.06	-.09	.11**	.03	.60***	-.06	.11**
Sex	-.11*	.11*	-.26***	.04	-.22**	.11	.06
P. edu.	-.02	-.06	.06	.02	-.05	.17***	.02
Language b.	-.07	.10	-.03	.04	.10**	-.03	.10

Note. **p* < .05. ***p* < .01. ****p* < .001. PAV = Performance-avoidance goals. ACH = math achievement test score. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. The residual variance of the T4 interest factor score was fixed to zero as the T4 factor score showed a slightly negative, non-significant, residual variance.

Results for the REM model. Regression coefficients for the REMs, including all achievement goals and one outcome variable (i.e., interest, achievement test scores, and math grades) are reported in Tables B10a, B10b, and B10c in Appendix B. In these analyses, the cross-lagged relations between the goals and the outcomes are of particular interest. Out of 24 possible cross-lagged effects from achievement goals to interest, test achievement, and grades, 3 effects reached significance. T1 mastery goals positively predicted T2 interest (β = .11; p = .009), and T1 PAP goals positively predicted T2 interest (β = .09; p = .038). T2 mastery goals positively predicted T3 grades (β = .12; p = .005). Out of 24 possible cross-lagged effects from interest, test achievement, and grades to achievement goals, 6 effects reached significance. T1 test achievement positively predicted T2 mastery goals (β = .11; p = .023) and T3 grades positively predicted T4 mastery goals (β = .15; p = .014). T1 interest positively predicted T2 PAP goals (β = .10; p = .013) and T3 grades positively predicted T4 PAP goals (β = .10; p = .013). T2 interest positively predicted T3 PAV goals (β = .07; p = .022) and T3 grades positively predicted T4 PAV goals (β = .09; p = .033).

4.5. Methods Study 2

4.5.1. Participants and procedure

In Study 2, we conducted a similar series of analyses as in Study 1 in a different, larger data set collected over a longer time interval. In this data set a different achievement goal measure was used. We obtained four measurement waves of the "AVG" project, an ongoing longitudinal study focusing on the development of student motivation upon entering

Gymnasium. The “AVG” project was approved by the Supervision and Services Directorate of Rhineland-Palatinaten (“Aufsichts- und Dienstleistungsdirektion”; protocol number: 32-03 405/29/05). The initial sample comprised 1477 students (744 boys, 661 girls, 72 missing information on sex; *M* age at T1 in years = 10.58 *SD* = 0.51). 57 students who provided no data on all investigated variables at all measurement points, except for demographics, were excluded from analyses. Students attended five schools in two federal states of Germany (i.e., Rhineland-Palatinate = 1217; Bavaria = 203). Regarding parents’ highest educational degree, 0.20% had no school degree, 3.36% had a certificate of secondary education (“Hauptschulabschluss”), 19.75% had a general certificate of secondary education (“Realschulabschluss”), 19.75% had a general qualification for university entrance (“Abitur”), 41.18% had a university degree, and 15.76% of the students’ parents had a PhD. Regarding language background, 86.73% of students were native German speakers. Data collection started with the 2005/2006 school year (i.e., Fall 2005).

Students were tested in groups in their classrooms at the beginning of Grade 5 (T1), in the middle of Grade 5 (T2), in the middle of Grade 6 (T3), and in the middle of Grade 8 (T4). On average, the measurements took place 60 days (T1), 173 days (T2), 544 days (T3), and 1265 days (T4) after entry into Gymnasium. This longitudinal study is ongoing; *a priori* we decided to use the data for T1 to T4 in the analyses, as in Study 1. Trained research assistants carried out the data collection. Participation was voluntarily and written parental consent was obtained for all students. As in Study 1, there were no experimental manipulations, all data exclusions and variables analyzed are reported, and data collection was completed before any analyses were conducted.

4.5.2. Materials

Achievement goals. Mastery goals, PAP goals, and PAV goals for math were assessed with three items each from the achievement goal measure used in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007). Students responded to the items on a 1 (“not true”) to 5 (“true”) scale. Cronbach’s Alpha ranged from $\alpha = .72$ to $\alpha = .84$ across measurement points. Item wording and reliability coefficients are provided in Table B11 in Appendix B.

Math interest. Interest in mathematics was assessed with three items (e.g., “I am interested in math.”; answered on a 5-point rating scale: “not true” to “true”) from the PALMA Project (Pekrun et al., 2006). Cronbach’s Alpha ranged from $\alpha = .79$ to $\alpha = .82$ across measurement points. Item wording and reliability coefficients are provided in Table B11 in Appendix B.

Grades in math. Students self-reported their mathematics grades from their latest midterm record one to two weeks after having received the record in Grade 5 (T2), Grade 6 (T3), and Grade 8 (T4). At T1, students had not received any grades yet. As noted earlier, math grades in Germany range from 1 to 6, with higher numbers indicating lower achievement; for ease of interpretation, grades were reflected such that higher numbers represented better math attainment.

Cognitive ability. Cognitive ability was assessed in the first term of Grade 5 (about 3 month after the school start) with the same measure used in Study 1 (the KFT 4-12+R; Heller & Perleth, 2000). Internal consistency of the measure score was α = .92. Row values were transformed into IQ scores based on the norming sample of the KFT.

Demographic variables, SES, and language background. At the first wave of measurement, students reported their age, gender, and language background in a questionnaire. As the vast majority of students reported a German language background (86.73%), language background was coded dichotomously (native German speaker: yes = 1, no = 0). Students further reported the highest educational degree of their parents as a proxy for SES (mother and father; 1 = "no school degree"; 2 = "lowest school degree"; 3 = "middle school degree"; 4 = "high school degree"; 5 = "university degree"; 6 = "PhD degree"). We created a variable that reflected the highest educational degree that was achieved by at least one of the parents.

4.5.3. Data Analysis

The data were analyzed with the same software and the same approaches as in Study 1. At any wave of measurement, 26% to 100% of the students provided data to estimate variances or covariances (covariance coverage: mastery goal items: 49-72%; PAP goal items: 49-72%; PAV goal items: 49-72%; math interest items: 49-72%; grades in math: 55-79%; cognitive ability: 26-67%; sex: 67-100%; parents' highest educational degree: 26-34%; language background: 30-77%).

Attrition. We conducted the same attrition analyses as in Study 1.

Factor structure and invariance. We conducted the same analyses for testing the factor structure and measurement invariance over time as in Study 1.

Development of achievement goals, interest and, achievement. We conducted the same analyses testing the development of achievement goals and interest as in Study 1. In the linear LGCMs, the factor loadings of the latent intercept factor were fixed to 1 and the factor loadings of the linear slope factor were fixed to .00, 0.50, 1.50, and 3.50 for T1, T2, T3, and T4, respectively, to account for the different interval durations between measurement points.

In the unspecified LGCMs, the factor loadings of the slope factor were fixed to 1, the factor loadings of the slope factor for T1 and T4 were fixed to .00 and 3.50 as in the linear model, and the factor loadings for the slope factor for T2 and T3 were freely estimated.

Joint development of achievement goals, interest, and achievement. We conducted the same analyses testing the development and correlated change of achievement goals, interest, and achievement (i.e., math grades) as in Study 1, except that achievement tests scores were not included because achievement test data were not available in Study 2. Further, we conducted the same REMs as in Study 1 to test the direction of the longitudinal relations between the achievement goals and interest and grades.

4.6. Results Study 2

4.6.1 Preliminary Analyses

Correlations for the investigated variables are reported in Table 11 and descriptive statistics are reported in Table 12.

Attrition. Missing values at each time point are reported in Table B12 in the supplementary materials. Effect sizes Cohen's *d* for mean differences in all study variables between attrition groups are reported in Table B13 in Appendix B. There were a few significant differences ($d > .20$). Students who dropped out had lower cognitive ability and had worse grades at T2 and T3 compared to students who reported T1 and T4 values. Furthermore, more boys than girls dropped out. Compared to students who joined the study at a later time point students who dropped out reported higher PAV goals and worse grades at T3. Students who joined the study at a later time point, reported higher PAP goals and lower PAV goals at T3 than students who participated at T1 and T4.

Table 11. *Correlations of the Investigated Variables in Study 2*

	M T1	M T2	M T3	M T4	PAP T1	PAP T2	PAP T3	PAP T4	PAV T1	PAV T2	PAV T3	PAV T4	INT T1	NT T2	NT T3	INT T4	Grades T2	Grades T3	Grades T4	Cog. a.	Sex	P. edu.
M T1	1																					
M T2	.58*	1																				
M T3	.43*	.51*	1																			
M T4	.27*	.35*	.39*	1																		
PAP T1	.39*	.31*	.17*	.18*	1																	
PAP T2	.32*	.41*	.26*	.23*	.66*	1																
PAP T3	.28*	.31*	.41*	.25*	.53*	.63*	1															
PAP T4	.14*	.17*	.17*	.47*	.32*	.40*	.43*	1														
PAV T1	.47*	.32*	.22*	.12*	.61*	.47*	.38*	.14*	1													
PAV T2	.36*	.47*	.25*	.15*	.47*	.54*	.40*	.17*	.63*	1												
PAV T3	.25*	.24*	.39*	.18*	.40*	.42*	.60*	.25*	.49*	.48*	1											
PAV T4	.14*	.10*	.10*	.40*	.22*	.22*	.20*	.51*	.25*	.27*	.35*	1										
INT T1	.67*	.50*	.36*	.30*	.31*	.26*	.24*	.16*	.22*	.18*	.12*	.04	1									
INT T2	.50*	.66*	.44*	.30*	.25*	.35*	.28*	.17*	.16*	.23*	.09*	.01	.64*	1								
INT T3	.35*	.41*	.67*	.36*	.18*	.20*	.36*	.16*	.11*	.11*	.17*	.00	.45*	.54*	1							
INT T4	.21*	.30*	.34*	.68*	.17*	.19*	.24*	.40*	.04	.04	.08*	.14*	.34*	.45*	.54*	1						
Grades T2	.01	.07	.04	.11	-.03	.05	.01	.07	-.13*	-.10*	-.11*	-.10*	.10*	.17*	.10*	.13*	1					
Grades T3	.08*	.14*	.16*	.17*	.01	.04	.03	.09*	-.03	-.03	-.12*	-.14*	.16*	.19*	.28*	.23*	.46*	1				
Grades T4	.06	.11*	.12*	.24*	-.03	.03	-.02	.03	-.04	-.04	-.13*	-.09*	.14*	.16*	.18*	.36*	.35*	.50*	1			
Cog.a.	-.04	-.01	.06	.06	-.07*	-.04	-.03	.03	-.10*	-.13*	-.12*	-.08*	.03	.07*	.11*	.11*	.28*	.40*	.29*	1		
Sex	-.10*	-.08*	-.02	-.09*	-.25*	-.24*	-.18*	-.16*	-.12*	-.05	-.04	.01	-.14*	-.09*	-.11*	-.16*	-.03	-.05	.03	.01	1	
P. edu.	-.13*	-.07	-.12*	.00	-.11	-.15*	-.15*	-.07	-.10	-.08	-.10	-.06	-.08	-.09	-.10	.07	.10	.09	.05	.00	-.06	1
Language	-.09*	-.06	.07*	.07*	-.13*	-.10*	-.02	-.03	-.07*	-.04	-.01	.01	-.02	-.01	.10*	.01	.08*	.08*	.02	.11*	-.03	-.07

Note. **p* < .05. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. INT = Interest in math. ACH = math achievement test score. Grades = grades in math. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Correlations including grades or parents' highest educational degree were estimated by Spearman's rank order correlation. Sex: 0 = boys, 1 = girls.

Table 12. *Descriptive Statistics of the Investigated Variables in Study 2*

	n	Min	Max	*M*	*SD*	Skew	Kurtosis	ICC
M T1	1014	1	5	3.94	0.88	-0.66	-0.03	.07
M T2	1020	1	5	3.78	0.91	-0.54	-0.08	.06
M T3	987	1	5	3.58	0.96	-0.39	-0.31	.04
M T4	970	1	5	3.17	0.90	-0.10	-0.32	.04
PAP T1	1014	1	5	2.70	1.18	0.32	-0.90	.07
PAP T2	1024	1	5	2.56	1.10	0.42	-0.61	.04
PAP T3	992	1	5	2.53	1.16	0.48	-0.70	.07
PAP T4	979	1	5	2.28	1.00	0.70	-0.03	.05
PAV T1	1014	1	5	3.37	1.12	-0.28	-0.66	.04
PAV T2	1021	1	5	3.26	1.07	-0.17	-0.64	.04
PAV T3	990	1	5	3.03	1.05	-0.03	-0.57	.05
PAV T4	973	1	5	2.81	0.97	0.21	-0.35	.02
Int. T1	1014	1	5	3.61	1.05	-0.43	-0.59	.10
Int. T2	1023	1	5	3.31	1.04	-0.13	-0.72	.08
Int. T3	993	1	5	3.19	1.06	-0.05	-0.76	.07
Int. T4	979	1	5	2.70	1.00	0.26	-0.58	.09
Grades T2	1010	2	6	4.85	0.76	-1.70	3.07	.09
Grades T3	985	1	6	4.26	0.95	-0.10	-0.37	.08
Grades T4	960	1	6	4.37	1.15	-0.72	-0.75	.08
Cog.a.	944	54	148	107.92	11.2	-0.47	1.84	.06

Note. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Int. = Interest in math. Grades = grades in math. Cog. a. = cognitive ability score. ICC = intraclass correlation between the school classes.

Factor structure and invariance. Confirmatory factor analyses indicated that at all measurement points a three-correlated-factor model (mastery goals vs. PAP goals vs. PAV goals) fit the data better than a two-correlated-factor model (mastery goals vs. a common factor for all PAP and PAV goal items), or a one-factor model (for fit indices, see Table B14 in Appendix B). McDonald's omega coefficients of each goal factor ranged from ω = .73 to ω = .95 across measurement points (see Table B11 in Appendix B).

We found partial scalar measurement invariance over time for all scales. That is, we had to set free some intercepts or factor loadings in the partial invariant models. For fit indices of the measurement invariance tests and more information about the coefficients that were set free, see Table B15 in Appendix B.

Table 13. *Model Parameters and Fit Indices of the Latent Growth Curve Models in Study 2*

Parameters	Mastery unspecified LGCM			PAP linear LGCM			PAV unspecified LGCM			Interest linear LGCM		
	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*
Means												
Intercept	4.00	.03	< .001	2.63	.04	< .001	3.34	.03	< .001	3.58	.04	< .001
Slope	-.24	.01	< .001	-.07	.01	< .001	-.15	.01	< .001	-.25	.01	< .001
Variance												
Intercept	.35	.02	< .001	.76	.03	< .001	.58	.02	< .001	.56	.03	< .001
Slope	.02	.00	< .001	.03	.00	< .001	.03	.00	< .001	.03	.00	< .001
Factor loadings												
slope on T1	.00 fix	-	-	.00 fix	-	-	.00 fix	-	-	.00 fix	-	-
slope on T2	.87*	.05	< .001	.50 fix	-	-	.52*	.07	< .001	.50 fix	-	-
slope on T3	1.72*	.08	< .001	1.50 fix	-	-	1.93*	.08	< .001	1.50 fix	-	-
slope on T4	3.50 fix	-	-	3.50 fix	-	-	3.50 fix	-	-	3.50 fix	-	-
Covariances												
I with S	-.05	.00	< .001	-.11	.01	< .001	-.09	.01	< .001	-.07	.01	< .001
Correlations												
I with S	-.55	.02	< .001	-.67	.03	< .001	-.67	.02	< .001	-.53	.02	< .001
Fit indices												
	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*	CFI	χ^2 (df)	*p*
	.942	97.99 (3)	< .001	.941	177.80 (5)	< .001	.986	35.30 (3)	< .001	.962	63.85 (5)	< .001

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. LGCMs were conducted separately for mastery, PAP, and PAV goals, and interest a. For, mastery and PAV goals, unspecified LGCMs are presented. For PAV goals and interest linear LGCMs are presented. In linear models, the factor loadings of the slope factor were fixed to 0, .5, 1.5, and 3.5 for T1, T2, T3, and T4, respectively. In the unspecified models, the factor loadings of the slope factor were fixed to 0 and 3.5 for T1 and T4 respectively, whereas factor loadings for T2* and T3* were freely estimated.

4.6.2 Main Analyses

Development of achievement goals and interest. The fit indices of the linear and unspecified LGCMs for mastery goals, PAP goals, PAV goals, and interest are presented in Table B16 in Appendix B. The unspecified LGCM fit the data better than the linear LGCM for mastery and PAV goals (ΔCFI < .01). For PAP goals and interest, there was no significant difference in model fit of the linear LGCM and the unspecified LGCM, indicating that the linear model adequately described the change in PAP goals and interest over time.

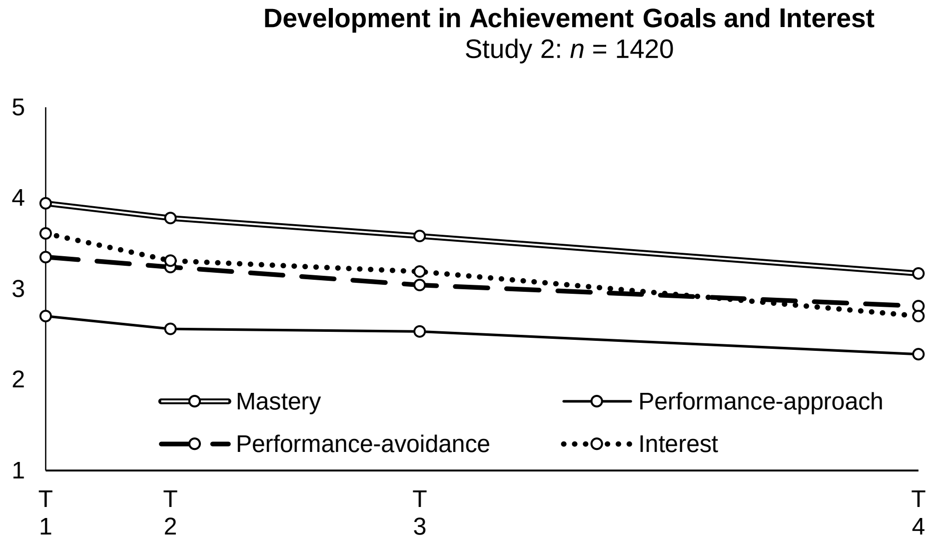


Figure 7. Developmental trajectories of achievement goals and interest based on factor means in Study 2.

Model coefficients of the LGCMs are reported in Table 13 and the developmental trajectories based on means are depicted in Figure 7. All achievement goals and interest showed a significant decrease over time (mastery goal slope: b = -.24, p < .001; PAP goal slope: b = -.07, p < .001; PAV goal slope: b = -.15, p < .001; interest slope: b = -.25, p < .001). The decrease in mastery and PAV goals was not linear. For mastery goals, the factor loadings from the unspecified LGCM of the slope factor for T2 and T3 were larger than could be expected based on a linear LGCM (T2: b = .87 vs. b = .50; T3: b = 1.72 vs. b = 1.50). Thus, the mastery goal decrease was particularly strong between T1 and T2. For PAV goals, the estimated factor loading of the slope factor for T2 was nearly the same as expected according to the linear LGCM (b = .52 vs b = .50). However, the factor loading of the slope factor for T3 was larger than expected (b = 1.9 vs. b = 1.5). Thus, the PAV goal decline was particularly strong between T2 and T3. As in Study 1, a latent change score model for math grades based on the first and the last available time points indicated a significant decrease in math grades ($p < .001$). Frequencies of slope estimates are depicted in Figure B8. Note that 96.73% of the mastery goal slopes, 67.97% of the PAP goal slopes, 84.60% of the PAV goal slopes, and 95.31% of interest slopes were negative, indicating a decrease in students' achievement goals and interest over time.

Joint development of achievement goals, interest, and achievement. Regression coefficients of the time variant control variables are presented in Tables B17a, B17b, and B17c in Appendix B and correlation coefficients between the control variables are presented in Tables B18a, B18b, and B18c in Appendix B. Fit indices of the SEMs are reported in Table B19 in Appendix B.

Of central interest, the slope of mastery goals was positively related to the slope of interest ($r = .74$; $p < .001$) and the difference score of grades ($r = .14$; $p < .001$). That is, students with larger declines in mastery goals showed larger declines in math interest and math grades over time. The slope of PAP goals was positively related to the slope of interest ($r = .14$; $p < .001$), but was not significantly related to the difference score of grades. That is, students with larger declines in PAP goals showed larger declines in math interest, but not in math grades over time. The slope of PAV goals was negatively related to the slope of interest ($r = -.14$; $p < .001$) and the difference score of grades ($r = -.08$; $p = .008$). That is, students with larger declines in PAV goals showed larger increases in math interest and math grades over time.

Table 14a. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of Mastery Goals, Interest, and Achievement in Study 2*

	Mastery intercept	Mastery slope	Interest intercept	Interest slope	Grades diff.
Correlations with					
Mastery slope	-.51***				
Interest intercept	.77***	-.39***			
Interest slope	-.46***	.74***	-.55***		
Grade diff. score	.07*	.14***	.07	.21***	
Regressions on					
Cog. a.	.05	.03	.10*	-.01	.10**
Sex	-.04	-.02	-.05	-.02	.04
P. edu.	-.06	.09	-.05	.13*	.01
Language b.	-.03	.09**	.03	-.01	-.03

Note. *$p < .05$. **$p < .01$. ***$p < .001$. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background.

Table 14b. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of PAP Goals, Interest, and Achievement in Study 2*

	PAP intercept	PAP slope	Interest intercept	Interest slope	Grades diff.
Correlations with					
PAP slope	-.62***				
Interest intercept	.14***	-.05			
Interest slope	-.04	.14***	-.53***		
Grade diff. score	-.04	.02	.04	.18***	
Regressions on					
Cog. a.	.01	.02	.11*	-.03	.11**
Sex	-.30***	.16***	-.09*	-.01	.03
P. edu.	-.21***	.16**	-.10	.13*	.00
Language b.	-.13***	.09*	.03	-.05	-.03

Note. **p* < .05. ***p* < .01. ****p* < .001. PAP = Performance-approach goals. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. The residual variance of the T2 interest factor score and the residual variance of the T4 PAP goals factor score were fixed to zero as these factor scores showed slightly negative, non-significant, residual variances.

Table 14c. *Standardized Correlation and Regression Coefficients of the SEM Investigating the Correlated Change of PAV Goals, Interest, and Achievement in Study 2.*

	PAV intercept	PAV slope	Interest intercept	Interest slope	Grades diff.
Correlations with					
PAV slope	-.53***				
Interest intercept	-.23***	.05			
Interest slope	.09*	-.14***	-.53***		
Grade diff. score	.02	-.08**	.05	.19***	
Regressions on					
Cog. a.	-.12***	.01	.12**	-.03	.11**
Sex	.10***	.03	-.05	-.01	.04
P. edu.	-.03	-.01	-.05	.11*	.02
Language b.	.03	-.02	.05	-.06	-.02

Note. **p* < .05. ***p* < .01. ****p* < .001. PAV = Performance-avoidance goals. Grades diff. = difference score of math grades. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. The residual variance of the T2 interest factor score was fixed to zero as this factor score showed a slightly negative, non-significant, residual variance.

Regarding the relation of initial level and change over time, the intercept of each construct was negatively related to the slope of the same construct. The intercept of mastery goals was negatively related to the slope of interest ($r = -.46$; $p < .001$). The intercept of PAV goals was positively related to the slope of interest ($r = .09$; $p = .032$). Finally, the intercept of interest was negatively related to the slope of mastery goals ($r = -.39$; $p < .001$).

Results for the REM model. Regression coefficients for the REMs, including all achievement goals and one outcome variable (i.e., interest or math grades) are reported in Table B20a and Table B20b in Appendix B. The cross-lagged relations between the achievement goals and the outcomes are of particular interest. Out of 15 possible cross-lagged effects from achievement goals to interest and grades, 5 effects reached significance. T1 mastery goals positively predicted T2 interest ($\beta = .06$; $p = .005$), and T1 PAP goals positively predicted T2 interest (β = .05; p = .005). T3 PAP goals positively predicted T4 interest (β = .06; p = .040). T2 mastery goals positively predicted T3 grades (β = .14; p < .001), and T3 mastery goals positively predicted T4 grades (β = .10; p < .001). Out of 15 possible cross-lagged effects from interest and grades on achievement goals, 7 effects reached significance. T1 interest positively predicted T2 mastery goals (β = .09; p < .001), T2 interest positively predicted T3 mastery goals (β = .10; p = .003), and T3 interest (β = .13; p < .001) and T3 grades (β = .07; p < .001) positively predicted T4 mastery goals. T3 grades positively predicted T4 PAP goals (β = .05; p = .011). T2 interest negatively predicted T3 PAV goals (β = -.06; p = .039). T3 grades negatively predicted T4 PAV goals (β = -.04; p = .048).

4.7. General Discussion

4.7.1. Results of Our New Studies

Longitudinal research on adolescents' achievement goals and their relation to the development of academic interest and achievement is sparse. Herein we reported two such longitudinal studies. The studies were comparable in methodology, thereby allowing us to discern which of the observed findings were most robust (i.e., consistent across studies).

With regard to change in achievement goals over time, our findings were highly consistent across the two studies. Specifically, all three of the focal achievement goals—mastery, PAP, and PAV—showed a decrease over time, indicating an overall drop in the *quantity* of motivation during secondary school. In addition, Study 1 revealed that for all three goals the decline was larger at the beginning of the first year of secondary school and then tapered off over time. In Study 2, this finding was also observed for mastery goals, whereas PAP goals decreased linearly and PAV goals showed a particularly large drop after the middle of Grade 5.

Regarding the joint development of achievement goals and math interest, we found positive relations between change in mastery and PAP goals and change in interest in both studies. Change in PAV goals, was either negatively related (Study 2) or unrelated (Study 1) to change in interest. Regarding the joint development of achievement goals and math achievement, change in PAP goals was unrelated to change in achievement in both studies, whereas change in mastery goals was either positively related (Study 2) or unrelated (Study 1) to change in achievement, and change in PAV goals was either negatively related (Study 2) or unrelated (Study 1) to change in achievement. The relations observed in the cross-lagged analyses were largely consistent with the aforementioned patterns from the joint development analyses (excepting the null findings for PAV goals in both studies), and provided a more detailed resolution of the relations over time. The cross-lagged results also showed some evidence of interest and achievement influencing achievement goal pursuit (particularly with regard to mastery and PAP goals). In short, the findings across the two studies were either highly consistent (exactly the same in each study) or reasonably consistent (e.g., positively related in one study and unrelated in the other, rather than positive in one study and negative in the other).

4.7.2. Results of Our New Studies in the Context of Prior Studies and Theoretical Predictions

We preceded our presentation of the two new longitudinal studies with a review of existing longitudinal work and an overview of various theoretical predictions. In the following, we compare our findings with those of prior work to discern the consistency therein and to determine which of the various theoretical predictions are best supported by the data. In doing so, it is important to bear in mind both the sparseness of the prior empirical work and the considerable diversity in samples, procedures, and measures across the prior and new studies.

Longitudinal change in achievement goals. With regard to the development of achievement goals, prior research has observed a decrease in mastery, PAP, and PAV goals over time (e.g., Paulick, Watermann, and Nückles, 2013; Shim, Ryan, and Anderson, 2008; for a broader review of competence motivation development, including achievement goals, see also Scherrer & Preckel, 2019). This is the same pattern that we observed in both of our studies—an overall drop in the quantity of competence-based goal pursuit. So, the findings for achievement goal development are not only robust across our studies, but are also robust across the full empirical corpus. The decrease in mastery goals during adolescence is predicted by social-cognitive and SEF theorizing; the decrease in performance-based goals during adolescence is not predicted by these theories, as both anticipate an increase in performance-based goals (with no distinction made between approach and avoidance; Anderman, Austin, & Johnson, 2002; Nicholls, 1989). However, at a broader level, the findings observed in our and the prior studies may be seen as concordant with the fundamental premise of social-cognitive and SEF theorizing, namely that secondary school is detrimental for adolescents' motivation (Eccles et al., 1993; Nicholls, 1989). Detrimental in this instance simply means a complete drop in all forms of competence-based striving, rather than a drop in mastery goal pursuit only. A general divestment from competence striving (akin to work avoidance; see Nicholls, Patashnick, & Nolen, 1985) is arguably even more problematic than a specific divestment from one form of competence striving (i.e., mastery goals). So, from this broader perspective, the observed findings are strongly supportive of social-cognitive and SEF theorizing.

Prior work and theorizing has largely ignored the distinction between linear and nonlinear patterns of achievement goal development. Only Paulick et al. (2013) have tested for nonlinear patterns and they, like us, found evidence for nonlinear change in all three goals. For PAP and PAV goals, the pattern was the same in their study and our Study 1—the drop in both of these goals was largest at the beginning of secondary school (i.e., the first half of the fifth grade) and became less pronounced over time. For mastery goals, we also found particular strong decreases between the two first measurement points in both of our studies, whereas in the study of Paulick et al. (2013) the drop in mastery goals was apparent at the beginning of secondary school but accelerated over time. The more precipitous drop in competence striving at the beginning of secondary school may be seen as concordant with SEF theory (Eccles et al., 1993; Wigfield & Eccles, 2002), especially our broader interpretation of this theory. The mastery goal finding in Paulick et al. (2013) is not what one would expect from an SEF perspective, but given that this result has only emerged once, it seems prudent to wait for it to be replicated before rigorously pursuing an explanation. The nonlinear findings in our and Paulick et al.'s (2013) work highlight the value of looking beyond linear patterns in longitudinal analyses of achievement goals; these findings provide more detail and nuance than those afforded by the more commonplace focus on linear patterns alone.

Longitudinal relations of achievement goals, interest, and achievement. In discussing our findings on the longitudinal relations of achievement goals with academic interest and achievement, we address each of the two outcome variables (i.e., interest and achievement) separately. For academic interest, prior research has found that positive development of mastery goals is related to a positive change in interest, and we observed these same result in both of our studies. Prior research has yielded null results for correlated change in PAP goals and interest, whereas we found evidence for positive relations in both of our studies. No prior research has examined correlated change in PAV goals and interest; in our Study 1, these variables were unrelated, whereas we found a negative relation between the change scores of PAV goals and interest in Study 2. These findings offer strong evidence for positive mutual relations in the development of mastery and PAP goals and interest in mathematics, and further provide suggestive evidence of a negative longitudinal relation between PAV goals and interest. In both studies, we found that the initial levels of mastery goals and interest were negatively related to change in mastery goals and interest, respectively. Musu-Gillette, Wigfield, Harring, and Eccles (2016) observed similar patterns between the initial level and change in interest. Students who initially report a higher level of interest seem to be prone to a larger decrease in interest over time as compared to students who initially report a lower level of interest. Possible explanations for this and the comparable finding for mastery goals, might be regression to the mean over time or a compensation effect in which educational resources are primarily allocated to the promotion of less able students (Schroeders, Schipolowski, Zettler, Golle, & Wilhelm, 2016).

For academic achievement, prior research has found that positive development of mastery goals is related to a positive change in achievement, and we observed these same results in one of our two studies (Study 2; a null result was observed in Study 1). Prior research has found a positive relation between change in PAP goals and change in achievement, but we observed no relation in both of our studies. No prior research has examined correlated change in PAV goals and achievement; in our Study 1, these variables were unrelated, whereas we found a negative relation between these variables in Study 2. These findings provide suggestive evidence for positive longitudinal relations between mastery goals and achievement, and negative longitudinal relations between PAV goals and achievement. They raise questions regarding the robustness of the previously documented positive longitudinal relations between PAP goals and achievement, and suggest the need to search for moderator variables (e.g., demonstration vs. normative focus, domain of study, country of origin, age of student, etc.; see Hulleman et al., 2010). The initial level of the achievement goals was only related to change in achievement in one of our studies for one of the goals; in Study 2, the intercept of mastery goals was positively related to the change in achievement.

The results from the correlated change SEMs that were the focus of our research provide important information on the relations between the three achievement goals and interest and achievement over time. However, these analyses do not test the causal ordering of the relations. Therefore, to complement these analyses we examined cross-lagged effects between the achievement goals and both interest and achievement over all measurement points using REMs. These analyses provide a more detailed resolution of relations over time and allow a test of causal ordering. However, in interpreting these results it is important to bear in mind that compared to correlated change SEMs, REMs underestimate the shared development over the whole time period and do not account for non-linear developmental trends.

In the REMs, mastery and PAP goals were positive or non-significant predictors of interest in both studies, whereas PAV goals did not significantly predict interest in either study. These findings for mastery and PAV goals are consistent with cross-lagged findings from prior empirical work, but our findings for PAP goals differ from prior work, which has obtained null findings for these goals. For achievement, mastery goals were positive or non-significant predictors in both studies, whereas PAP and PAV goals were no significant predictors in either study. These findings for mastery goals are in accordance with prior research. Prior work with PAP goals has found positive, null, and negative results, and prior work with PAV goals has found them to be a negative or non-significant predictor of interest. Thus, here too questions are raised regarding the robustness of the PAP and PAV relations, and subsequent work would do well to investigate potential moderator variables (see our comparable point above for candidate moderators) that could unpack these diverse findings. Interestingly, several reverse findings were observed in which interest or achievement influenced an achievement goal. This suggests that the development of achievement goals and our focal outcomes of interest and achievement may be reciprocal, a possibility clearly in need of additional, and more targeted investigation.

In sum, several points are noteworthy regarding the patterns of relations for the three achievement goals. First, taken together, our new and the prior findings for adolescent students map quite nicely onto the findings observed in the undergraduate and adult literature on achievement goals in general—mastery goals are the most adaptive for academic interest and achievement, PAV goals are the least adaptive, and PAP goals fall in between (and tend to be adaptive). These patterns are presumed to be grounded in the differential sets of processes evoked by the different achievement goals (see Elliot, 1999; 2005) and the implications of these goal processes appear to be generalizable across adolescence and adulthood. Second, there is a noteworthy degree of consistency in the prior and new results, especially for academic interest, and the inconsistency that *is* present almost exclusively involves a finding being significant (positive or negative) in one instance and null in another. Third, cross-lagged analyses indicate that relations between the achievement goals on the one hand, and interest and achievement on the other, appear to be reciprocal rather than unidirectional. Fourth, at least to some degree, inconsistency in findings across studies is to be expected, given the different sample demographics, procedures, measures, and achievement contexts across the various studies in the literature. Nevertheless, a clear priority for subsequent empirical work is to consider methodological and substantive variables carefully and systematically that might moderate longitudinal relations between achievement goals, interest, and achievement.

4.7.3. Implications

Overall, the empirical and theoretical work on achievement goal development and predictive utility over time clearly identifies secondary school as an important period of motivational vulnerability for adolescents. Mastery goals are clearly beneficial for academic interest and achievement, and PAP goals seem somewhat adaptive for these outcomes as well, but both of these goals show a decline during secondary school. Accordingly, educators would do well to target this period as an important time for intervention.

Several different types of interventions may be considered. One type could focus broadly on the documented mistmatch between adolescents' needs and general school practices (Eccles & Roeser, 2009). For example, smaller classes that afford better access to teachers, the provision of extra-curricular activities that signal an interest in student well-being (Eccles & Templeton, 2002; Lee & Smith, 2007), and later starting times that better align with adolescent sleep cycles (Scherrer, Roberts, & Preckel, 2016; Preckel, et al., 2013) may help students feel more supported and engaged in the educational process. This, in turn may result in a sustained commitment to competence in the form of mastery, and perhaps PAP, goal pursuit. A second type of intervention could focus on structural aspects of the achievement context per se that are expected to facilitate mastery goal pursuit (Ames, 1992). For example, providing adolescents with autonomy in course selection, moderately challenging and intrinsically interesting tasks, and private and intrapersonal competence feedback may bolster students' pursuit of mastery goals (Anderman, Maehr, & Midgley, 1999; Maehr & Midgley, 1996; O'Keefe, Ben-Eliyahu, & Linnenbrink, 2012). A third type of intervention could focus on explicitly and directly encouraging students to pursue mastery goals (Elliot & Hulleman, 2017; Martin, 2008; Smeding, Darnon, Souchal, Toczek-Capelle, & Butera, 2013). This explicit encouragement toward mastery goals could be accompanied by an equally explicit guiding students away from PAV goal pursuit. Performance-approach goals are believed to be most beneficial when they emerge of their own accord, rather than being encouraged from an external source (Elliot & Moller, 2003).

4.7.4. Strengths and Limitations

Our new longitudinal studies presented herein have both strengths and weaknesses. One strength is that we conducted two multi-year longitudinal studies, which is a rarity in the literature (see also Wang & Pomerantz, 2009). A second strength is that both of our studies used large sample sizes—745 and 1420 students in Studies 1 and 2, respectively. This is undoubtedly responsible for the substantial level of consistency in findings across the two studies. A third strength is that our analyses on the predictive utility of achievement goals were quite rigorous in that they controlled for cognitive ability, sex, parents' highest educational degree, and language background. A fourth strength is that our longitudinal analyses were nuanced, in that they examined both linear and nonlinear change in achievement goals, relations of both intercepts and slopes with the two central outcome variables, and reciprocal cross-lagged relations among the achievement goal and outcome variables.

One limitation of our research is that it focused on only one type of adolescent school experience, that of German students in the highest track of the secondary school system. A second, related, limitation is that our studies were conducted in only one country—Germany. The studies were conducted in three different federal states in Germany that have been shown to have different levels of selectivity and rankings of educational quality (Anger, Plünnecke, & Schüler, 2018), and this may, in part, account for the differences that did emerge across studies. Nevertheless, the fact that both studies focused on students in Germany means that the scope of our work was limited to a Western European educational context. A third limitation is that our research focused on the three goals of the trichotomous achievement goal model, leaving out mastery-avoidance goals of the 2 x 2 model and the differentiation of task- and self-based mastery goals of the 3 x 2 model. Our work was the first to investigate the longitudinal influence of interindividual differences in PAV goals on academic interest and achievement, but these other goal constructs were not used in our investigations and, indeed, have yet to be investigated in the literature. Finally, we only used a limited set of control variables (e.g., students' cognitive ability, sex, parents' highest educational degree, and language background). Additional variables such as academic self-concept might have influenced both the development of the achievement goals and the development of the outcomes. Therefore, additional potentially confounding variables should be included in the future research agenda.

4.7.5. Conclusion

In closing, adolescence is an important period for the development of achievement motivation, including achievement goal pursuit. Educators, be they administrators, teachers, or parents, would do well to attend to the unique challenges facing students during this time. Longitudinal research is essential to document changes in achievement goals during adolescence and their implications for important achievement-relevant outcomes. Equally essential is intervention research that seeks to provide clear guidelines for how to best structure schools, classrooms, and home learning environments to maximize the likelihood that the adolescents under our care will develop to their full potential.

Chapter 5. Article 3

High-Ability Grouping: Benefits for Gifted Students' Achievement Development Without Costs in Academic Self-Concept

Franzis Preckel, Isabelle Schmidt, Eva Stumpf, Monika Motschenbacher,
Katharina Vogl, Vsevolod Scherrer, and Wolfgang Schneider

5.1. Abstract

Effects of full-time ability grouping on students' academic self-concept (ASC) and mathematics achievement were investigated in the first three years of secondary school (four waves of measurement; students' average age at first wave: 10.5 years). Students were primarily from middle and upper class families living in southern Germany. The study sample comprised 148 (60% male) students from 14 gifted classes and 148 (57% male) students from 25 regular classes (matched by propensity score matching). Data analyses involved multi-level and latent growth curve analyses. Findings revealed no evidence for contrast effects of class-average achievement or assimilation effects of class type on students' ASC. ASC remained stable over time. Students in gifted classes showed higher achievement gains than students in regular classes.

Key words: ability grouping, giftedness, academic achievement, academic self-concept, longitudinal study, latent growth curve modeling, propensity score matching

5.2. Introduction

Numerous studies have investigated the effects of ability grouping on students' academic achievement (for a review, see Steenbergen-Hu, Makel, & Olszewski-Kubilius, 2016) and students' academic self-perceptions (for a review, see Seaton, Marsh, & Craven, 2009). Perceptions of one's own academic ability including one's academic self-concept (ASC) influence achievement, motivation, affect, and general well-being; therefore, they are central for understanding and fostering students' development (Trautwein & Möller, 2016). However, researchers in the field have not been able to reach a consensus about the usefulness of ability grouping for achievement or academic self-concept – especially for the grouping of high-ability or gifted students. Regarding achievement, some researchers have concluded that no one gains academically from ability grouping (Oakes, 2008), whereas others have suggested that high-ability students benefit from it (Colangelo, Assouline, & Gross, 2004; Rogers, 2007). Regarding ASC, some researchers have warned of costs of high-ability grouping (Belfi, Goos, De Fraine, & Van Damme, 2012; Marsh et al., 2008). Students use the ability of their classmates as a frame of reference against which they evaluate their own standing to form their ASC (Huguet et al., 2009). When the ability level of the reference group increases, ASC decreases. The effect of the reference-group on individual ASC is usually of medium size (e.g., for the PISA 2003 data, effect size of .49; Seaton, Marsh, & Craven, 2009; for the PISA 2006 data, effect size of .40; Nagengast & Marsh, 2012). By lowering ASC, high-ability grouping might therefore impair students' achievement, motivation, affect, or well-being (Marsh et al., 2008). However, other researchers have argued that the reference group effect of high-ability grouping on students' ASC is short-term and not universal (Dai, 2004). Overall, there is a lack of empirical evidence informing this discussion. Despite the wealth of studies on the effects of ability grouping, few studies in the field of high-ability grouping have investigated achievement and ASC simultaneously and even fewer have used longitudinal data. The present study fills this research gap. We carried out a 3-year longitudinal study that investigated the development of high-ability students' academic achievement and ASC over time as a function of ability grouping in gifted classes, as compared to heterogeneous classrooms.

When studying the effects of high-ability grouping, one has to deal with the problem of selection bias. On average, high-ability students in settings of ability grouping like gifted classes differ from students in regular classrooms not only with regard to their higher academic ability and achievement, but also regarding other variables such as age, gender, or family background. For example, students in gifted classes frequently are younger and come from families with a higher socioeconomic status (Schneider, Preckel, & Stumpf, 2014; Vogl & Preckel, 2014; Roznowski, Reith, & Hong, 2000). In addition, more boys than girls attend gifted classes (Peterson, 2013). A meta-analysis by Goldring (1990) on achievement differences between ability-grouped and ungrouped gifted students revealed that the number of variables used to match students in both settings influenced the effect sizes for mean achievement differences in favor of grouping. Studies that matched the students using fewer variables obtained larger effects sizes. This suggests that a priori differences between samples need to be taken into account when investigating the effects of ability grouping on the achievement and ASC of high-ability students. The present study controlled for these differences by matching students on relevant variables using propensity score matching.

To conclude, our study is the first to investigate the joint development of high-ability students' academic achievement and ASC over time as a function of full-time ability grouping in gifted classes or regular classes with matched student samples.

5.2.1. Ability Grouping and Achievement of Gifted Students

Most reviews and meta-analyses indicate that ability grouping generally does not provide substantial academic benefits. For instance, Kulik and Kulik (1992) carried out a meta-analysis on ability grouping procedures that included elementary as well as secondary school students. The results showed that the benefits were rather small for comprehensive grouping programs (see also Kulik & Kulik, 1982, 1984; Slavin, 1990). Similarly, in a more recent review of meta-analyses based on over 300 studies, Hattie (2009) found an average effect size of ability grouping on educational outcomes of d = 0.11, indicating minimal effects on learning outcomes. However, effect sizes vary depending on the form of ability grouping. In their second-order meta-analyses, Steenbergen-Hu et al. (2016) synthesized the outcomes of 13 ability grouping meta-analyses. Results indicated that students benefited from within-class grouping (Hedges' g = 0.25) and cross-grade subject grouping (Hedges' g = 0.26), whereas the benefits were negligible from between-class grouping (Hedges' g = -0.03). It is important to note, however, that these findings do not necessarily generalize to the grouping of gifted students. While some studies found no effects of ability grouping on gifted students' achievement development (e.g., Adelson, McCoach, & Gavin, 2012), many studies reported positive effects for achievement outcomes (Adelson & Carpenter, 2011; Hattie, 2002; Rogers, 2007; Tieso, 2005). Kulik and Kulik (1992) identified 25 studies of special programs for the gifted and found that ability grouping had a positive effect (Cohen's d = 0.41) on the achievement of gifted elementary school children. Further supportive evidence came from a study by Lou, Abrami, and Spence (2000) who analyzed 51 studies of ability grouping and found that high-ability students particularly benefited from the grouping procedure. These and further meta-analyses were included in the second-order meta-analyses of Steenbergen-Hu et al. (2016). The original mean effect sizes of the six included meta-analyses ranged from Hedges' $g = 0.32$ to 0.47 with a statistically significant, integrated effect size of 0.37. Further, the original effect sizes appeared to be homogeneous, "suggesting that gifted students benefited from being placed in special groups or programs that were specifically designed to serve those with initial high achievement levels or learning potential" (Steenbergen-Hu et al., 2016, p. 876). This finding also corresponds to results reported in Hattie's (2009) review of meta-analyses, which showed that there is a higher likelihood of success in learning when ability grouping is provided for gifted and talented students. Overall, Hattie reported an average effect size of $d = 0.30$ for gifted education programs. However, the size of the effect in favor of ability grouping for the gifted is influenced to a large extent by the number of variables used to match students in ability-grouped versus ungrouped settings (Goldring, 1990). This finding indicates that the achievement advantages of high-ability grouping for gifted students can partly be explained by a priori differences between comparison groups regarding achievement, ability, or other achievement-related variables (e.g., SES).

5.2.2. Ability Grouping and Academic Self-Concept of Gifted Students

Ability grouping can lower high-ability students' ASC, which can be explained by the "big-fish-little-pond-effect" (BFLPE; Marsh & Parker, 1984). The BFLPE describes a frame of reference model according to which self-perceptions in educational settings are largely shaped by social comparison processes (Huguet et al., 2009). Students' achievement and ASC are positively related. However, when controlling for this relation, the average achievement of the reference group is negatively related to students' ASC (see Seaton et al., 2008, for an overview). The most salient reference group is usually students' own class (Wouters, Colpin, van Damme, de Laet, & Verschueren, 2013). When students engage in social comparisons with a highly able reference group such as a gifted class, they experience a negative contrast effect and their ASC decreases. Many findings support negative contrast effects of gifted ability grouping (Preckel, Zeidner, Goetz, & Schleyer, 2008; Craven, Marsh, & Print, 2000; Zeidner & Schleyer, 1998). However, belonging to a gifted class can also make students feel positively about their abilities by assimilation effects (Seaton et al., 2008). Perceived similarities should motivate students to draw closer to the high-ability in-group by associating themselves with its positive qualities (Mussweiler, 2003). Accordingly, for gifted classes, Preckel and Brüll (2010) found a strong positive assimilation effect that fully compensated for the negative contrast effect. That is, grouping in gifted classes had no significant negative

BFLPE on ASC. In a recent study, Herrmann, Schmidt, Kessels, and Preckel (2016) investigated contrast and assimilation effects in mathematics and the verbal domain in regular and gifted classes. In mathematics, an assimilation effect of membership in gifted classes compensated for negative contrast effects of class-average achievement on ASC even after controlling for previous ASC. In the verbal domain, there was no evidence for a significant contrast or assimilation effect. Some investigations of gifted education programs showed only short-term declines in ASC or even no decline at all (Dai & Rinn, 2008; Makel, Lee, Olszewki-Kubilius, & Putallaz, 2012). Moreover, the ASC of high-achieving students appears to be less affected by the BFLPE than that of lower achieving students (Marsh & Rowe, 1996; Trautwein, Lüdtke, Marsh, & Nagy, 2009). However, more research is needed to clarify the extent to which high achievement or positive assimilation effects buffer negative contrast effects of a high-ability reference group (Makel et al., 2012; see also Chmielewski, Dumont, & Trautwein, 2013). Further, no study of gifted grouping modeled contrast effects of class-average achievement and assimilation effects of class type, while controlling for selection bias by matching students in the different class types. However, a priori differences in cognitive or academic ability, age, socio-economic status, or gender might relate to these effects. For example, socio-economic status might interact with assimilation effects. Therefore, in our study we investigated contrast and assimilation effects of gifted ability grouping with a matched student sample.

5.2.3. Relation of Academic Achievement and Academic Self-Concept

Academic achievement and ASC show reciprocal, mutually reinforcing relations. That is, prior ASC has a positive effect on subsequent achievement beyond what can be explained by prior academic achievement. In addition, subsequent ASC is also positively affected by prior achievement beyond what can be explained by prior ASC (Huang, 2011; Marsh & Martin, 2011). In a recent study, Seaton, Marsh, Parker, Craven, and Yeung (2015) investigated reciprocal effects between achievement and ASC comparing academically selective and mixed-ability comprehensive schools. The findings supported the reciprocal effects model for both school types indicating its generalizability over heterogeneous and high-ability environments.

5.3. The Present Study

We investigated the development of self-concept and achievement in high-ability students in both gifted classes (with full-time ability grouping) and regular classes in the highest educational track (i.e., “Gymnasium”) of the German three-track secondary school system. In Germany, the placement of students to Gymnasium is based on their academic performance, teachers' recommendations, and parents' preferences. Of note, Gymnasium is itself not a gifted track given that approximately one third of the total student cohort attends these schools (Federal Statistical Office, 2015). Within this track, schools in our study offered full-time ability grouping in gifted classes ("school-within-a-school") starting in Grade 5. In these gifted classes the standard curriculum was presented at a faster pace (acceleration) and more in depth (enrichment) than in the regular classes. In detail, the schools offered a compacted curriculum, bilingual lessons, additional lessons in scientific subjects (e.g., computer science or experiments) and interdisciplinary projects. We assessed students' achievement and ASC in the domain of mathematics four times over a three-year period at the beginning of secondary school. We controlled for selection bias by using propensity score matching (see *Participants*).

Past studies investigating the BFLPE for ability grouping of gifted students have reported mixed findings (negative BFLPE: Craven et al., 2000; Zeidner & Schleyer, 1998; no negative BFLPE: Preckel & Brüll, 2010; Herrmann et al., 2016; Makel et al., 2012). Using the BFLPE framework (i.e., multilevel analyses), we therefore investigated contrast and assimilation effects of ability grouping in either regular classes or gifted classes at each wave of measurement of our study. We conducted multilevel analyses with a matched student sample to rule out alternative explanations of our findings by a priori group differences. Additionally, we controlled for students' prior ASC in our analyses so that our findings refer to *change* in ASC. We thus conducted a rigorous test of contrast and assimilation effects of gifted ability grouping, in response to the inconsistent research findings concerning negative reference-group effects and positive assimilation effects of gifted ability grouping (e.g., Chmielewski et al., 2013; Makel et al., 2012).

Further, we compared the development of academic achievement and ASC over time as a function of class type (gifted vs. regular). We investigated group differences in the development (i.e., change) in achievement and the level of achievement at the last wave of measurement. There is meta-analytic evidence for positive effects of gifted ability grouping on students' achievement (Steenbergen-Hu et al., 2016) but most available findings were based on unmatched samples and thus might in part be explained by selection bias (Goldring, 1990). For ASC, we expected decreases in both class types because a number of studies have described a normative pattern of developmental decline in multiple self-concept domains including math ASC from middle childhood into adolescence (Gest, Molloy, & Ram, 2015; Wigfield & Wagner, 2005). We explored whether the decrease in ASC was stronger in gifted classes than in regular classes with no specific hypothesis in mind because, as stated above, research findings for negative reference-group effects of gifted ability grouping are not clear. Based on evidence for the reciprocal effects model for high-ability students (Niepel, Brunner, & Preckel, 2014; Seaton et al., 2015), we expected positive relations between the development of ASC and the development of achievement over time.

5.4. Method

5.4.1. Procedure and Participants

The data used in this study were collected as part of the project "PULSS", which assessed developmental patterns in academic achievement and motivation in the early secondary school years (i.e., from Grade 5 to 7) in regular high-track schools (Gymnasiums) of the German three-track secondary school system that offered regular classes and gifted classes (i.e., full-time ability grouping). The gifted classes were part of the regular secondary school setting and thus could be compared with the regular classes in the same school.

Data collection took place in class in group sessions. Students were tested at the beginning and at the end of Grade 5, at the end of Grade 6, and in the middle of Grade 7. Before starting the regular or gifted classes in Grade 5, all students had attended a regular primary school. The average number of days between the transition from elementary school and measurement was 39 days for the first wave, 246 days for the second wave, 606 days for the third wave, and 917 days for the fourth wave of measurement, with no differences by class type. Students' cognitive ability was assessed once at the beginning of Grade 5. Their academic achievement and ASC were assessed at all four waves of measurement (see *Material*).

Students' participation was voluntary, and written parental consent was obtained for all participants. In addition, parents filled out questionnaires at home. Student and parent data as well as student data across waves of measurements were matched by using a pseudo-anonymized identification code (i.e., self-generated alphanumeric code according to a predefined procedure). On average, 25 students in regular classes and 20 students in gifted classes participated in the study (response rates: 82% for regular classes, 81% for gifted classes).

The overall sample comprised 922 students from two successive cohorts (beginning with the school year 2008/2009) attending seven regular high-track schools in southern Germany. Four schools were located in the federal state of Baden-Württemberg (one in a small city, two in mid-sized cities, and one in a large city); three schools were located in the federal state of Bavaria (one in a small city, one in a mid-sized city, and one in a large city). The cohorts did not differ significantly on any variable included in the present study. Month' of data collection were October 2008/2009, February 2009/2010, May 2009/2010, and February 2010/2011 (cohort 1/2). Most students had a German language background (89%); the language background of the other students was Russian (4%), English (2%), Turkish (2%), or other (<1%: Chinese, Italian, French, and Spanish). The majority of students came from middle and upper class families. In 61% of these families, at least one parent had a college/university degree or a PhD (see Table 15 for exact information on parents' highest educational degree).

From the overall sample of 922 students, 283 (186 male) students attended 14 gifted classes and 639 (355 male) students attended 26 regular classes. Those students who entered the gifted classes in Grade 5 had to apply for these classes, usually based on parents' initiative. That is, application for gifted classes was based on self-selection, and therefore not all gifted students attended gifted classes. For various reasons, some gifted students attended regular classes. For those students who applied for the gifted classes, all schools employed multistage selection procedures. The selection procedures were comparable over schools and included intelligence tests (minimum IQ of 120), teacher observations of students' behavior during one or two days of probationary class, and interviews with parents. Applicants were then selected in a conference among teachers, school psychologists, and school board members based on a partly compensatory strategy (i.e., high achievement could partly compensate for an IQ of slightly below 120 and vice versa).

Propensity score matched sample. To control for differences between the students in gifted classes and regular classes, we used propensity score matching. We estimated a propensity score (PS) for each student by boosted regression (McCaffrey, Ridgeway, & Morral, 2004). A PS is defined as "the conditional probability of assignment to a particular treatment given a vector of observed covariates" (Rosenbaum & Rubin, 1983, p. 41). In our study, the treatment comprised attending a gifted class. As covariates, we used students' cognitive ability, age, gender, parents' highest educational degree (either mother or father), the particular school the student attended, student test achievement (math, reading speed, text comprehension), ASC (general, math, verbal), and need for cognition defined as the "tendency for an individual to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). A detailed description of the PS estimation is given in the Appendix C. The covariates were assessed at the first wave of measurement and were selected based on theoretical considerations; they all showed significant positive relations with the treatment variable.

Table 15. *Description of the Unmatched Sample (Overall and by Class Type) and the Matched Sample (Model 3) on the Covariates Used for Matching with Effect Sizes Cohen's d for Mean Differences between GC and RC (d_1) and between Matched Students from GC and all Students from GC (d_2)*

	Unmatched Sample				**Matched Sample**			
Covariates	**All** *N*=922	**RC** *n*=639	**GC** *n*=283	d_1	**GC** *n*=148	**RC** *n*=148	d_1	d_2
Cognitive ability *M (SD)*	111.98 (12.31)	107.21 (9.90)	122.61 (10.35)	1.49	118.77 (8.68)	117.98 (6.75)	.12	.37
Age in months *M (SD)*	126.34 (6.62)	127.90 (5.16)	122.83 (8.04)	.63	125.64 (5.96)	125.94 (5.84)	.05	.35
Math ACH *M (SD)*	-.57 (.96)	-.87 (.83)	.10 (.90)	1.08	-.15 (.85)	-.23 (.71)	.12	.27
Reading speed *M (SD)*	7.81 (5.43)	7.22 (5.91)	9.21 (3.71)	.54	8.78 (3.78)	8.23 (3.28)	.17	.12
Text comprehension *M (SD)*	23.48 (6.01)	22.10 (5.95)	26.62 (4.85)	.93	25.60 (4.63)	24.55 (5.37)	.20	.21
General ASC *M (SD)*	4.14 (.65)	4.11 (.66)	4.23 (.62)	.19	4.19 (.65)	4.24 (.58)	.09	.06
Math ASC *M (SD)*	4.09 (.86)	3.98 (.89)	4.32 (.75)	.45	4.24 (.82)	4.24 (.75)	.01	.11
Verbal ASC *M (SD)*	4.09 (.73)	4.03 (.74)	4.22 (.70)	.26	4.19 (.71)	4.16 (.63)	.05	.04
Need for cognition *M (SD)*	3.52 (.68)	3.42 (.66)	3.74 (.67)	.48	3.64 (.68)	3.59 (.61)	.10	.14
Gender male *n*	541	355	186		89	85		
female *n*	381	284	97		59	63		
Parents' highest educational degree (degree: *n* parents)	1: 2 2: 27 3: 155 4: 119 5: 343 6: 132	1: 2 2: 26 3: 123 4: 98 5: 207 6: 70	1: 0 2: 1 3: 32 4: 21 5: 136 6: 62		1: 0 2: 1 3: 22 4: 15 5: 59 6: 36	1: 1 2: 2 3: 19 4: 16 5: 56 6: 32		

Note. GC = gifted classes. RC = regular classes. ACH = achievement. ASC = academic self-concept. *d* = standardized mean difference Cohen's *d* (difference of *M* SC and *M* RC divided by *SD* SC). Educational degree: 1 = no educational degree, 2 = certificate of secondary education, 3 = secondary school level I certificate, 4 = vocational/university entrance diploma, 5 = college/university degree, 6 = PhD.

Using the PS, matching of students in gifted and regular classes was implemented with the MatchIt Package (Ho, Imai, King, & Stuart, 2011) in the statistical program R. Overall, we applied five different matching strategies in various combinations resulting in five different matching solutions or models. A detailed description of the matching strategies and resulting models is given in the Appendix C. Of note, the student subsamples selected from the gifted and regular classes varied by model. For all models, Table C1 in the Appendix C reports standardized mean differences in covariates between students in the gifted and regular classes before and after the matching. Table C1 further reports standardized mean differences in covariates between all students in gifted classes and those students in gifted classes selected by the matching procedure. These values indicate whether the matched students from the gifted classes were representative of the full student population in the gifted classes or whether a systematic distortion accrued due to the exclusion of some students.

We compared the matching solutions from Models 1 to 5 and chose Model 3 as the best model (n = 148 students per group; see Appendix C for a detailed description of the model comparison). It offered the best solution taking into account (a) the balancing of covariates (i.e., effect sizes $d \leq .20$ for mean differences between groups) and (b) the representativeness of the gifted class subsample for all students in the gifted classes. There were no relevant mean differences (all differences were $d \leq .20$) between the matched students from regular classes and gifted classes and very few small differences (d = .20 to .50) between the students from gifted classes selected by the matching procedure and all students in the gifted classes. Therefore, all further analyses were computed using the subsample selected by Model 3, which comprised 148 (89 male) students from 14 gifted classes and 148 (85 male) students from 25 regular classes (see Table 15).

To test the robustness of findings from the different matching solutions, we additionally ran all analyses with the subsamples from Model 1, 2, and 4. Model 5 was excluded from further consideration because of the extreme weights of some students in regular classes (e.g., two high-performing students from regular classes accounted for 90 students from gifted classes). Robustness checks revealed that findings were very robust over the four matching solutions (results are reported in Appendix D). We found no differences in the results of the multilevel analyses testing the BFLPE. In addition, the results for the level and the development of academic achievement and ASC as a function of class type (gifted vs. regular) were comparable over the four matching solutions. However, there was one exception: For Model 1, which included all students from gifted classes, the effect of class type on the development of math achievement shortly missed the 5%-significance level (p = .063; see Appendix D, Table D5).

Before the matching, effect sizes for mean differences between all 639 students in regular classes and all 283 students in gifted classes were large for cognitive ability, math achievement, and text comprehension ($d_1 > .80$), of medium size for age and reading speed (d_1 = .50 to .80), and small to non-relevant for differences regarding the other covariates ($d_1 <$ .50) (see Table 15). After the matching, all effect sizes for mean differences between regular and gifted classes were small ($d \leq .20$). Moreover, the distribution of gender and parents' highest educational degree was comparable over class type.

5.4.2. Materials

Math achievement. We used standardized achievement tests (Götz, Lingel, & Schneider, 2013a, 2013b) that assessed competencies in arithmetic, algebra, geometry, and word problems. The test content matched the objectives of the math curriculum showing curricular validity. Separate test versions with partly overlapping material were used for each of the four measurement points (test version 1/2/3/4: 26/31/37/39 items). The minimum and maximum scores for the test versions ranged from 0 to 14 (Version 1), from 0 to 16 (Version 2), from 0 to 25 (Version 3), and from 0 to 27 (Version 4). All students got Version 1 at the first wave of measurement, Version 2 at the second wave of measurement, etc. Internal consistencies of all four test versions were satisfactory ($\alpha = .73–.87$). There were no ceiling effects (skewness values for the four versions lay between 0.031 and 0.377 for the regular classes and between -0.116 and -0.761 for the gifted classes).

Academic self-concept. ASC in math was assessed with four items from the Self-Description Questionnaire (SDQ; Marsh, 1990) which is considered to be one of the best self-concept instruments available (Byrne, 2002). Items were: *Mathematics is one of my best subjects*; *I get good marks in mathematics*; *I have always done well in mathematics*; and *I learn mathematics quickly*. The first three items represent the math self-concept scale used in PISA 2000; the fourth item was added in PISA 2003. Item translation into German was carried out in the context of the PISA study (Kunter et al., 2002). Past research has documented the reliability and validity of measuring ASC with a short scale (e.g., 3-item scales; Gogol et al., 2014). Students responded to the items on a 5-point rating scale ranging from 1 (*strongly disagree)* to 5 (*strongly agree),* with higher scores indicating a higher ASC. Sample internal consistencies were high, ranging from .87 to .92.

Cognitive ability. We used the Cognitive Ability Test for Grades 4 to 12 (Kognitiver Fähigkeitstest für 4. bis 12. Klassen; KFT 4-12+R; Heller & Perleth, 2000) which is a German adaptation of the Cognitive Abilities Test developed by Thorndike and Hagen (1971). In Germany, the KFT 4-12+R is one of the most frequently used cognitive ability tests in the research on giftedness and education. The KFT 4-12+R can be used to assess verbal, numerical, and figural reasoning as well as a composite score. Reasoning ability or the "capacity for processing power/formal logical thinking and judgment ability" (Carroll, 1993, p. 64) is a core construct of general intelligence (Carroll, 1993). The 90-minute short version of the test was administered in class using a paper-and-pencil format. The sample alpha was .95 for the composite score.

Parents' educational background. Parents reported their highest educational degree in a separate parents' questionnaire on an ordinal scale with the following answer options: *no educational degree, certificate of secondary education, secondary school level I certificate, vocational/university entrance diploma, college/university degree, PhD*.

Student background variables. Students reported their gender and birthday in a demographic questionnaire that accompanied the tests.

5.4.3. Data Analysis

Equating of math test scores. To compare achievement over the four versions of the math test, we used vertical linear equating under the common item equivalent groups design (mean-mean equating; Kolen & Brennan, 2004). Table C3 in the Appendix C shows the design. We used an IRT 1PL-Model (Embretson & Reise, 2000) for the linking procedure. All common items were taken into account with the same weight. Item and person parameters were estimated with the software ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007). We used the data of all participants who worked on the respective test version and estimated person parameters by weighted maximum likelihood estimation (WLE; Warm, 1989). In the equating, the multilevel structure of the data was not taken into account. Table C3 in the Appendix C shows the item parameters of the vertically linked test versions as well as the reliability of person parameters per test version.

BFLPE: Multilevel analyses. For the investigation of the BFLPE, we conducted manifest multilevel analyses (Raudenbush & Bryk, 2002) *using Mplus* version *7.4* (Muthén & Muthén, 1998-2015). Continuous variables were standardized over the whole sample of 922 students (M = 0, SD = 1); the dichotomous variable class type remained in its original metric (regular classes = 0; gifted classes = 1). Math test scores were aggregated at the class level from the standardized achievement measures including all students from the overall sample of 922 students to obtain an index of the average achievement per class and were not re-standardized. By doing so, Level 1 and Level 2 coefficients can be read as standardized outcomes and the Level 2 coefficient for class-average achievement directly depicts the contrast effect (Raudenbush & Bryk, 2002). Missing data were handled using the full information maximum likelihood (FIML) algorithm in Mplus. Further analyses were conducted with the matched student sample (N = 148 students per class type). For reasons of comparability with traditional BFLPE research, we also report the results from a BFLPE analyses with the whole sample in the Appendix C (Table C2).

To calculate the manifest mean scale score of the ASC measure from each wave of measurement as dependent variable, we computed a random-intercept multilevel model including individual math ability as Level 1 predictor (i.e., individual level) and class-average math ability as well as class type as Level 2 predictors (i.e., class level). From the second wave of measurement on, we controlled for ASC from the prior wave of measurement by including it as an additional Level 1 predictor. The equation for our final model was:

Level-1: (ASC) = $\beta_{0j} + \beta_{1j} *$ (prior ASC_{T-1}) + $\beta_{2j} *$ (individual achievement) + r_{ij} (1)

Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}$ (whole class-average achievement) + γ_{02} (class type) + μ_{0j}

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

Legend:

β_{0j} = random intercept of the regression at Level-1

β_{1j} = effect of prior ASC_{T-1} on ASC at Level-1

β_{2j} = effect of individual achievement on ASC at Level-1

r_{ij} = residual at Level-1

γ_{00} = average intercept

γ_{10} = slope of Level-1 predictor prior ASC_{T-1}; pooled within-class regression coefficient for prior ASC_{T-1}

γ_{20} = slope of Level-1 predictor individual achievement; pooled within-class regression coefficient for individual achievement

γ_{01} = effect of whole class-average achievement on β_{0j}

γ_{02} = effect of class type on β_{0j}

μ_{0j} = residual at Level-2

Common development of achievement and ASC: Confirmatory analyses. The common development of achievement and ASC was investigated with confirmatory methods (factor analyses, CFA; structural equation models, SEM; latent growth curve models, LGCM). We used the maximum likelihood estimator with robust standard errors (MLR; Kaplan, 2009) in Mplus 7.4 (Muthén & Muthén, 1998-2015). 75% to 100% of the students provided data to estimate variances or covariances at any time of measurement (covariance coverage; ASC items: 79.6-100%; math tests: 75-95.5%; cognitive ability test: 79.6-100%). Missing data were handled by the full information maximum likelihood (FIML) approach implemented in Mplus. To correct for the clustered sample structure (“students within classes”), we used the “type is complex” option in Mplus.

Measurement invariance testing. As a precondition for our analyses, we had to ensure measurement invariance of the ASC measure for the different class types and for longitudinal analysis. First, we ensured measurement invariance of the ASC measure *across class types per wave of measurement* in a multi-group CFA. We applied a stepwise strategy to test three levels of measurement invariance: (1) configural invariance (same pattern of factor loadings across groups), (2) metric factorial invariance (factor loadings were additionally constrained to be equal across groups), and (3) scalar factorial invariance (invariant item intercepts were additionally constrained to be equivalent across groups) (e.g., Meredith, 1993; Meredith & Horn, 2001). Second, within each class type, measurement invariance *across time* was tested using a four-wave SEM and the same stepwise strategy described above. We successively tested increasing measurement invariance levels against each other. Measurement invariance was assumed if the more restrictive model showed only minor deteriorations in model fit. Following Chen’s (2007) recommendation, we used the difference between comparative fit indices (ΔCFI) to compare two models, in addition to χ^2-difference tests. ΔCFI values of .01 or less were interpreted as a tolerable deterioration in model fit.

Latent growth curve modeling. We estimated individual LGCMs to represent change in achievement and ASC across the first three years in secondary school based on four waves of measurement. As indicators, we used person parameters from achievement test equating and – given the scalar measurement invariance of the ASC measure (see *Results*) – latent factor scores of ASC. With linear LGCM, two latent variables are estimated, namely a level growth parameter (i.e., intercept) and a growth rate parameter (i.e., linear slope). We also estimated quadratic growth curves that can identify more complex trajectories than linear growth curve models, as they include one extra quadratic growth curve parameter (i.e., the quadratic slope). The quadratic slope factor indicates the amount of acceleration or deceleration of the constant rate of change indicated by the mean of the linear growth curve parameter (Newsom, 2015).

The analyses were conducted in two steps. First, we fitted univariate LGCMs; that is, separate LGCMs for achievement and ASC. In each class type, we determined the adequate form of each growth trajectory by comparing the fit between two nested models, namely one specifying a linear form and the other specifying a quadratic form. Because we were interested in the *effects* of ability grouping, the last wave of measurement was chosen as the reference point, meaning that the intercept referred to the level of achievement or ASC in the middle of Grade 7. The unequal intervals between the different waves of measurement were specified in the model in order to obtain adequate estimations of the growth curve parameters. Second, we investigated the common development of achievement and ASC and their dependence on class type by a conditional dual-process LGCM (McArdle & Nesselroade, 2003). For all latent growth factors we used class type as a categorical predictor variable (regular class = 0; gifted class = 1). The repeated measures were represented by the following latent growth factors: an intercept, a slope, and if applicable, a quadratic factor; the different factors were allowed to co-vary with each other. To evaluate model fit, the following criteria were used: (a) the normed version of the chi-square statistic (Bollen, 1989), (b) the Root Mean Square Error of Approximation (RMSEA), (c) the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), and (d) the Standardized Root Mean Square Residual (SRMR; Hu & Bentler, 1999).

5.5. Results

Table 16. *Means and Standard Deviations for Academic Self-Concept (ASC) and Achievement (ACH; Person Parameters) by Class Type and Wave of Measurement (Matched Sample, Model 3)*

	Gifted Classes			**Regular Classes**			
	n	*M*	*SD*	*n*	*M*	*SD*	***d***
ASC1	138	4.236	0.819	137	4.245	0.751	-0.011
ASC2	133	4.105	0.865	132	4.098	0.806	0.008
ASC3	131	3.832	0.956	133	3.942	0.856	-0.121
ASC4	122	3.902	1.052	116	4.036	0.835	-0.141
ASC1_FS	148	0.342	0.973	147	0.267	0.848	0.082
ASC2_FS	148	0.094	0.996	147	0.112	0.961	-0.018
ASC3_FS	148	-0.213	1.124	147	-0.104	1.017	-0.102
ASC4_FS	148	-0.193	1.241	147	-0.021	0.955	-0.155
ACH1	138	-0.160	0.880	131	-0.233	0.705	0.092
ACH2	141	0.428	1.095	137	-0.056	0.913	0.480
ACH3	137	0.818	1.018	128	0.449	1.089	0.350
ACH4	117	0.945	1.177	116	0.555	0.821	0.384

Note. Numbers following ASC and ACH refer to wave of measurement. _FS = Latent factor scores. *d* = Standardized mean difference Cohen’s *d*.

Table 16 reports descriptive statistics as well as standardized mean differences for ASC manifest scale scores, ASC latent factor scores, and achievement scores (person parameters from achievement test equating) for the matched subsample by class type and wave of measurement. Because of the PS matching, at the first wave of measurement there were negligible mean differences between both class types (ASC scale score: d = -0.011; ASC factor score: d = 0.082; math achievement: d = 0.092). For the following three waves of measurement, mean differences in ASC were also negligible (*ds* between 0.008 and -0.141) while differences in achievement were of medium size (*ds* between 0.350 and 0.480) with higher levels in the gifted classes.

5.5.1. Results for Multilevel Analyses Testing the BFLPE

We investigated the relation of ASC with class-average math achievement (i.e., contrast effect) and class type (i.e., assimilation effect) controlling for individual math achievement, and – from the second wave of measurement on – for prior ASC (see Table 17 for results). Intraclass correlations at wave of measurement 1/2/3/4 for the ASC measure were .078/.076/.092/.054, respectively; for the math achievement test these correlations were .270/.365/.307/.290, respectively. Regarding the predictors at the individual level (Level-1), math achievement and prior ASC were always positively and significantly related to ASC. Regarding the predictors at class level (Level-2), our findings indicated neither significant contrast effects of class-average achievement nor significant assimilation effects of class type. That is, regarding the four waves of measurement there was no evidence for a negative BFLPE and no evidence for positive assimilation effects of gifted ability grouping in special classes.

Table 17. *Regression Coefficients from Multilevel Analyses for each Wave of Measurement (T1 – T4) relating Math Academic Self-Concept (ASC) to Math Ability, Prior Math Academic Self-concept, Class-Average Math Ability, and Class type (Gifted vs. Regular; Matched Sample, Model 3)*

	Math ASC T1			**Math ASC T2**			**Math ASC T3**			**Math ASC T4**		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.21	.08	.007	.28	.08	<.001	.23	.08	.002	.24	.05	<.001
Prior ASC_{T-1}	--	--	--	.44	.06	<.001	.54	.07	<.001	.61	.06	<.001
Class level												
Math ability	-.23	.21	.284	-.15	.17	.376	-.09	.13	.486	-.07	.12	.556
Class type[a]	.18	.20	.379	-.02	.17	.922	-.03	.13	.815	-.19	.18	.274

Notes. [a] Class type coded as 0 = regular classes vs. 1 = gifted classes.

5.5.2. Results for the Confirmatory Analyses

Measurement invariance testing. To prepare our longitudinal analysis, we investigated measurement invariance of the ASC measure over class type and time. Results are reported in Tables A.3 and A.4 in the Appendix C. The inspection of the model fit indices as well as the differences in these indices between models suggested that a scalar invariant model specification provided a good approximation of the data. This level of measurement invariance allows comparing latent factor means (Meredith, 1993). Subsequent analyses were therefore conducted with latent factor scores of ASC, which were derived from a 4-wave, multiple-group model (i.e., gifted vs. regular classes) assuming scalar measurement invariance.

ASC and achievement over time: Latent growth curve modeling. Initially, we determined the adequate form of the growth trajectories of ASC or achievement (linear or quadratic form) within each class type. Results are reported in the Appendix C (see Table C5 for model fit and parameter estimates). In regular classes, the χ^2-difference test for testing a LGCM with a linear factor against a LGCM with a quadratic factor was significant ($\Delta\chi^2$ = 12.729, $\Delta df = 4$, $p = .013$; see Table C5: ASC, Model B). Growth in achievement was sufficiently modeled as purely linear as indicated by the nonsignificant χ^2-difference test (gifted classes: $\Delta\chi^2 = 7.015$, $\Delta df = 4$, $p = .135$; regular classes: $\Delta\chi^2 = 4.752$, $\Delta df = 4$, $p = .314$; see Table C5: achievement, Model B).

Based on these findings, the conditional dual-process LGCM was conducted with modeling a quadratic growth trajectory for ASC, and a linear growth trajectory for achievement (see Figure 8 and Table 18 for results). The model showed a good fit, χ^2 = 27.681, $df = 19$, $p = .090$, RMSEA = .039 (90% CI .000-.069), CFI = .988, TLI = .977, SRMR = .033.

Regarding ASC, at the fourth wave of measurement the level of ASC was comparable over class type (i.e., there was no significant effect of class type on intercept ASC: $\beta = -0.199$, $p = .184$). ASC showed no significant linear decrease over time (slope linear ASC: regular classes $\beta = -0.003$, $p = .949$; gifted classes $\beta = -0.066$, $p = .302$). The variance of the quadratic growth factor was significant indicating a non-linear change of ASC over time. Change in ASC showed no significant relation with class type (class type on slope linear ASC: $\beta = -0.069$, $p = .355$; class type on slope quadratic ASC: $\beta = -0.004$, $p = .758$). That is, students in both class types showed comparable developmental trends in ASC over time.

Regarding achievement, at the fourth wave of measurement students in gifted classes showed significantly higher achievement levels (class type on intercept ACH: $\beta = 0.441$, $p = .016$). The average development of achievement was positive and significant in both class types (slope linear ACH: regular classes $\beta = 0.159$, $p < .001$; gifted classes $\beta = 0.224$, $p < .001$). Further, change in achievement was significantly related to class type (class type on slope linear ACH: β = 0.065, p = .010). That is, students in gifted classes both showed higher achievement levels in the middle of Grade 7 and a better development of achievement over time than comparable students in regular classes.

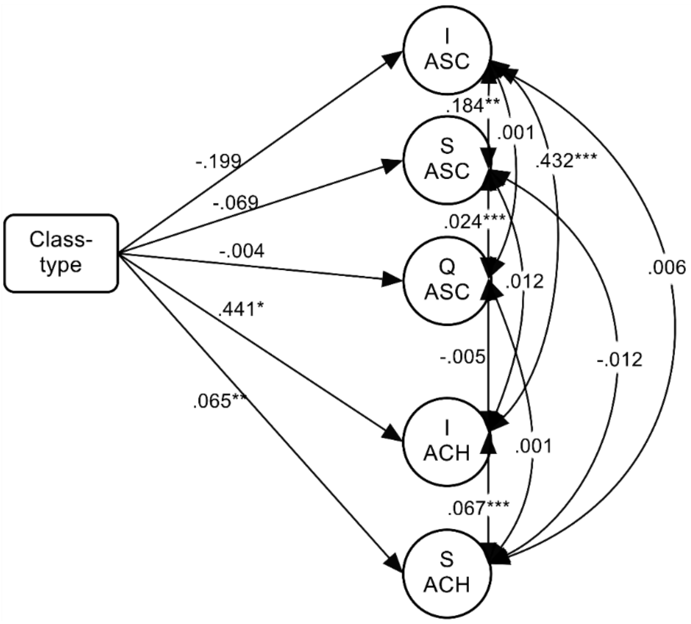


Figure 8. Results (unstandardized estimates) of the conditional dual-process LGCM with an intercept for ASC (I ASC), a linear slope for ASC (S ASC), a quadratic slope for ASC (Q ASC), an intercept for achievement (I ACH), a linear slope for achievement (S ACH), and class type (regular classes = 0, gifted classes = 1) as predictor (* $p < .05$. ** $p < .01$. *** $p < .001$)

Concerning the common development of ASC and achievement, two of the six correlations between intercepts and slopes were significant. Students with a higher ASC showed a higher level of achievement (intercept ASC with intercept ACH: r = .432, p < .001) and a better development of achievement (intercept ASC with slope linear ACH: r = .037, p = .003). The development of ASC was unrelated to the level of achievement at the last wave of measurement (slope linear ASC with intercept ACH: r = .012, p = .735; slope quadratic ASC with intercept ACH: r = -.005, p = .439) and the development of both constructs was independent of each other (no significant correlations between slopes). To conclude, higher levels of ASC at the fourth wave of measurement was related to higher levels of achievement at the fourth wave of measurement and a better development of achievement over time. Beyond that, both constructs were independent of each other. That is, changes in ASC were not related to changes in achievement, and vice versa.

Table 18. *Results of the Conditional Dual-Process Latent Growth Curve Model for Achievement (ACH) and Academic Self-Concept (ASC) (Matched Sample, Model 3)*

	Unstandardized Estimates		S.E.	
	Regular classes [a]	Gifted classes [a]	Regular classes [a]	Gifted classes [a]
Means				
Intercept ASC	-0.144, *p* = .081	-0.34, *p* = .006	.082	.126
Slope linear ASC	0.003, *p* = .949	-0.066, *p* = .302	.038	.064
Slope quadratic ASC	0.013, *p* = .058	0.009, *p* = .403	.007	.011
Intercept ACH	0.530, *p* < .001	0.972, *p* < .001	.111	.147
Slope linear ACH	0.159, *p* < .001	0.224, *p* < .001	.017	.018
Residual Variances				
Intercept ASC	1.066, *p* < .001		.126	
Slope linear ASC	0.181, *p* < .001		.046	
Slope quadratic ASC	0.004, *p* = .011		.001	
Intercept ACH	0.695, *p* < .001		.105	
Slope linear ACH	0.015, *p* = .005		.005	
Covariances				
Intercept ASC with Slope linear ASC	.184, *p* = .002		.059	
Intercept ASC with Slope quadratic ASC	.010, *p* = .294		.009	
Slope linear ASC with Slope quadratic ASC	.024, *p* = .001		.007	
Intercept ASC with Intercept ACH	.432, *p* < .001		.070	
Intercept ACH with Slope linear ASC	.012, *p* = .735		.034	
Intercept ACH with Slope quadratic ASC	-.005, *p* = .439		.006	
Intercept ASC with Slope linear ACH	.037, *p* = .003		.012	
Slope linear ASC with Slope linear ACH	.006, *p* = .419		.008	
Slope quadratic ASC with Slope linear ACH	.001, *p* = .713		.001	
Intercept ACH with Slope linear ACH	.067 , *p* < .001		.018	
Class type on ASC [a]				
Intercept ASC	-0.199, *p* = .184		.150	
Slope linear ASC	-0.069, *p* = .355		.074	
Slope quadratic ASC	-0.004, *p* = .758		.013	
Intercept ACH	0.441, *p* = .016		.184	
Slope linear ACH	0.065, *p* = .010		.025	

Note. [a] = Because of the use of the categorical predictor class type (0= regular classes, 1 = gifted classes) means of latent factors refer to the group coded as 0. Therefore, the model was tested twice, once with regular classes coded as 0 and once with gifted classes coded as 0. Estimates for (co-) variances are unaffected by this coding. The size of the effects of the categorical predictor class type is also unaffected by this coding but not the prefix; the prefixes reported here refer to the coding of regular classes as 0.

5.6. Discussion

Ability grouping has been the subject of heated debate, partially because empirical evidence for achievement gains seems to be limited. Further, the grouping of high-ability students has been criticized due to potentially negative reference group effects on students' ASC (i.e., BFLPE), which among other things should impair students' achievement development. When studying effects of ability grouping of high-ability students, selection bias presents a specific challenge and many studies in this field suffer from the lack of suitable control groups (Dai, Swanson, & Cheng, 2011). We investigated the effects of full-time ability grouping of high-ability students in gifted classes as compared to regular classes on students' ASC and achievement in math over a 3-year period at the beginning of secondary school. To control for selection bias when comparing students in both class types, we matched students by propensity score matching. By doing, our study makes three key contributions. First, it studies achievement and ASC simultaneously. Second, it employs a longitudinal design, which is rare in intervention studies, particularly in grouping studies. Third, the study contributes a new investigation to the vast, but aging research literature on ability grouping, employing more rigorous statistical controls and design features than many previous studies on the effects of ability grouping in schools.

Our main findings for our matched student sample were as follows: There was no evidence for significant negative reference-group effects or for significant positive assimilation effects of ability grouping on high-ability students' ASC. ASC remained stable over time in both class types. Further, ability grouping showed a significant and positive impact on students' achievement development. Students in gifted classes showed a significantly better development of achievement over time and, accordingly, a higher level of achievement at the last wave of measurement than comparable students in regular classes. Higher levels of ASC at the fourth wave of measurement related to higher levels of achievement and a greater development of achievement over time. However, higher levels of achievement at the fourth wave of measurement were not significantly related to changes in ASC. Moreover, changes in ASC were independent of changes in achievement, and vice versa.

5.6.1. Limitations

Before discussing these findings further, we would like to point to some limitations of our study. Our sample only comprised students from the highest academic track of the German secondary school system. In comparison to the other scholastic tracks, the Gymnasium differs with respect to school quality and learning environment, with more cognitively demanding and activating instruction (Becker, Lüdtke, Trautwein, Köller, & Baumert, 2012). Generalizability of our findings to other school systems therefore warrants further investigation.

Moreover, our sample only included schools that offered both regular and gifted classes. Thus, it remains an open question if the within-school comparisons used in the present study would generalize to between-school comparisons. Students typically use their classmates as a reference group (Wouters et al., 2013). Thus, the class context seems more relevant to students' social comparisons than the school context. However, one cannot rule out the possibility that the presence of gifted classes at the schools affected the self-perceptions and achievements of students in regular classes and vice versa. We did not have any schools without gifted classes in our study, and therefore could not examine this research question. Our matching design, on the other hand, did allow us to control for the influence of each particular school (see e.g., Hattie, 2009, for school effects).

Further, we investigated our research question in the domain of mathematics only. Although the data of our study stemmed from a project with a rich data base, only the math achievement tests could be equated over time. Thus, our findings cannot be generalized to other achievement domains without further investigation. However, most of the published research on reference group effects on ASC has used the domain of mathematics (Herrmann et al., 2016; Chiu, 2012), which enhances the comparability of our findings with those of other studies.

Finally, although we used propensity score matching we could not control for all possibly relevant sample differences. In applied research such as in our study, there is no experimental control and no way to randomly assign students to class type. However, by using propensity score matching we could control for some bias a posteriori. Furthermore, our matching strategy resulted in the exclusion of the most capable students in the gifted classes because there were not enough students in the regular classes showing similarly high IQs. On average, the students selected by the matching were lower in cognitive ability by three IQ-points (small effect). Further, they were about three month older (small effect size). We conducted robustness checks and our findings were stable across multiple matching solutions including a model in which no student was discarded from the gifted classes (Model 1, Table C1 in the Appendix C). However, there was one notable exception: For Model 1, which included all students from gifted classes, the effect of class type on the development of math achievement was not significant (p = .063; see Appendix D, Table D5). This indicates, that the positive achievement effects only hold for some part of the distribution, namely the matched students from gifted classes who were marginally older and somewhat lower in intelligence than all students from gifted classes. However, these differences were of small effect size.

5.6.2. Ability Grouping and its Effects on the Development of Academic Self-Concept and Achievement

We found very little evidence for specific costs or gains of grouping high-ability students regarding their ASC. In our matched sample, mean differences in ASC between class types were negligible and class type had no significant effect on the level or development of ASC. There was neither evidence for significant negative contrast effects of class-average achievement nor for positive assimilation effects of class type. Rather, in both class types ASC stayed stable over time. Our findings regarding the ability grouping of gifted students offer no support for the BFLPE and no support for the notion of a general decline in students' academic self-perceptions after the transition to secondary school (e.g., Wigfield & Wagner, 2005). This decline has been frequently explained by psychological factors such as an increasingly realistic self-concept and by contextual factors such as stricter evaluative feedback, less teacher support, or a stronger performance orientation in secondary school as compared to elementary school (e.g., Gest et al., 2015). However, independent of class type, high-ability students in our sample did not show this decline in academic self-concept. Given the design of our study, it cannot be ruled out that students experienced a decline in ASC during the first month in secondary school. As noted above, our first wave of measurement took place approximately one month after the students' transition to the new school. However, the BFLPE is also assumed to operate in stable environments (Marsh, 1987) and longitudinal studies offer empirical evidence for negative reference-group effects in stable environments (e.g., Marsh, 1994; Marsh, Kong, & Hau, 2000). Therefore, the finding of no significant changes in ASC from the first month in secondary school until the middle of Grade 7, which covers a period of two and a half years, is noteworthy. It indicates that highly able students in both regular and gifted classes do not suffer in their academic self-perceptions in the first years in secondary school, possibly because they receive more positive evaluative feedback, and possibly because an increasingly realistic self-concept would mirror their high ability. Our findings are well aligned with those of other studies that showed that the ASC of high-achieving students appears to be less affected by reference-group effects than that of lower achieving students (Marsh & Rowe, 1996; Trautwein, Lüdtke, Marsh, & Nagy, 2009). However, our findings need replication in other school settings and achievement domains.

Controlling for achievement differences between class types at the beginning of Grade 5, we found that in Grade 7 students in gifted classes showed significantly higher levels of achievement than comparable students in regular classes. That is, students in gifted classes showed greater achievement gains over time. Because of our matching procedure, these achievement gains cannot be explained by a priori differences between students but rather point to an impact of class type. One crucial element of gifted education (or education in general) is appropriate developmental placement that places gifted students in learning environments that correspond to their abilities and preferences (Lubinski & Benbow, 2000). The gifted classes in our study were characterized not only by ability grouping but they were also specifically designed to serve gifted students with high achievement levels and high learning potential. That is, the specific curriculum offered more differentiation, problem-based learning, independent study, enrichment, and acceleration (Schneider et al., 2014). Research on gifted education supports the effectiveness of these instructional strategies (e.g., Rogers, 2007; see, Robinson, Shore, & Enersen, 2007, for an overview). The ability grouping of gifted students in special classes allows students to progress at a higher speed. It offers advantages for teachers because they know in advance about the potential of their students. This increases the likelihood that teachers develop high achievement expectations for their students and that they more effectively instruct students in their zone of proximal development (Tieso, 2005). All teachers in our study did not only teach in gifted classes but also taught in regular classes. However, it is very likely that teachers were able to influence in which class type they taught. Thus, it cannot be excluded that more motivated teachers taught in gifted classes. All of these factors increase the probability of more favorable achievement trends.

5.6.3. Conclusion

Our findings in the domain of mathematics offer further empirical evidence that ability grouping combined with a specific curriculum is beneficial for the development of gifted students: we found gains for the development of their achievement and no specific costs for the development of their academic self-perceptions. Even after applying a strict matching procedure, we found higher achievement gains for high-ability students in gifted classes than in regular classes. These achievement gains cannot be explained by a priori differences between students from both class types but rather point to a positive impact of ability grouping in gifted classes. We found no evidence for specific costs for students' academic self-perceptions due to negative reference-group effects, which in turn could impair students' achievement development. Rather, the development of achievement and academic self-perceptions were independent of each other. Thus, our findings for high-ability students only partly support the reciprocal effects model of ASC and achievement (Marsh & Martin, 2011). Students in both class types had comparably high academic self-perceptions, which were stable over time.

Chapter 6. Discussion

Besides student cognitive ability, motivational variables are central to the prediction of achievement-relevant outcomes in school (e.g., Steinmayr & Spinath, 2009; Wigfield & Eccles, 2002). As the predictive power of cognitive ability has been very well established and precisely estimated (Roth et al., 2015), a trend towards investigating alternative predictors of achievement has been observed in current educational research (Grigorenko, 2016). Grigorenko (2016, p. v) went so far as to describe this trend as:

“The King is dead, long live the King!”

By the old king she meant all ability-driven factors such as the g-factor. By the new king she meant all other constructs that predict educational outcomes beyond cognitive ability. Undoubtedly, motivational variables are a major part[11] of the new king (e.g., Steinmayr & Spinath, 2009). In this dissertation, questions regarding the new king’s well-being were raised. Does the royal court treat the king majestically and does the king feel well during students’ school careers? In other words, how does student achievement motivation develop during the school career (research question 1)? Which circumstances anger the king and what happens when the king becomes mad? In other words, which variables predict achievement motivation development (research question 2a) and which changes are associated with changes in achievement motivation (research question 2b)? Finally, does “divide and conquer” work for the king? That is, does ability grouping affect student motivational development (research question 3)?

Three articles aimed at addressing these questions were presented in Chapters 3, 4, and 5. A short summary of each article is provided in the following. Then, the key findings regarding the three research questions are discussed and embedded in previous theoretical and empirical research. Based on these findings, conclusions regarding theoretical and practical implications are made. Finally, strengths and limitations are weighed against one another, subsequent research is suggested, and a short outlook is given.

[11] Prior knowledge has also been discussed as a further major factor that is able to predict educational achievement outcomes (e.g., Schneider, Körkel, & Weinert, 1989; Stern, 2001).

6.1. Summary of the Findings

6.1.1. Summary of Article 1

In Article 1, a meta-analysis of a total of 107 independent longitudinal studies was presented. In each of the considered longitudinal studies, school-aged students reported on their self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, or achievement goals at at least two different time points during their school careers. Average mean-level changes (Glass's Δ) in these motivational constructs were conducted using meta-regressions, while the dependence of effect sizes within studies was controlled by Robust Variance Estimation (Hedges, Tipton, & Johnson, 2010). Taking all constructs together, an overall decrease of Glass's $\Delta = -.11$ was found over an average duration of 1.65 years. Significant differences between the considered constructs were noted: In self-esteem, self-efficacy, and performance-avoidance goals, no significant mean-level change was found. In academic self-concept, meta-regressions indicated decreases in math and language self-concepts, but no significant change in the general self-concept (math: Glass's $\Delta = -.18$ over an average duration of 1.65 years; language: Glass's $\Delta = -.15$ over an average duration of 1.20 years). Declines were found in intrinsic motivation, mastery goals, and performance-approach goals, respectively (intrinsic motivation: Glass's $\Delta = -.19$ over an average duration of 1.66 years; mastery goals: Glass's $\Delta = -.20$ over an average duration of 1.10 years; performance-approach goals: Glass's $\Delta = -.14$ over an average duration of 1.09 years).

School transition, school type (elementary, middle, or high school), grade level, academic domain (general, math, or language), questionnaire, duration of the interval, year of data collection, year of publication, and geographic location were tested as moderators of the mean-level development. School transition and school type were not related to the mean-level development of any construct. Grade level only moderated the mean-level development of mastery goals (mastery goals decreases became smaller with rising grade level). Academic domain was only relevant for the development of academic self-concept (math and language academic self-concepts decreased, while no change was found in general academic self-concept). Similarly, the usage of a questionnaire was only related to the mean-level development of academic self-concept (larger declines in studies that used the scales by Eccles et al., 1983; smaller declines in studies that used the scales by Harter, 1985). The duration of the intervals was associated with the mean-level development of academic self-concept, intrinsic motivation, and mastery achievement goals (a longer duration was related to larger declines). The year of data collection was only related to the mean-level development of performance-approach goals (smaller declines in older data). Year of publication was only related to the mean-level development of intrinsic motivation (larger declines in older publications) and performance-approach goals (smaller declines in older publications). Finally, when all constructs were taken together, larger declines were found in Europe than in Asia and North America.

6.1.2. Summary of Article 2

In Article 2, two longitudinal studies with German adolescents from the highest track of the German three-track secondary school system (i.e., Gymnasium) were reported. In Study 1, 745 students in grades 5-7 were investigated in four waves. In Study 2, 1420 students in grades 5-8 were assessed in four waves. In both studies, the (joint) development of achievement goals (mastery, performance-approach, and performance-avoidance goals), interest, and achievement (grades in Studies 1 and 2; achievement test scores in Study 1) were investigated using growth curve modeling. Cognitive ability, sex, parents' highest educational degree, and language background were investigated as additional predictors of motivational development[12].

Findings regarding mean-level development were consistent across the two studies. Mean levels of mastery, performance-approach, and performance-avoidance goals as well as interest all decreased over time. In addition, the declines of all four constructs decreased with rising grade level in Study 1. In Study 2, a similar trend was observed for mastery goals. In both studies, mastery and performance-approach goal change was positively related to change in interest. Change in performance-avoidance goals was either unrelated (Study 1) or negatively related (Study 2) to change in interest. Regarding the achievement variables, change in mastery goals showed either no relation (Study 1) or a positive relation (Study 2) and change in performance-approach goals was not associated with change in achievement. Change in performance-avoidance goals was either unrelated (Study 1) or negatively related (Study 2) to change in achievement. Regarding the predictors of motivational development, mastery goals development was not predicted by any variable except for language background in Study 2 (students with a German background showed smaller declines in mastery goals than students with another language background). Performance-approach goals development was positively predicted by a higher parental education level in both studies and additionally predicted by sex and language background in Study 2 (girls and students with a German language background showed smaller declines than boys and students with another language background). Performance-avoidance goals development was not predicted by any variable

[12] These variables were considered as control variables and were not the focus of Article 2.

except for sex in Study 1 (girls showed smaller declines than boys). Finally, interest development was not predicted by any variable except for parental education level in Study 1, which was a positive predictor.

6.1.3. Summary of Article 3

In Article 3, effects of ability grouping were investigated in a longitudinal study of high-ability students from the highest middle school track in Germany (i.e., Gymnasium). The sample of grade 5-7 students was assessed in four waves and comprised 148 students from regular classes and 148 students from special classes for gifted students. These students were selected from a larger dataset by using propensity score matching. During the matching process, a priori differences (i.e., T1 differences) between these students in numerous control variables (e.g., differences in cognitive ability, age, and parents' highest educational degree) were minimized. Latent growth curve modeling was applied to investigate student mean-level development in math self-concept and achievement (standardized math tests). In addition, big-fish-little-pond and assimilation effects were investigated using multilevel analyses.

Latent growth curve models indicated no significant change of student self-concept over time in regular or special classes. Thus, grouping was not related to mean-level development of academic self-concept. Student achievement increased over time and increases in special classes were larger than in regular classes. Interestingly, change in academic self-concept was not related to change in achievement. Finally, multilevel analyses revealed no significant effects of class average achievement or class type (regular versus special classes) on student self-concepts at any time point. Thus, neither a big-fish-little-pond effect nor an assimilation effect was evident for the duration the study.

6.2. Discussion of the Key Findings

6.2.1 Research Question 1: How Does Student Achievement Motivation Develop During the School Career?

Research question 1 refers to general achievement motivation development as well as the development of particular motivational constructs. According to stage-environment fit theory (Eccles et al., 1993), a mismatch between student needs and opportunities in school causes declines of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, interest, and mastery goals. Student mean-level development was investigated in all three presented articles.

Regarding general achievement motivation development, the meta-analysis in Article 1 tested overall change of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals. When all constructs were taken together, a small decline was found. This finding is in accordance with the general assumption of stage-environment fit theory (Eccles et al., 1993). Note that in this meta-analysis the magnitude of the decline varied by particular construct.

For example, in the meta-analysis in Article 1, no mean-level change was observed for student self-esteem over the course of the school career. This finding does not support the assumption that a mismatch between student needs and opportunities in school affects general self-esteem. This finding has also been confirmed by a further meta-analysis recently conducted by Orth, Erol, and Luciano (2018). Orth and colleagues (2018) investigated the development of self-esteem from age 4 to 94 and found that self-esteem increased from age 4 to age 11, remained stable from age 11 to age 15, and increased from age 15 until age 60. Although the age groups of the two meta-analyses are not fully comparable (i.e., the meta-analysis presented in Article 1 was limited to school-aged students), both meta-analyses showed a lack of self-esteem declines over the course of students' school careers.

Student academic self-concept was investigated in the meta-analysis reported in Article 1 and in the longitudinal study reported in Article 3. The meta-analysis revealed decreases of domain-specific self-concepts (i.e., math and language), but no change of general self-concept. The longitudinal study showed no change of (math) self-concept.
Thus, only the meta-analytic findings regarding domain-specific self-concept were in accordance with stage-environment fit theory. The differences between domain-specific and general self-concepts found in the meta-analysis can plausibly be explained by age-related differentiation processes of academic self-concept. Note that young students' competence beliefs do not much differ between domains, whereas, with rising age, students tend to assume that they are competent in some, but less competent in other domains (Denissen, Zarrett, & Eccles, 2007; Marsh, Craven, & Debus, 1998; Schmidt et al., 2017). That is, with rising age, internal comparison processes (e.g., How good am I at math in comparison to German?) become more important for the formation of student academic self-concepts. Internal comparison processes can lower self-concept regarding the intraindividually weaker domain and strengthen self-concept regarding the intraindividually stronger domain (Möller & Marsh, 2013). Note that general self-concept is typically unaffected by internal comparison processes (Schmidt et al., 2017). Thus, the decreases in domain-specific self-concepts could be explained by the fact that students carry out more internal comparison with rising age. At the

same time, the absence of decreases of general self-concept is explained by the fact that internal comparison processes do not affect general self-concept. Finally, it is noteworthy that in the longitudinal study reported in Article 3, no decrease in math self-concept was found. The most plausible explanation of this finding is that the examined sample comprised exclusively high-ability students. It is possible that highly able students suffer less from evaluative middle school environments, because they show better achievement and consequently receive more positive evaluative feedback than regular students.

Development of academic self-efficacy was investigated in the meta-analysis presented in Article 1. In disagreement with stage-environment fit theory, no significant change could be observed in this meta-analysis. Theoretically, this finding can be explained by the fact that academic self-efficacy refers to specific tasks and thus is less affected by a general stage-environment mismatch than broader constructs, such as academic self-concept. In addition, in the meta-analysis, only eight independent longitudinal studies examined self-efficacy, resulting in restricted test power.

Student intrinsic motivation was investigated in the meta-analysis and student math interest was investigated in the two longitudinal studies presented in Article 2. In accordance with stage-environment fit theory, the meta-analysis and the two longitudinal studies revealed declines in student intrinsic motivation and interest, respectively. Note that the average magnitude of the declines quantified by the meta-analysis was small (i.e., on average intrinsic motivation decreased by .19 *SD* over 1.66 years). Thus, in order to adequately depict small changes in intrinsic motivation, large sample sizes are required in future research. Similarly, researchers should not expect notable changes when examining short time intervals.

Regarding student achievement goals development, different predictions can be derived based on stage-environment fit theory. That is, a shift from mastery to performance goals (approach and avoidance), decreases in all three goals (i.e., a quantitative decline), or decreases in mastery and performance-approach goals and increases in performance-avoidance goals (i.e., a qualitative decline). Achievement goals development was investigated in the meta-analysis presented in Article 1 and in the two longitudinal studies reported in Article 2. Regarding mastery and performance-approach goals, decreases were observed in the meta-analysis and the two longitudinal studies, respectively. Regarding performance-avoidance goals, no significant change was found in the meta-analysis, whereas both new longitudinal studies observed declines. Note that in the meta-analysis, performance-avoidance goals descriptively showed even larger declines than performance-approach goals, however these changes did not reach significance. This could be explained by lower test power when examining performance-avoidance goals compared to the other goals (i.e., only 9 independent longitudinal studies investigated performance avoidance goals, whereas 19 and 16 independent longitudinal studies investigated mastery and performance-approach goals, respectively). Results of the meta-analysis and the two longitudinal studies unambiguously reject the assumption that student mastery goals shift to performance goals (approach and avoidance). Although the concept of a qualitative decline cannot be completely ruled out, the results tend to support a quantitative decline. That is, it seems like the mismatch between student needs and opportunities might lead to a drop in all forms of competence-based pursuit (i.e., mastery, performance-approach, and performance-avoidance goals).

6.2.2. Research Question 2: What Are the Correlates of Achievement Motivation Development?

Research question 2 refers to possible predictors of motivational development (a) and to constructs whose development is associated with motivational development (b). Regarding research question 2a, numerous predictors of motivational development were investigated in the meta-analysis presented in Article 1 and the two longitudinal studies reported in Article 2. According to stage-environment fit theory, the mismatch between student needs and opportunities in school primarily occurs after the transition to middle school. Surprisingly, the meta-analysis indicated no effects of school transition or school type (elementary, middle, or high school) on motivational development (i.e., development of self-esteem, academic self-concept, academic self-efficacy, intrinsic motivation, and achievement goals). This finding does not contradict the general assumption that the mismatch between student needs and opportunities causes motivational declines, as motivational declines were found (at least in some constructs). However, it also does not support the assumption that the mismatch occurs primarily during and after school transition. On the one hand, it is possible that stage-environment fit differences between elementary, middle, and high school have been minimized as a function of administrative changes in the educational system (e.g., Au, 2011; Borman, Hewes, Overman, & Brown, 2003; Ertmer & Ottenbreit-Leftwich, 2010). On the other hand, the meta-analysis and both longitudinal studies indicated that decreases of mastery goals became smaller with rising grade level. As the two longitudinal studies investigated students at the beginning of middle school, sinking decreases may be seen as concordant with stage-environment fit theory (i.e., initial decreases after the school transition fade out over time). In addition, it must be noted that only a few studies included in the meta-analysis focused on school transition. Thus, test validity regarding transition and school type was limited (i.e., in the meta-analysis, declines during the school transition were compared to declines during other school stages). More primary studies that examine the same students before and after a school transition are required to test whether the stage-environment mismatch primarily occurs after school transition.

Interestingly, the meta-analysis showed that older publications reported larger declines in intrinsic motivation than more recent publications. This finding supports the assumption that administrative changes in the educational system might have minimized the stage-environment mismatch. In addition, the meta-analysis indicated larger performance-approach declines in recent publications than in older publications. This finding is interesting as researchers formerly expected a shift from mastery to performance goals over the course of the school career, whereas current research rather supports a quantitative decline of achievement goals (see section 6.2.1). Thus, it is possible that decades ago the school environment was beneficial for student performance goals, whereas today, student performance goals decline over the course of the school career. Another interesting finding from the meta-analysis is that student motivational declines were larger in Europe than in North America or Asia. This finding can be explained by varying school politics between countries (e.g., school-tracking; Schnabel et al., 2002; Chmielewski, Dumont, & Trautwein, 2013; Nagy et al., 2010) and culture-specific ideologies (Tweed & Lehman, 2002).

In addition, the two longitudinal studies reported in Article 2 showed that sex, parents' highest educational degree, and students' language background were predictive of student development of achievement goals and interest. Girls showed smaller declines of performance-approach and performance-avoidance goals than boys. This rather surprising finding could be explained by the fact that boys started with significantly higher performance-approach and performance-avoidance goals (i.e., boys had greater potential for performance goal declines than girls). Parents' highest educational degree was positively associated with development in performance-approach goals and interest. It is plausible that parents who achieved a high educational degree themselves encourage their children to enjoy academic activities as well as to be better than others in school. Furthermore, native German speakers showed smaller declines in mastery and performance-approach goals than students with other language backgrounds. It is possible that the stage-environment mismatch is less destructive for native speakers, because they can more quickly adjust to the evaluative middle school environment. Finally, in both longitudinal studies, student cognitive ability was unrelated to motivational development. This finding is consistent with the assumption that cognitive ability and achievement motivation do not share common variance (Bipp, Steinmayr, Spinath, 2008). Thus, it makes perfect sense to investigate both constructs as complementary factors that are both important on their own.

Regarding research question 2b, the two longitudinal studies reported in Article 2 investigated the joint development of achievement goals and interest or achievement and the longitudinal study presented in Article 3 tested the joint development of academic self-concept and achievement. According to achievement goal theory, interest and achievement are typical outcomes of achievement goals (Elliot & Hulleman, 2017). More specifically, mastery goals are thought to have positive effects, performance-approach goals are posited to have some positive and some negative effects, and performance-avoidance goals are assumed to have negative effects (Elliot & Hulleman, 2017). Both longitudinal studies presented in Article 2 revealed positive relations between development of mastery and performance-approach goals and development of interest. Development of performance-avoidance goals, was negatively related to development of interest in one study and unrelated in the other study. Regarding the development of achievement, positive or null relations were found for the development of mastery goals, null relations observed for the development of performance-approach, and negative or null relations were noted for the development of performance-avoidance goals. Overall, these results are mostly in accordance with achievement goal theory and illustrate the importance of mean-level changes of motivational constructs. That is, students that show larger declines of mastery and performance-approach goals are also more likely to show larger declines of interest and achievement than students that show smaller declines or increases of these goals.

As reciprocal relations between academic self-concept and achievement were found in previous research (Marsh & Craven, 1997, 2007; Preckel et al., 2017), a joint development of these variables could be also expected. Surprisingly, no correlated change between academic self-concept and achievement was found in the longitudinal study reported in Article 3. Note that Article 3 does not focus on the investigation of joint development. Furthermore, as only high-ability students were investigated in this study, this null finding can be explained by restricted variance in student academic self-concept and achievement (i.e., average academic self-concept and achievement were very high).[13]

[13] Note that Preckel and colleagues (2017) observed a reciprocal effect between academic self-concepts in regular as well as in gifted students. Thus, restricted variance in gifted students did not mask the reciprocal effects in their investigation.

6.2.3. How Does Ability Grouping Affect Student Motivational Development?

The longitudinal study reported in Article 3 investigated the effects of ability grouping on student development of academic self-concept and achievement in a sample of high-ability students. In general, positive effects of ability grouping can be assumed based on stage-environment fit theory. That is, grouping students regarding their ability levels could ensure a better fit between student needs and their learning environment (e.g., an adapted level of task difficulty). However, regarding student academic self-concept, positive (assimilation) as well as negative (contrast) effects of grouping can be assumed for high-ability students based on the “big-fish-little-pond-effect” (Marsh & Parker, 1984).

In the longitudinal study presented in Article 3, neither negative contrast effects, nor positive assimilation effects of ability grouping on student academic self-concept were found. Thus, this study gained no support for the “big-fish-little-pond-effect”. Furthermore, the development of student academic self-concept was not affected by ability grouping. On the one hand, this finding seems to contradict the assumption that a better person-environment fit prevents motivational declines. One the other hand, no declines of academic self-concept, regardless of the grouping, were found (i.e., no declines in regular or special classes, see section 6.2.1). It seems as though the investigated high-ability students in both class types did not suffer from the evaluative middle school environment. Therefore, the question whether ability grouping could prevent (potential) motivational declines in those cases in which students do suffer from the new school environment remains. In addition, a positive effect on student achievement development was found. That is, high-ability students in special classes showed stronger increases in achievement than high-ability students in regular classes, whereas no differences in academic-self-concept development were noted. Overall, the new findings support the utility of high-ability grouping, as gains in student achievement development and no “cost” for student self-concepts were found.

6.3. Conclusions

To sum up, in the presented research, achievement motivation slowly declined over the course of students’ school careers and the magnitude of the decline differed by construct. School stage and school transition were not related to the magnitude of the decline. In part, declines were predicted by sex, parents’ highest educational degree, and students’ language background. Declines in achievement goals went along with declines in interest and achievement. Finally, high-ability grouping had no “costs” for student self-concept development, but fostered gains in student achievement.

6.3.1. Theoretical Conclusions.

Several theoretical conclusions can be derived from these findings. The assumption of declining motivation that is in accordance with several theoretical models (i.e., cognitive-developmental approach, environmental influences, stage-environment fit theory, and self-determination theory) was confirmed (for the first time) by the meta-analysis of previous longitudinal studies and by two new longitudinal studies. For example, motivation declines over the course of students' school careers can be plausibly explained by a growing gap between students' rising needs and opportunities in school (i.e., stage-environment fit theory; Eccles et al., 1989; 1993). However, some assumptions of stage-environment fit theory were not supported by the findings.

Stage-environment fit theory offers the same prediction for the development of various motivational constructs (i.e., a general motivational decline; Eccles et al., 1989; 1993). However, different development trends were found for different constructs in the meta-analysis. General constructs that are less related to school (i.e., self-esteem) and specific constructs that focus on particular school tasks (academic self-efficacy) seemed to be less affected by the stage-environment mismatch than constructs that refer to school as a whole or particular school domains (i.e., academic self-concept, intrinsic motivation, achievement goals). Furthermore, it was found that student general self-concept suffered less than self-concept in particular domains and that high-ability students' self-concept seemed to be scarcely affected by the evaluative middle school environment. Regarding achievement goals, it turned out that the former assumption of a shift from mastery to performance goals over the course of the school career in accordance with stage-environment fit theory (Anderman & Midgley, 1997; Eccles et al., 1993), is obsolete. The findings rather suggest a decline in all achievement goals, indicating a quantitative decline of competence-based goal pursuit. From a broader perspective, these findings are in agreement with the fundamental premise of the stage-environment fit, which states a rising stage-environment mismatch is detrimental to student motivational development. However, the understanding of what detrimental means may need to be reconsidered. That is, declines of all forms of goal pursuit might be seen as even worse for student development than a shift from mastery to performance goals. Regarding intrinsic motivation and interest, the findings clearly indicated decreases that are in accordance with stage-environment fit theory. Although extrinsic motivation was not investigated in the presented articles, the theoretical idea of a quantitative decrease could be assumed for intrinsic as well as extrinsic motivation development (i.e., decreases in extrinsic motivation in addition to the found decreases in intrinsic motivation). In accordance with self-determination theory (Ryan & Deci, 2000a; 2000b), this development trend could be seen as *amotivational* development.

The finding that the development of some constructs was predicted by students' sex, parents' highest educational degree, and students' language background could be additionally integrated into the stage-environment fit framework as well as further possible predictors of motivational development. That is, it warrants further investigation which interpersonal or family factors can help students cope with the stage-environment mismatch in school.

The found joint development of achievement goals, interest, and achievement is in accordance with achievement goal theory (Elliot and Hulleman, 2017) and illustrates the importance of motivational development. That is, decreases in mastery and performance goals goals go along with decreases in interest and achievement. Furthermore, these findings also support the concept of a quantitative decrease. That is, it seems like some motivational constructs decrease simultaneously (i.e., all achievement goals, intrinsic motivation, and interest), whereas other constructs are not a part of this quantitative decrease (i.e., self-esteem, academic self-efficacy).

The findings regarding ability grouping provided evidence of its utility for gifted students. Gifted students that were grouped together (combined with a specific curriculum) showed stronger achievement gains than gifted students in regular classes. For student academic self-concept, no advantages or disadvantages of ability grouping were found.

6.3.2 Practical Implications.

The main findings are sobering. Student achievement motivation declines with each additional school year – slowly but continuously. Fortunately, student achievement motivation is open to various types of interventions that evidently foster positive motivational development (Bernacki, Nokes-Malach, Richey, & Belenky, 2016; Eccles & Roser, 2009; Lazowski & Hulleman, 2016). Thus, an obvious practical implication would be to apply interventions that foster student achievement motivation?

According to stage-environment fit theory, the mismatch between student needs and opportunities that causes motivation declines can be avoided on several ecological levels (i.e., classroom, school, school district, and community level; Eccles & Roser, 2009). At the lowest ecological level, teachers' general sense of efficacy and expectancies of student achievement, classroom climate, and teacher-student relationships are central variables that might prevent student achievement motivation from declining. Thus, to improve the lower ecological levels, intervention programs that target teachers' efficacy or social relatedness have been suggested (e.g., Nunn & Jantz, 2009). At a higher ecological level, school start time can be considered for a possible intervention. With rising age, most students develop a preference towards evening-orientation, whereas, in most countries, school start time stays constant throughout the entire course of students' school careers (Roennenberg et al., 2007). Student preference towards eveningness is negatively connected to achievement (Preckel et al., 2011) and motivational variables (Preckel et al., 2013; Scherrer, Roberts, & Preckel, 2016). Beginning school later with rising grade level could ensure a better fit between student preferences and the school environment.

In addition, specific intervention programs regarding particular motivational constructs that were found to decrease have been suggested (e.g., academic self-concept: Craven, Marsh, & Raymound, 1991; intrinsic motivation: Guthrie et al., 2006; achievement goals: Bernacki, Nokes-Malach, Richey, & Belenky, 2016). For example, Bernacki, Nokes-Malach, Richey, and Belenky (2016) instructed students to write self-assessments of their interest and understanding in science lessons. The results indicated that the intervention improved student mastery goals and interest. Such intervention programs can be recommended based on our results, as they foster a positive development of mastery goals and additionally ensure gains in further academic outcomes such as interest and achievement.

6.4. Strengths and Limitations

The three presented articles have both strengths and limitations. The fact that numerous motivational constructs that are covered in several motivational theories (see section 2.1) were investigated in this dissertation is a big strength. Thereby, the results can be discussed against different theoretical approaches (e.g., achievement goals theory, self-determination theory). Furthermore, similarities and differences of the particular constructs could be compared. However, the articles were unable to cover the full variety of motivational constructs that were discussed in previous research with regard to motivational decreases and stage-environment fit theory (Cassady & Johnson, 2002; Eccles et al., 1993; Wigfield & Eccles, 2002). That is, the present investigations need to be extended by examining further important constructs such as extrinsic motivation, amotivation, attainment value, utility value, perceived costs, and test anxiety. Similarly, the dissertation covered several domains of motivational constructs (i.e., general, language, and math domains were investigated in the meta-analysis). Again, additional domains such as sport or music that were not covered by the dissertation need to be investigated in subsequent research. Note also that the three new longitudinal studies presented in Articles 2 and 3 exclusively referred to the math domain. Accordingly, the significance of these studies is limited to the math domain.

A further strength of the dissertation is that a large body of previous longitudinal research could be integrated and an average magnitude of motivation change could be quantified by the conducted meta-analysis. Furthermore, several moderators of development were investigated in this meta-analysis. However, two limitations regarding the moderator analyses must be mentioned. First, for some motivational constructs, only a few independent studies could be assigned to the particular moderator categories. Therefore, the test power of the moderator analyses was limited regarding particular motivational constructs. Second, most of the moderator analyses conducted in the meta-analysis refer to the differences of reported motivational development between studies. Differences between persons within particular studies and within persons were almost completely ignored by these analyses. To some extent, this limitation is compensated by the two longitudinal studies reported in Article 2, in which factors that varied between persons (i.e., cognitive ability, sex, parents' highest educational degree, and language background) were used to predict interindividual differences in intraindividual motivational development. However, these factors were simply used as control variables and additional predictors of motivational development may be suggested based on different theories. For example, based on self-determination theory, differences in needs for autonomy, achievement, and competence could be suggested as possible predictors of motivational development.

The found joint development of achievement goals, interest, and achievement in Article 2 demonstrated why motivation development matters. However, the conducted analyses of correlated change do not allow statements about the causal order of the investigated variables. To compensate for this weakness, reciprocal effect models were additionally conducted in the studies reported in Article 2. For the most part, these analyses revealed that the relations between achievement goals, interest, and achievement were reciprocal.

Regarding the investigation of ability grouping, two strengths of Article 3 can be highlighted. First, a longitudinal design was used, which has rarely been applied in previous ability grouping research. Second, a priori differences between students in special and regular classes were minimized by using propensity score matching. Thus, for the large part, the differences in student development that were found can be attributed to ability grouping. Furthermore, two variables that are central to "big-fish-little-pond-effect" research (i.e., academic self-concept and achievement; Marsh & Parker, 1984) were investigated. However, when considering ability grouping with regard to stage-environment fit theory (i.e., as a particular approach to reduce the person-environment mismatch), additional motivation variables could also be taken into account. That is, further research could focus on the question of how ability grouping affects the development of self-esteem, academic self-efficacy, intrinsic motivation, interest, and achievement goals.

A further weakness is that all three articles focused on student achievement motivation that was assessed exclusively by questionnaire-based student self-reports. Self-reports are a common method to capture individuals' feelings, thoughts, beliefs, attitudes, states, and traits (Baumeister, Vohs, & Funder, 2007). However, methodological limitations of this method such as common method variance and socially desirable or consistent responses must be considered when interpreting the results (e.g., Podsakoff & Organ, 1986).

Finally, although this dissertation focused on stage-environment fit theory to explain motivational development in school-aged students, no direct test of the theoretical assumptions was conducted in the presented articles. This is because student needs and opportunities were not measured in either study. Thus, the dissertation's empirical investigation was limited to the assumed consequences of stage-environment fit (i.e., motivational development) that were theoretically explained in accordance with stage-environment fit theory.

To sum up, this dissertation provides a broad overview of achievement motivation development over the course of students' school careers. The term achievement motivation comprises a large variety of distinct, but sometimes overlapping, constructs and it is not possible to investigate all achievement motivation facets at once. Therefore, subsequent research must investigate additional constructs that were not included in the present investigations. The results were explained by and discussed with regard to stage environment fit theory. Additional research is needed to empirically test assumptions of this theory that were not supported by the presented results on motivational development.

6.5 Subsequent Research

This dissertation provides clear answers to several research questions (e.g., the exact magnitude of change was quantified for several motivational constructs). However, research often raises even more questions than it answers. Some exemplary subsequent investigations are suggested in the following.

As noted in the previous section, the presented studies should be extended with regard to additional motivational constructs. For example, a further meta-analysis of longitudinal studies or new longitudinal studies could focus on the development of intrinsic, extrinsic, and amotivation (i.e., on constructs derived from self-determination theory; Ryan & Deci, 2000a;

2000b) or on the development of intrinsic value, attainment value, utility value, and cost (i.e., on value constructs derived from expectancy value theory; Wigfield & Eccles, 2002). Based on the quantitative decreases of achievement goals found (i.e., decreases in all goals), one could assume similar tendencies with regard to extrinsic motivation, attainment value, utility value, and cost (i.e., an overall decrease in all mentioned constructs). If subsequent research confirms the assumption of a quantitative decrease of additional constructs, a need for interventions that also focus on the extrinsic part of student motivation could be derived.

As the results of the meta-analysis question the assumption of a particularly strong stage-environment mismatch after the school transition, subsequent research could also focus on the effects of school transition on student motivational development. Optimally, a new longitudinal study would capture student achievement motivation at several time points before and after the school transition. Thus, the magnitude of decreases before, during, and after the school transition could be compared. In addition, students' developing needs and school factors should be assessed. Student needs could refer to needs for autonomy, achievement, and relatedness or to student circadian preference. School factors could consist of class climate regarding autonomy, achievement, and relatedness. Thus, it could be tested whether the gap between student needs and opportunities indeed increases with rising age and the school transition. In addition, it could be clarified which unsatisfied need is related to decreases of which motivational construct. For example, according to self-determination theory (Ryan & Deci, 2000a; 2000b), one could assume that unsatisfied needs for autonomy will result in decreasing intrinsic motivation, but will not reduce extrinsic motivation. On the other hand, a stage-environment mismatch regarding needs for achievement and relatedness could be detrimental to intrinsic as well as extrinsic motivation.

Finally, as student motivational development differed by geographic location in the meta-analysis, more cross culture research can be suggested. In this regard, investigations of the particular types of educational practices that are associated with reduced stage-environment mismatch in some countries could be conducted. Subsequently, it could be examined whether similar practices could be applied in other countries.

6.6 Outlook

If motivational variables are supposed to be the new king of educational research (Grigorenko, 2016), it is alarming that the king's well-being worsens with every school year (i.e., a motivation decline over the course of students' school careers is evident). Is this because the individuals in schools (i.e., teachers, students, parents) do not pay as much attention to achievement motivation as educational researchers do? The king is powerful (e.g., achievement motivation is a powerful predictor of achievement) and this power can be used either productively or destructively. The king does not need to be mad. In fact, if the individuals in schools would treat him properly, the king is mild (i.e., achievement motivation is open to different types of interventions). Maybe, the teachers, students, and parents just have to get to know the king better. Maybe, the king just needs an image campaign like a regular politician. Maybe, researchers could help the individuals in schools and the king to get along with each other better. That is, when researchers, teachers, students, and parents collaborate, a happy kingdom (i.e., positive motivation development) can come true for more students.

Appendix A

Tables

Table A1. *Example Articles that Were Excluded Due to the Set Criteria (a) to (g)*

Article	Search strategy	Inclusion and Exclusion criteria
Becker, McElvany, and Kortenbruck (2010)	**(1)** PsycINFO search string	**(a)** Relevant constructs not gathered in the longitudinal section or no means/SDs given
Settles, Zapolski, and Smith (2014)	**(2)** Articles that cited Eccles et al. (1993)	**(a)** Relevant constructs not gathered in the longitudinal section or no means/SDs given
Pullmann and Allik (2000)	**(3)** Reference lists of relevant reviews and meta-analyses	**(a)** Relevant constructs not gathered in the longitudinal section or no means/SDs given
Chan and Moore (2006)	**(1)** PsycINFO search string	**(b)** Intervention study; experimental manipulation; study of extreme groups
Chaplin, Freiburger, Mayes, and Sinha (2010)	**(2)** Articles that cited Eccles et al. (1993)	**(b)** Intervention study; experimental manipulation; study of extreme groups
Goldberg, and Cornell (1998)	**(3)** Reference lists of relevant reviews and meta-analyses	**(b)** Intervention study; experimental manipulation; study of extreme groups
Bartle-Haring, Brucker, and Hock (2002)	**(1)** PsycINFO search string	**(c)** Sample consisted no school students
Roberts and Robins (2004)	**(2)** Articles that cited Eccles et al. (1993)	**(c)** Sample consisted no school students
Elliot (1996)	**(3)** Reference lists of relevant reviews and meta-analyses	**(c)** Sample consisted no school students
Antunes and Fontaine (2007)	**(1)** PsycINFO search string	**(d)** Means contain values of more than one different class level (e.g., t1= grades 5 to 7)
Jiang, Huebner, and Siddall (2012)	**(2)** Articles that cited Eccles et al. (1993)	**(d)** Means contain values of more than one different class level (e.g., t1= grades 5 to 7)
Guay, Boivin, and Hodges (1999)	**(3)** Reference lists of relevant reviews and meta-analyses	**(d)** Means contain values of more than one different class level (e.g., t1= grades 5 to 7)
Wade, Thompson, Tashakkori, and Valente (1989)	**(3)** Reference lists of relevant reviews and meta-analyses	**(e)** Published before 1994
Keltikangas-Järvinen (1990)	**(3)** Reference lists of relevant reviews and meta-analyses	**(e)** Published before 1994
Stein, Newcomb, and Bentler (1986)	**(3)** Reference lists of relevant reviews and meta-analyses	**(e)** Published before 1994
Getzoff (2004)	**(3)** Reference lists of relevant reviews and meta-analyses	**(f)** No peer reviewed journal in English language
Hoglund (1995)	**(3)** Reference lists of relevant reviews and meta-analyses	**(f)** No peer reviewed journal in English language
Yang (1997)	**(3)** Reference lists of relevant reviews and meta-analyses	**(f)** No peer reviewed journal in English language
Wentzel (1997)	**(4)** Reference lists of articles found by search strategies 1-3	**(g)** All scales used in article were excluded through classification process
Van der Werf, Opdenakker, and Kuyper (2008)	**(4)** Reference lists of articles found by search strategies 1-3	**(g)** All scales used in article were excluded through classification process
Stoel, Peetsma, and Roeleveld (2003)	**(4)** Reference lists of articles found by search strategies 1-3	**(g)** All scales used in article were excluded through classification process

Table A2. *Example Scales that Were Assigned to the Motivational Constructs*

Scale name	Source of the scale	Description of the scale or example item	Assigned construct
Self-esteem	Rosenberg (1989)	"I am satisfied with self."	Self-esteem
Global self-worth	Harter (1985)	"Some teenagers are happy with themselves most of the time BUT other teenagers are often not happy with themselves."	Self-esteem
Global self-esteem	DuBois et al. (1996)	"I am happy with myself as a person"	Self-esteem
Self-esteem	Coopersmith (1975)	"I often feel ashamed of myself" (reverse scored)	Self-esteem
Self-esteem	Bachman (1975)	"I feel that I have a number of good qualities."	Self-esteem
Scholastic competence	Harter (1982)	"Some teenagers do very well at their classwork BUT other teenagers don't do very well at their classwork."	Self-concept
Mathematics self-concept	Marsh (1983)	"I have always done well in mathematics."	Self-concept
Math self-concept	Eccles (1983)	"Compared to other kids your age, how well do you do in math?"	Self-concept
Competence beliefs	Spinath and Spinath (2005)	"How good are you at the things you do in school?"	Self-concept
Perceived academic performance	Whitesell, Mitchell, and Spicer (1989)	"Compared with your classmates, how well do you do in school?"	Self-concept
Self-efficacy	Midgley et al. (2000)	"I'm certain I can master the skills taught in school this year"	Self-efficacy
Mathematics self-efficacy	Pajares and Graham (1999)	"Suppose that you were asked to answer the following mathematics questions in a multiple choice test tomorrow. Please indicate how confident you are that you will give the correct answer to each question correctly."	Self-efficacy
Academic self-efficacy	Schwarzer and Jerusalem (1995)	"I am certain I can learn skills taught in school this year."	Self-efficacy
Academic self-efficacy beliefs	Bandura et al. (1996)	Perceived capability to successfully master different curricular areas.	Self-efficacy
Beliefs of academic self-efficacy	Pintrich and De Groot (1990)	"I am sure that I can do an excellent job on the problems and tasks assigned for [schoolwork/ subject class]"	Self-efficacy
Intrinsic value of school	Eccles (1983)	"In general, I find working on school work …" (very boring to very interesting)	Int. mot.
Intrinsic motivation	Spinath and Steinmayr (2008)	"How much do you like mental arithmetic?"	Int. mot.
Liking school	Barber (2004)	"Compared to last year, how much do you like school this year?"	Int. mot.
Academic intrinsic motivation	Gottfried (1985)	Measures enjoyment of school in general	Int. mot.
Affective attitude towards reading	Van Schooten, de Glopper, and Stoel (2004)	"I enjoy reading literary works."	Int. mot.

Scale name	Source of the scale	Description of the scale or example item	Assigned construct
Mastery goals	Roeser et al. (1996)	"The main reason why I do my work in [school/ subject class] is because I like to learn")	Mastery
Task goals approach	Midgley et al. (2000)	"I wanted to find out something new"	Mastery
Mastery	Spinath et al. (2002)	"In school, I want to get new ideas."	Mastery
Mastery	Harter (1981)	"Some kids read things because they are interested in the subject BUT other kids read things because the teacher wants them to"	Mastery
Task-Mastery Goal scale	Meece et al. (1988)	"I wanted to learn as much as possible"	Mastery
Performance-approach goal	Roeser et al. (1996)	"I would feel successful if I did better than most of the other students in my [school/subject class]"	PAP
Ability goals approach	Midgley et al. (2000)	"I wanted others to think I was smart."	PAP
Performance-approach	Spinath et al. (2002)	"In school, want to show that I am good at things."	PAP
Performance-approach goals	Köller and Baumert (1998)	"It is important to me that I know more than other students in my class"	PAP
Performance Goal scale	Meece et al. (1988)	"I wanted others to think I am smart"	PAP
Performance-avoidance goals	Schwinger and Wild (2006)	"It is important to me that others do not think I am stupid"	PAV
Performance-avoidance goal	Midgley et al. (1997)	"An important reason I do my work is so that the teacher doesn't think I know less than Others"	PAV
Avoidance goals in French	Bouffard et al. (1998)	"In French, I do no more work than what is necessary just to not fail"	PAV
Performance-avoidance	Spinath et al. (2002)	"In school, don't want the other students to think I am stupid."	PAV
Performance avoidance math	Seegers et al. (2002)	"I feel embarrassed when I have to ask for help during math lessons"	PAV

Note. Self-concept = Academic self-concept. Self-efficacy = Academic self-efficacy. Int. mot. = Intrinsic motivation. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals.

Table A3. *Averaged Autocorrelations of Longitudinal Studies Separated by Construct and Duration of Interval*

Construct	All interval durations		Interval duration < 1 year		Interval duration = 1 year		Interval duration > 1 year	
	r (*SD*)	*h* effect sizes/ *k* studies	*r* (*SD*)	*h* effect sizes/ *k* studies	*r* (*SD*)	*h* effect sizes/	*r* (*SD*)	*h* effect sizes/ *k* studies
Overall	.509 (.143)	403/65	.631 (.104)	81/11	.536 (.117)	155/46	.426 (.129)	167/42
Self-esteem	.518 (.111)	84/23	.550 (-)	1/1	.568 (.100)	43/18	.464 (.099)	40/18
Self-concept	.531 (.149)	113/23	.649 (.122)	27/6	.549 (.142)	33/14	.460 (.124)	53/15
Self-efficacy	.558 (.088)	21/4	.576 (.057)	19/1	.39 (.042)	2/2	-	-
Int. Mot.	.476 (.173)	102/22	.686 (.115)	18/4	.533 (.122)	33/13	.365 (.128)	51/13
Mastery	.472 (.138)	36/11	.578 (.076)	8/2	.488 (.092)	17/8	.369 (.166)	11/4
PAP	.549 (.104)	35/10	.639 (.042)	8/2	.535 (.107)	18/9	.499 (.095)	9/2
PAV	439 (.088)	12/5	-	-	.472 (.071)	9/5	.340 (.053)	3/1

Note. Self-concept = Academic self-concept. Self-efficacy = Academic self-efficacy. Int. mot. = Intrinsic motivation. PAP = Performance-approach achievement goals. PAV = Performance-avoidance achievement goals.

Figures

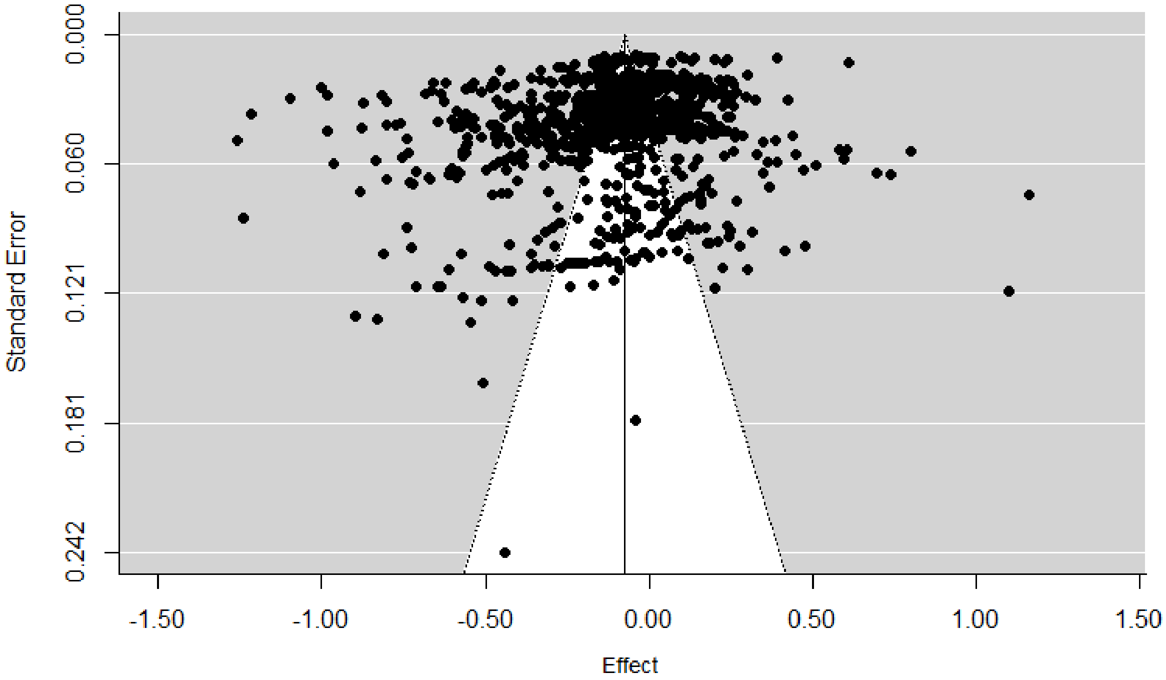
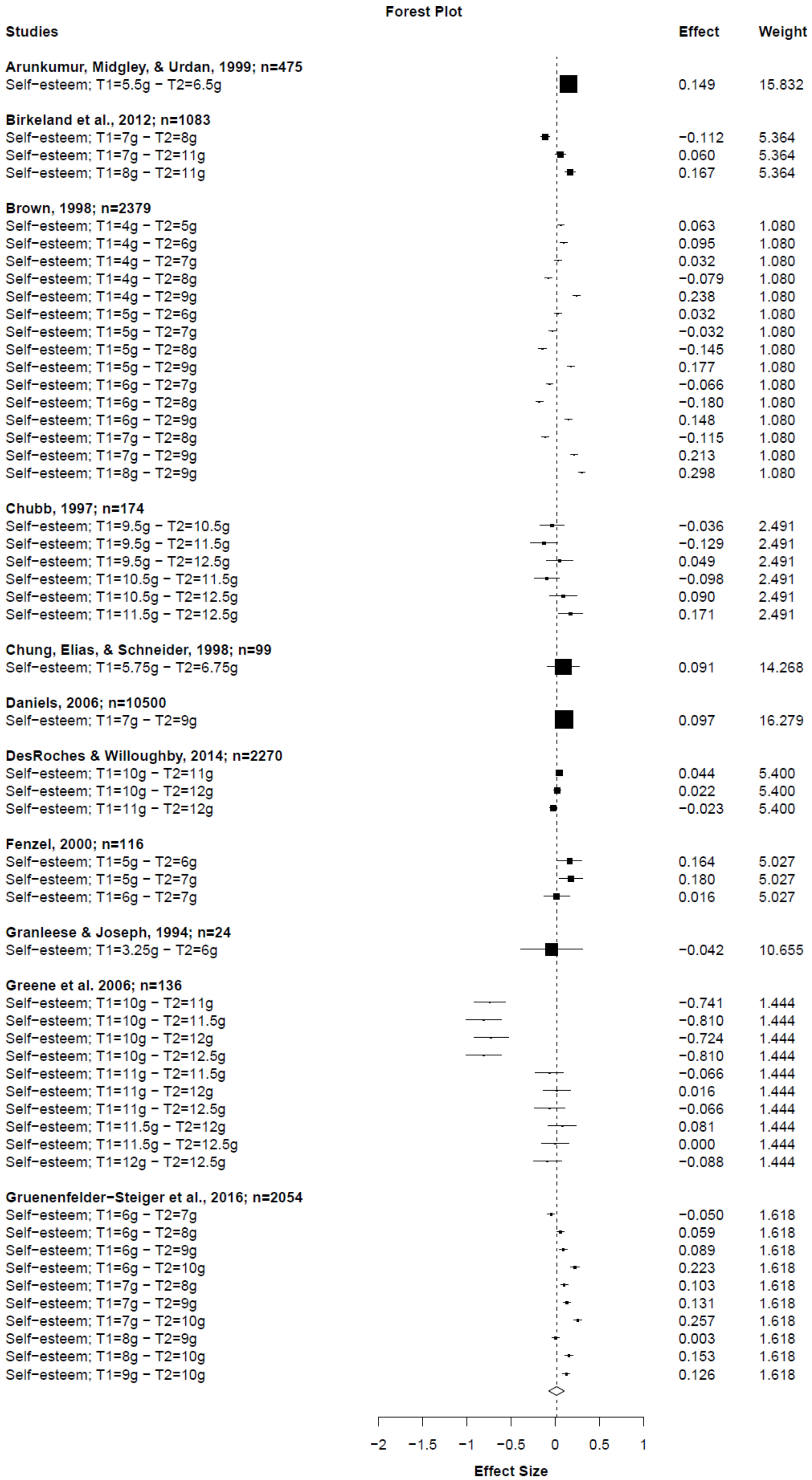
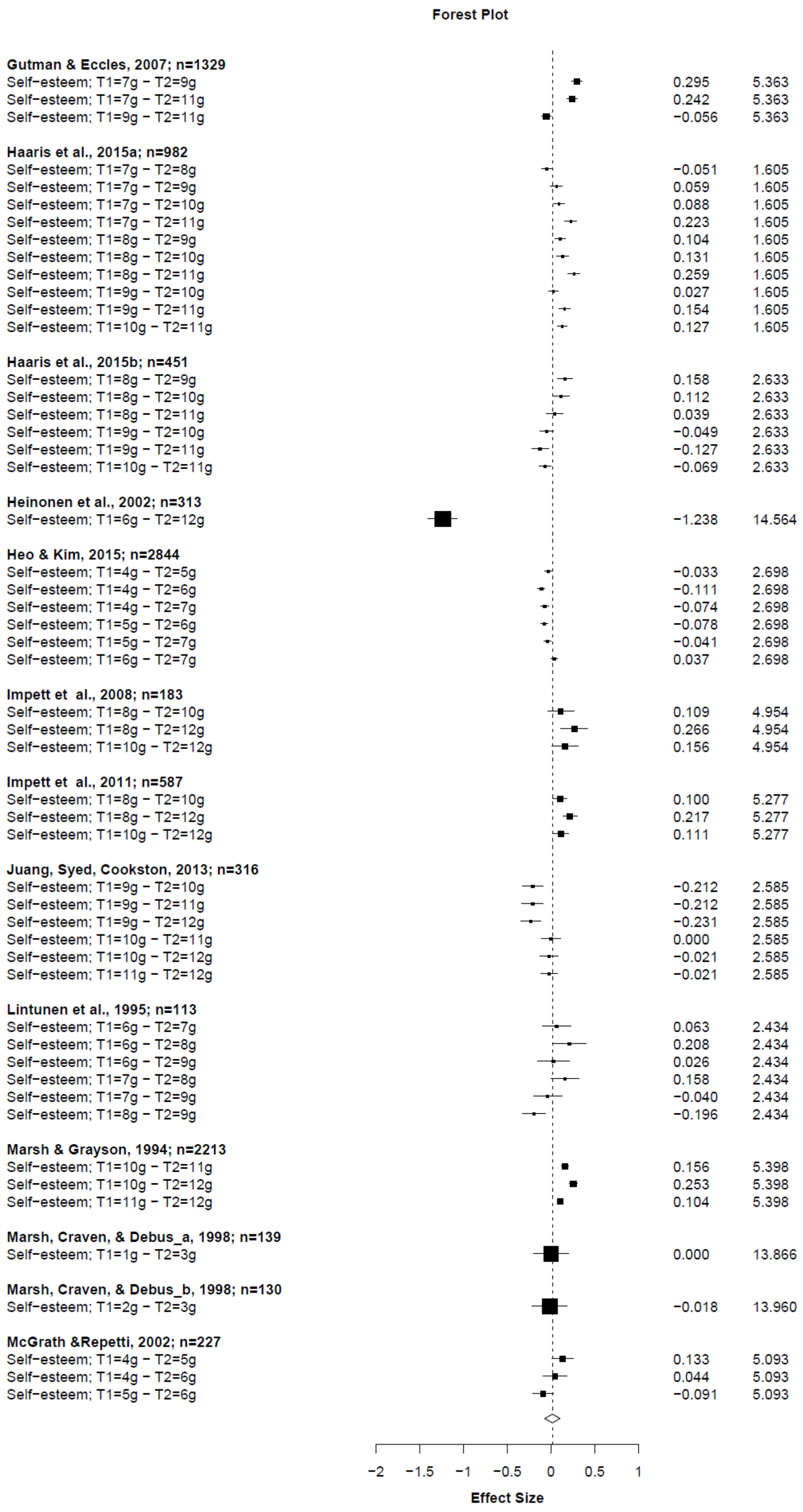
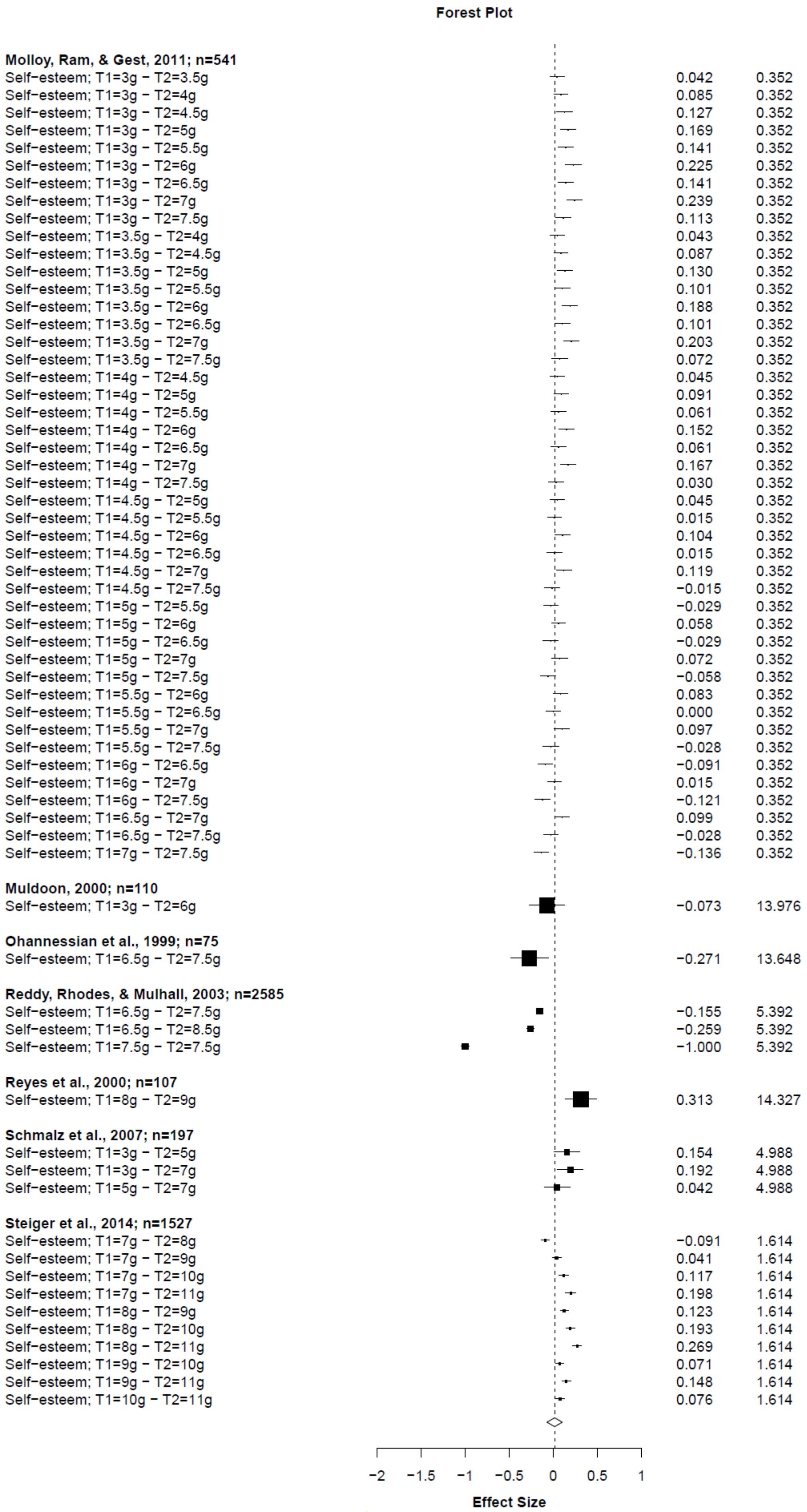
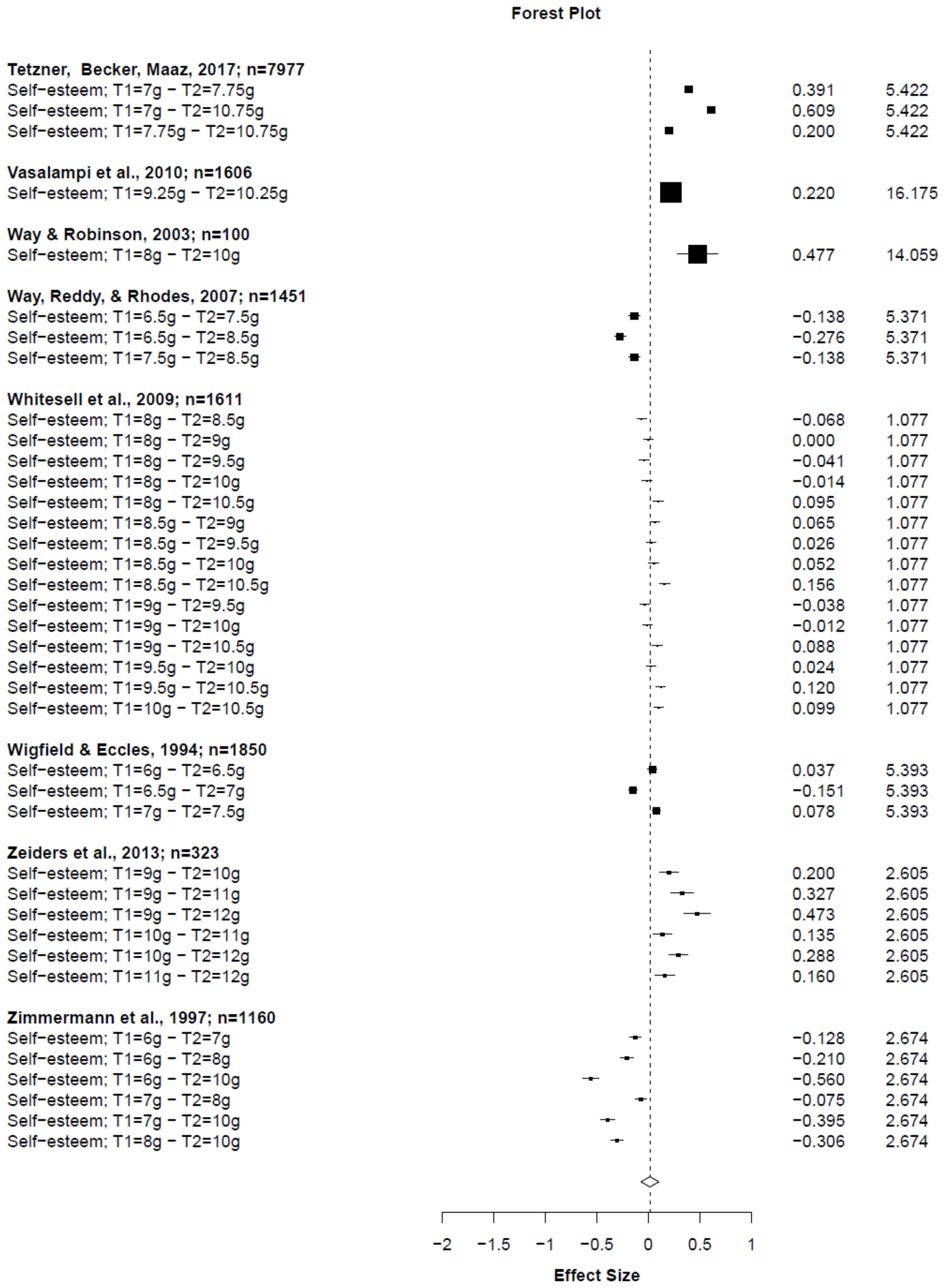


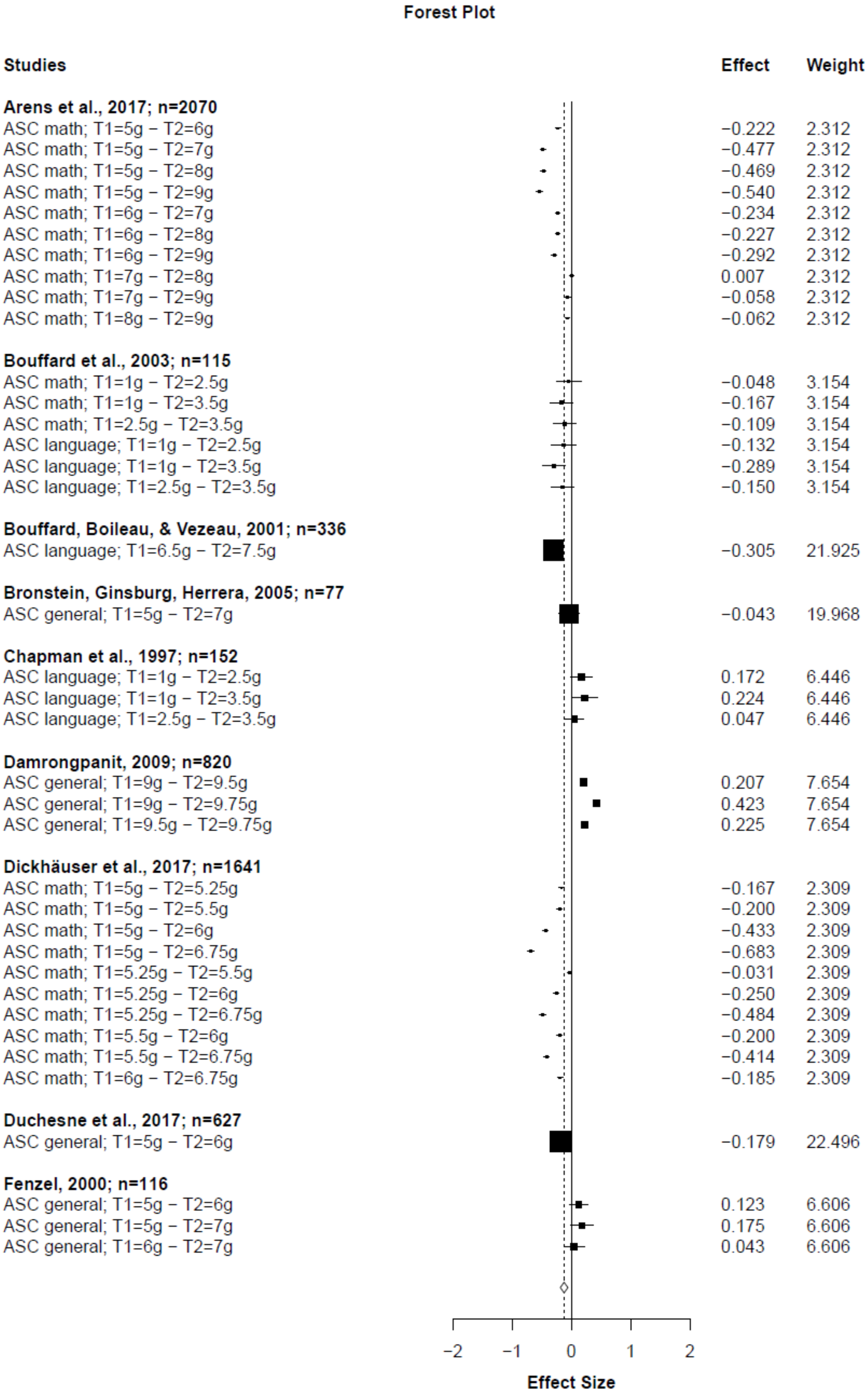
Figure A1. Funnel plot for 914 effect sizes from 107 independent studies. Points represent collected effect sizes of difference.

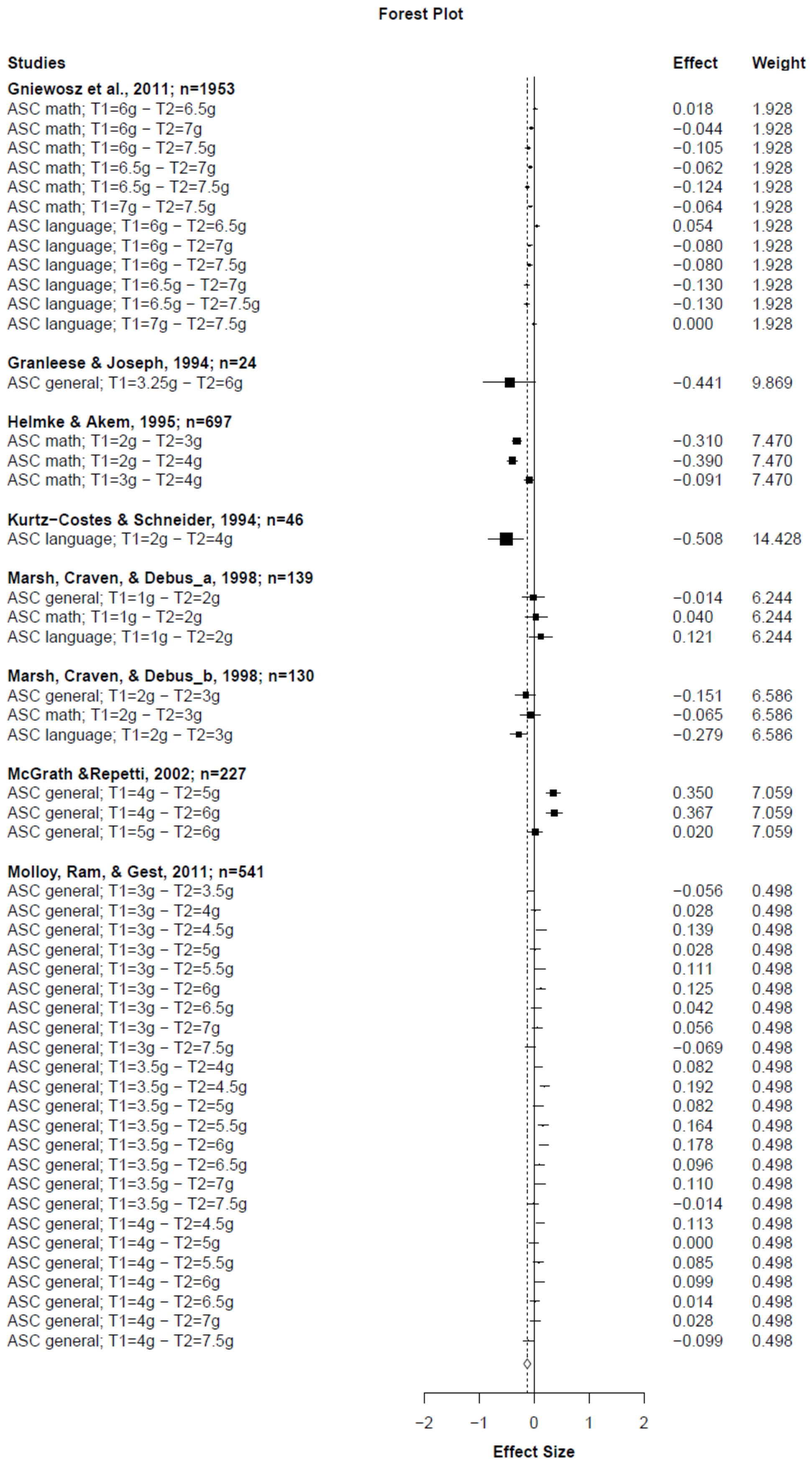

Forest Plot
Studies
Effect
Weight
Arunkumur, Midgley, & Urdan, 1999; n=475
Self−esteem; T1=5.5g − T2=6.5g 0.149 15.832
Birkeland et al., 2012; n=1083
Self−esteem; T1=7g − T2=8g −0.112 5.364
Self−esteem; T1=7g − T2=11g 0.060 5.364
Self−esteem; T1=8g − T2=11g 0.167 5.364
Brown, 1998; n=2379
Self−esteem; T1=4g − T2=5g 0.063 1.080
Self−esteem; T1=4g − T2=6g 0.095 1.080
Self−esteem; T1=4g − T2=7g 0.032 1.080
Self−esteem; T1=4g − T2=8g −0.079 1.080
Self−esteem; T1=4g − T2=9g 0.238 1.080
Self−esteem; T1=5g − T2=6g 0.032 1.080
Self−esteem; T1=5g − T2=7g −0.032 1.080
Self−esteem; T1=5g − T2=8g −0.145 1.080
Self−esteem; T1=5g − T2=9g 0.177 1.080
Self−esteem; T1=6g − T2=7g −0.066 1.080
Self−esteem; T1=6g − T2=8g −0.180 1.080
Self−esteem; T1=6g − T2=9g 0.148 1.080
Self−esteem; T1=7g − T2=8g −0.115 1.080
Self−esteem; T1=7g − T2=9g 0.213 1.080
Self−esteem; T1=8g − T2=9g 0.298 1.080
Chubb, 1997; n=174
Self−esteem; T1=9.5g − T2=10.5g −0.036 2.491
Self−esteem; T1=9.5g − T2=11.5g −0.129 2.491
Self−esteem; T1=9.5g − T2=12.5g 0.049 2.491
Self−esteem; T1=10.5g − T2=11.5g −0.098 2.491
Self−esteem; T1=10.5g − T2=12.5g 0.090 2.491
Self−esteem; T1=11.5g − T2=12.5g 0.171 2.491
Chung, Elias, & Schneider, 1998; n=99
Self−esteem; T1=5.75g − T2=6.75g 0.091 14.268
Daniels, 2006; n=10500
Self−esteem; T1=7g − T2=9g 0.097 16.279
DesRoches & Willoughby, 2014; n=2270
Self−esteem; T1=10g − T2=11g 0.044 5.400
Self−esteem; T1=10g − T2=12g 0.022 5.400
Self−esteem; T1=11g − T2=12g −0.023 5.400
Fenzel, 2000; n=116
Self−esteem; T1=5g − T2=6g 0.164 5.027
Self−esteem; T1=5g − T2=7g 0.180 5.027
Self−esteem; T1=6g − T2=7g 0.016 5.027
Granleese & Joseph, 1994; n=24
Self−esteem; T1=3.25g − T2=6g −0.042 10.655
Greene et al. 2006; n=136
Self−esteem; T1=10g − T2=11g −0.741 1.444
Self−esteem; T1=10g − T2=11.5g −0.810 1.444
Self−esteem; T1=10g − T2=12g −0.724 1.444
Self−esteem; T1=10g − T2=12.5g −0.810 1.444
Self−esteem; T1=11g − T2=11.5g −0.066 1.444
Self−esteem; T1=11g − T2=12g 0.016 1.444
Self−esteem; T1=11g − T2=12.5g −0.066 1.444
Self−esteem; T1=11.5g − T2=12g 0.081 1.444
Self−esteem; T1=11.5g − T2=12.5g 0.000 1.444
Self−esteem; T1=12g − T2=12.5g −0.088 1.444
Gruenenfelder−Steiger et al., 2016; n=2054
Self−esteem; T1=6g − T2=7g −0.050 1.618
Self−esteem; T1=6g − T2=8g 0.059 1.618
Self−esteem; T1=6g − T2=9g 0.089 1.618
Self−esteem; T1=6g − T2=10g 0.223 1.618
Self−esteem; T1=7g − T2=8g 0.103 1.618
Self−esteem; T1=7g − T2=9g 0.131 1.618
Self−esteem; T1=7g − T2=10g 0.257 1.618
Self−esteem; T1=8g − T2=9g 0.003 1.618
Self−esteem; T1=8g − T2=10g 0.153 1.618
Self−esteem; T1=9g − T2=10g 0.126 1.618
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size

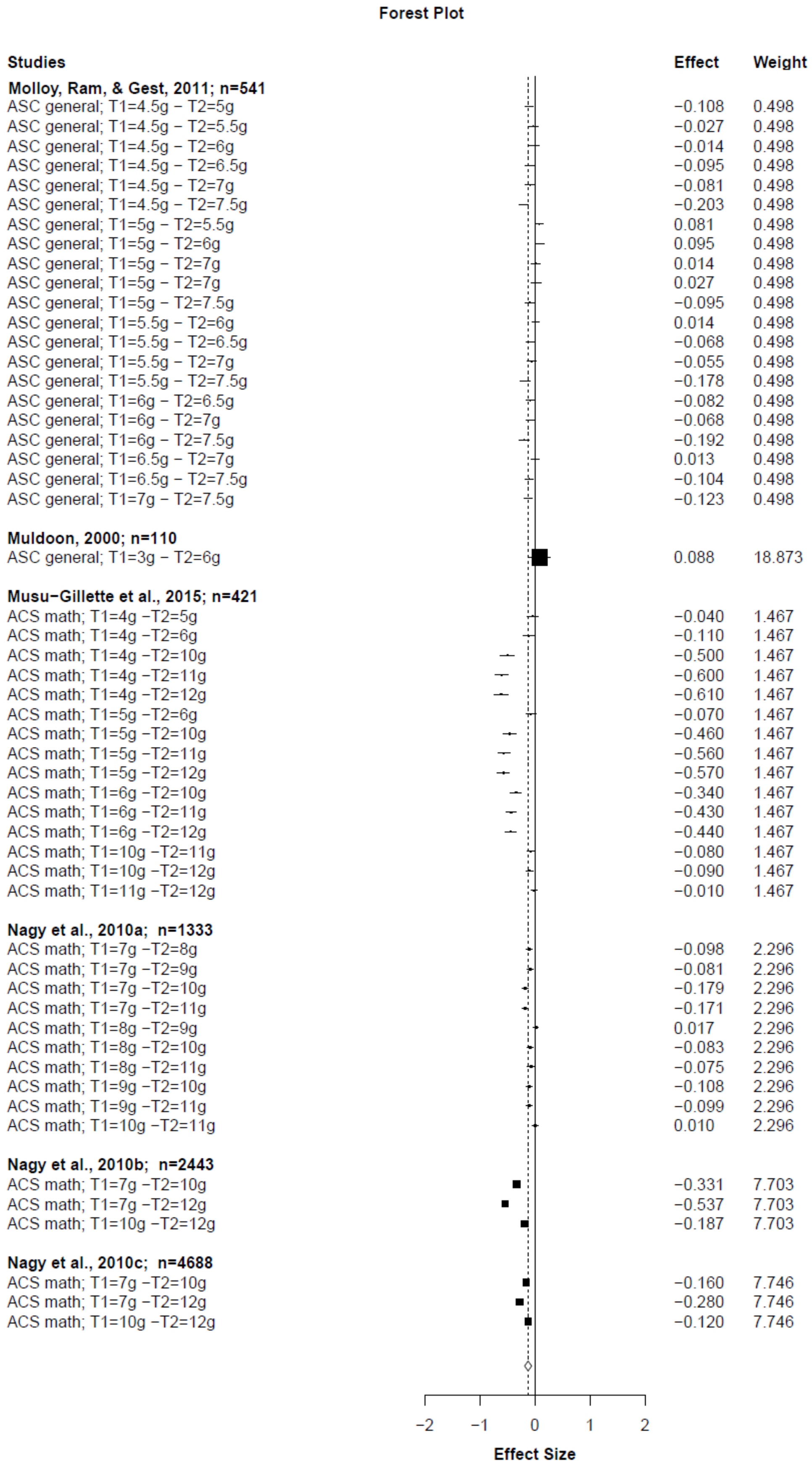

Forest Plot
Gutman & Eccles, 2007; n=1329
Self−esteem; T1=7g − T2=9g 0.295 5.363
Self−esteem; T1=7g − T2=11g 0.242 5.363
Self−esteem; T1=9g − T2=11g −0.056 5.363
Haaris et al., 2015a; n=982
Self−esteem; T1=7g − T2=8g −0.051 1.605
Self−esteem; T1=7g − T2=9g 0.059 1.605
Self−esteem; T1=7g − T2=10g 0.088 1.605
Self−esteem; T1=7g − T2=11g 0.223 1.605
Self−esteem; T1=8g − T2=9g 0.104 1.605
Self−esteem; T1=8g − T2=10g 0.131 1.605
Self−esteem; T1=8g − T2=11g 0.259 1.605
Self−esteem; T1=9g − T2=10g 0.027 1.605
Self−esteem; T1=9g − T2=11g 0.154 1.605
Self−esteem; T1=10g − T2=11g 0.127 1.605
Haaris et al., 2015b; n=451
Self−esteem; T1=8g − T2=9g 0.158 2.633
Self−esteem; T1=8g − T2=10g 0.112 2.633
Self−esteem; T1=8g − T2=11g 0.039 2.633
Self−esteem; T1=9g − T2=10g −0.049 2.633
Self−esteem; T1=9g − T2=11g −0.127 2.633
Self−esteem; T1=10g − T2=11g −0.069 2.633
Heinonen et al., 2002; n=313
Self−esteem; T1=6g − T2=12g −1.238 14.564
Heo & Kim, 2015; n=2844
Self−esteem; T1=4g − T2=5g −0.033 2.698
Self−esteem; T1=4g − T2=6g −0.111 2.698
Self−esteem; T1=4g − T2=7g −0.074 2.698
Self−esteem; T1=5g − T2=6g −0.078 2.698
Self−esteem; T1=5g − T2=7g −0.041 2.698
Self−esteem; T1=6g − T2=7g 0.037 2.698
Impett et al., 2008; n=183
Self−esteem; T1=8g − T2=10g 0.109 4.954
Self−esteem; T1=8g − T2=12g 0.266 4.954
Self−esteem; T1=10g − T2=12g 0.156 4.954
Impett et al., 2011; n=587
Self−esteem; T1=8g − T2=10g 0.100 5.277
Self−esteem; T1=8g − T2=12g 0.217 5.277
Self−esteem; T1=10g − T2=12g 0.111 5.277
Juang, Syed, Cookston, 2013; n=316
Self−esteem; T1=9g − T2=10g −0.212 2.585
Self−esteem; T1=9g − T2=11g −0.212 2.585
Self−esteem; T1=9g − T2=12g −0.231 2.585
Self−esteem; T1=10g − T2=11g 0.000 2.585
Self−esteem; T1=10g − T2=12g −0.021 2.585
Self−esteem; T1=11g − T2=12g −0.021 2.585
Lintunen et al., 1995; n=113
Self−esteem; T1=6g − T2=7g 0.063 2.434
Self−esteem; T1=6g − T2=8g 0.208 2.434
Self−esteem; T1=6g − T2=9g 0.026 2.434
Self−esteem; T1=7g − T2=8g 0.158 2.434
Self−esteem; T1=7g − T2=9g −0.040 2.434
Self−esteem; T1=8g − T2=9g −0.196 2.434
Marsh & Grayson, 1994; n=2213
Self−esteem; T1=10g − T2=11g 0.156 5.398
Self−esteem; T1=10g − T2=12g 0.253 5.398
Self−esteem; T1=11g − T2=12g 0.104 5.398
Marsh, Craven, & Debus_a, 1998; n=139
Self−esteem; T1=1g − T2=3g 0.000 13.866
Marsh, Craven, & Debus_b, 1998; n=130
Self−esteem; T1=2g − T2=3g −0.018 13.960
McGrath &Repetti, 2002; n=227
Self−esteem; T1=4g − T2=5g 0.133 5.093
Self−esteem; T1=4g − T2=6g 0.044 5.093
Self−esteem; T1=5g − T2=6g −0.091 5.093
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size

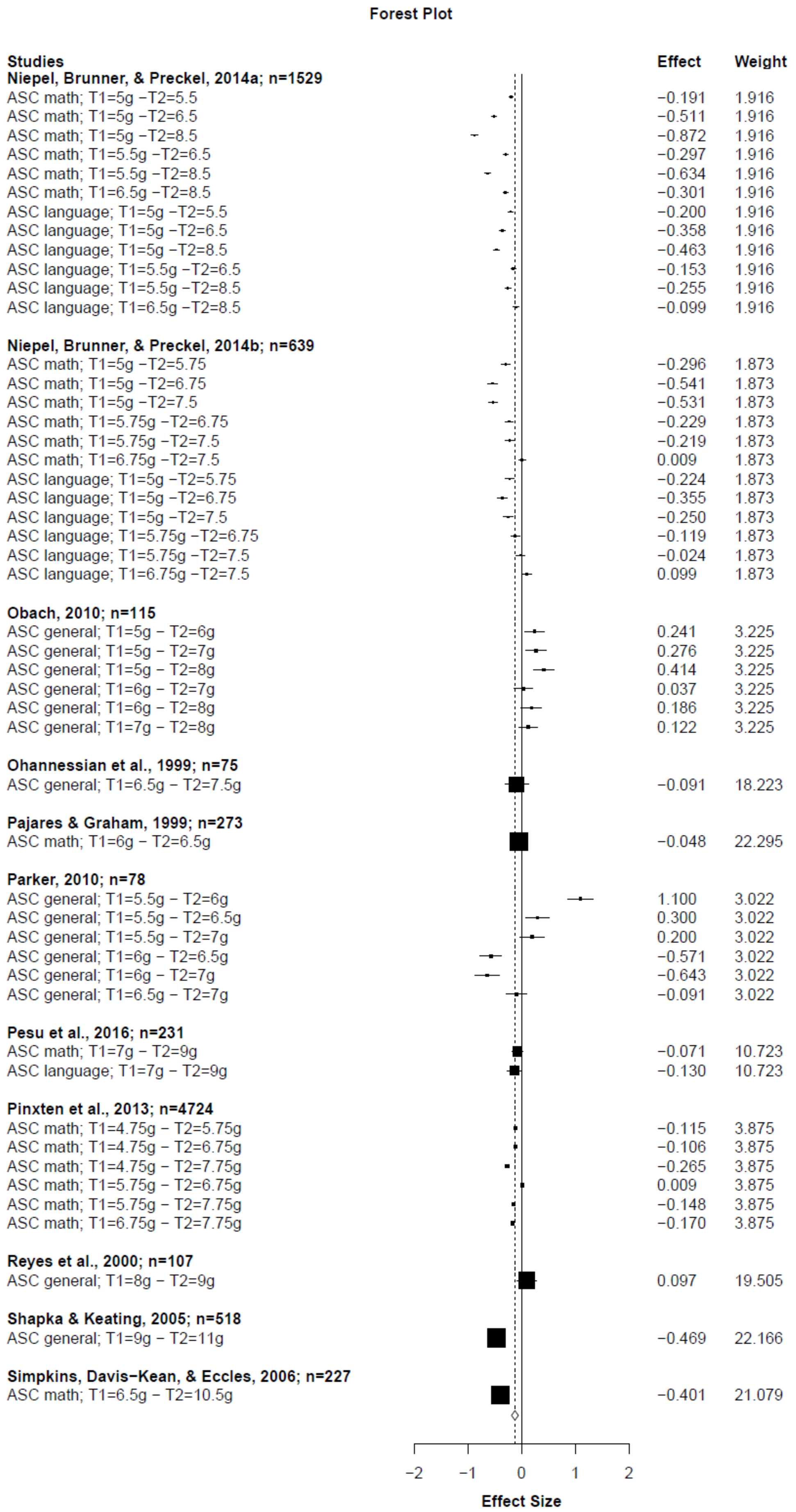

Forest Plot
Molloy, Ram, & Gest, 2011; n=541
Self−esteem; T1=3g − T2=3.5g 0.042 0.352
Self−esteem; T1=3g − T2=4g 0.085 0.352
Self−esteem; T1=3g − T2=4.5g 0.127 0.352
Self−esteem; T1=3g − T2=5g 0.169 0.352
Self−esteem; T1=3g − T2=5.5g 0.141 0.352
Self−esteem; T1=3g − T2=6g 0.225 0.352
Self−esteem; T1=3g − T2=6.5g 0.141 0.352
Self−esteem; T1=3g − T2=7g 0.239 0.352
Self−esteem; T1=3g − T2=7.5g 0.113 0.352
Self−esteem; T1=3.5g − T2=4g 0.043 0.352
Self−esteem; T1=3.5g − T2=4.5g 0.087 0.352
Self−esteem; T1=3.5g − T2=5g 0.130 0.352
Self−esteem; T1=3.5g − T2=5.5g 0.101 0.352
Self−esteem; T1=3.5g − T2=6g 0.188 0.352
Self−esteem; T1=3.5g − T2=6.5g 0.101 0.352
Self−esteem; T1=3.5g − T2=7g 0.203 0.352
Self−esteem; T1=3.5g − T2=7.5g 0.072 0.352
Self−esteem; T1=4g − T2=4.5g 0.045 0.352
Self−esteem; T1=4g − T2=5g 0.091 0.352
Self−esteem; T1=4g − T2=5.5g 0.061 0.352
Self−esteem; T1=4g − T2=6g 0.152 0.352
Self−esteem; T1=4g − T2=6.5g 0.061 0.352
Self−esteem; T1=4g − T2=7g 0.167 0.352
Self−esteem; T1=4g − T2=7.5g 0.030 0.352
Self−esteem; T1=4.5g − T2=5g 0.045 0.352
Self−esteem; T1=4.5g − T2=5.5g 0.015 0.352
Self−esteem; T1=4.5g − T2=6g 0.104 0.352
Self−esteem; T1=4.5g − T2=6.5g 0.015 0.352
Self−esteem; T1=4.5g − T2=7g 0.119 0.352
Self−esteem; T1=4.5g − T2=7.5g −0.015 0.352
Self−esteem; T1=5g − T2=5.5g −0.029 0.352
Self−esteem; T1=5g − T2=6g 0.058 0.352
Self−esteem; T1=5g − T2=6.5g −0.029 0.352
Self−esteem; T1=5g − T2=7g 0.072 0.352
Self−esteem; T1=5g − T2=7.5g −0.058 0.352
Self−esteem; T1=5.5g − T2=6g 0.083 0.352
Self−esteem; T1=5.5g − T2=6.5g 0.000 0.352
Self−esteem; T1=5.5g − T2=7g 0.097 0.352
Self−esteem; T1=5.5g − T2=7.5g −0.028 0.352
Self−esteem; T1=6g − T2=6.5g −0.091 0.352
Self−esteem; T1=6g − T2=7g 0.015 0.352
Self−esteem; T1=6g − T2=7.5g −0.121 0.352
Self−esteem; T1=6.5g − T2=7g 0.099 0.352
Self−esteem; T1=6.5g − T2=7.5g −0.028 0.352
Self−esteem; T1=7g − T2=7.5g −0.136 0.352
Muldoon, 2000; n=110
Self−esteem; T1=3g − T2=6g −0.073 13.976
Ohannessian et al., 1999; n=75
Self−esteem; T1=6.5g − T2=7.5g −0.271 13.648
Reddy, Rhodes, & Mulhall, 2003; n=2585
Self−esteem; T1=6.5g − T2=7.5g −0.155 5.392
Self−esteem; T1=6.5g − T2=8.5g −0.259 5.392
Self−esteem; T1=7.5g − T2=7.5g −1.000 5.392
Reyes et al., 2000; n=107
Self−esteem; T1=8g − T2=9g 0.313 14.327
Schmalz et al., 2007; n=197
Self−esteem; T1=3g − T2=5g 0.154 4.988
Self−esteem; T1=3g − T2=7g 0.192 4.988
Self−esteem; T1=5g − T2=7g 0.042 4.988
Steiger et al., 2014; n=1527
Self−esteem; T1=7g − T2=8g −0.091 1.614
Self−esteem; T1=7g − T2=9g 0.041 1.614
Self−esteem; T1=7g − T2=10g 0.117 1.614
Self−esteem; T1=7g − T2=11g 0.198 1.614
Self−esteem; T1=8g − T2=9g 0.123 1.614
Self−esteem; T1=8g − T2=10g 0.193 1.614
Self−esteem; T1=8g − T2=11g 0.269 1.614
Self−esteem; T1=9g − T2=10g 0.071 1.614
Self−esteem; T1=9g − T2=11g 0.148 1.614
Self−esteem; T1=10g − T2=11g 0.076 1.614
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size

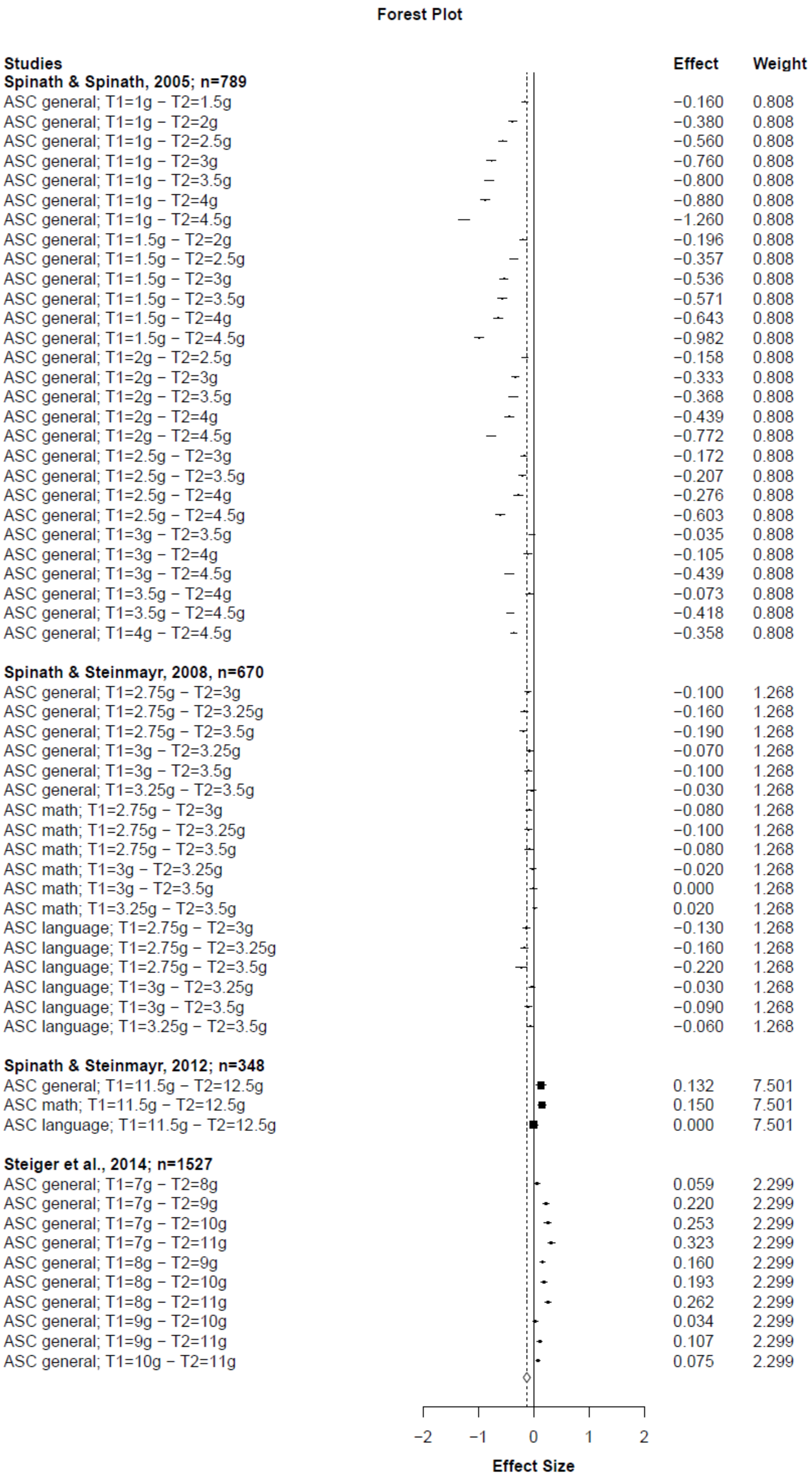

Forest Plot
Tetzner, Becker, Maaz, 2017; n=7977
Self−esteem; T1=7g − T2=7.75g 0.391 5.422
Self−esteem; T1=7g − T2=10.75g 0.609 5.422
Self−esteem; T1=7.75g − T2=10.75g 0.200 5.422
Vasalampi et al., 2010; n=1606
Self−esteem; T1=9.25g − T2=10.25g 0.220 16.175
Way & Robinson, 2003; n=100
Self−esteem; T1=8g − T2=10g 0.477 14.059
Way, Reddy, & Rhodes, 2007; n=1451
Self−esteem; T1=6.5g − T2=7.5g −0.138 5.371
Self−esteem; T1=6.5g − T2=8.5g −0.276 5.371
Self−esteem; T1=7.5g − T2=8.5g −0.138 5.371
Whitesell et al., 2009; n=1611
Self−esteem; T1=8g − T2=8.5g −0.068 1.077
Self−esteem; T1=8g − T2=9g 0.000 1.077
Self−esteem; T1=8g − T2=9.5g −0.041 1.077
Self−esteem; T1=8g − T2=10g −0.014 1.077
Self−esteem; T1=8g − T2=10.5g 0.095 1.077
Self−esteem; T1=8.5g − T2=9g 0.065 1.077
Self−esteem; T1=8.5g − T2=9.5g 0.026 1.077
Self−esteem; T1=8.5g − T2=10g 0.052 1.077
Self−esteem; T1=8.5g − T2=10.5g 0.156 1.077
Self−esteem; T1=9g − T2=9.5g −0.038 1.077
Self−esteem; T1=9g − T2=10g −0.012 1.077
Self−esteem; T1=9g − T2=10.5g 0.088 1.077
Self−esteem; T1=9.5g − T2=10g 0.024 1.077
Self−esteem; T1=9.5g − T2=10.5g 0.120 1.077
Self−esteem; T1=10g − T2=10.5g 0.099 1.077
Wigfield & Eccles, 1994; n=1850
Self−esteem; T1=6g − T2=6.5g 0.037 5.393
Self−esteem; T1=6.5g − T2=7g −0.151 5.393
Self−esteem; T1=7g − T2=7.5g 0.078 5.393
Zeiders et al., 2013; n=323
Self−esteem; T1=9g − T2=10g 0.200 2.605
Self−esteem; T1=9g − T2=11g 0.327 2.605
Self−esteem; T1=9g − T2=12g 0.473 2.605
Self−esteem; T1=10g − T2=11g 0.135 2.605
Self−esteem; T1=10g − T2=12g 0.288 2.605
Self−esteem; T1=11g − T2=12g 0.160 2.605
Zimmermann et al., 1997; n=1160
Self−esteem; T1=6g − T2=7g −0.128 2.674
Self−esteem; T1=6g − T2=8g −0.210 2.674
Self−esteem; T1=6g − T2=10g −0.560 2.674
Self−esteem; T1=7g − T2=8g −0.075 2.674
Self−esteem; T1=7g − T2=10g −0.395 2.674
Self−esteem; T1=8g − T2=10g −0.306 2.674
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size

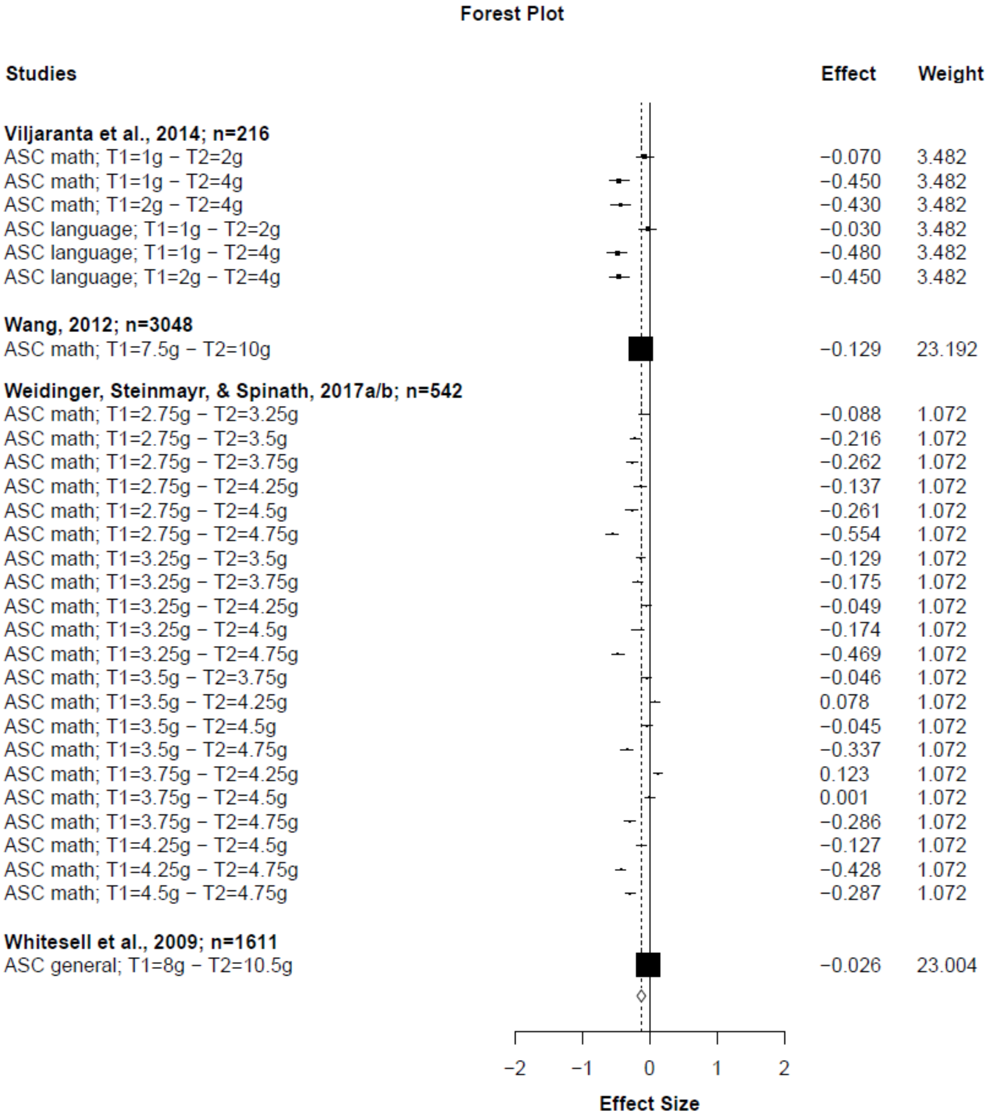

Forest Plot
Studies
Effect
Weight
Arens et al., 2017; n=2070
ASC math; T1=5g – T2=6g −0.222 2.312
ASC math; T1=5g – T2=7g −0.477 2.312
ASC math; T1=5g – T2=8g −0.469 2.312
ASC math; T1=5g – T2=9g −0.540 2.312
ASC math; T1=6g – T2=7g −0.234 2.312
ASC math; T1=6g – T2=8g −0.227 2.312
ASC math; T1=6g – T2=9g −0.292 2.312
ASC math; T1=7g – T2=8g 0.007 2.312
ASC math; T1=7g – T2=9g −0.058 2.312
ASC math; T1=8g – T2=9g −0.062 2.312
Bouffard et al., 2003; n=115
ASC math; T1=1g – T2=2.5g −0.048 3.154
ASC math; T1=1g – T2=3.5g −0.167 3.154
ASC math; T1=2.5g – T2=3.5g −0.109 3.154
ASC language; T1=1g – T2=2.5g −0.132 3.154
ASC language; T1=1g – T2=3.5g −0.289 3.154
ASC language; T1=2.5g – T2=3.5g −0.150 3.154
Bouffard, Boileau, & Vezeau, 2001; n=336
ASC language; T1=6.5g – T2=7.5g −0.305 21.925
Bronstein, Ginsburg, Herrera, 2005; n=77
ASC general; T1=5g – T2=7g −0.043 19.968
Chapman et al., 1997; n=152
ASC language; T1=1g – T2=2.5g 0.172 6.446
ASC language; T1=1g – T2=3.5g 0.224 6.446
ASC language; T1=2.5g – T2=3.5g 0.047 6.446
Damrongpanit, 2009; n=820
ASC general; T1=9g – T2=9.5g 0.207 7.654
ASC general; T1=9g – T2=9.75g 0.423 7.654
ASC general; T1=9.5g – T2=9.75g 0.225 7.654
Dickhäuser et al., 2017; n=1641
ASC math; T1=5g – T2=5.25g −0.167 2.309
ASC math; T1=5g – T2=5.5g −0.200 2.309
ASC math; T1=5g – T2=6g −0.433 2.309
ASC math; T1=5g – T2=6.75g −0.683 2.309
ASC math; T1=5.25g – T2=5.5g −0.031 2.309
ASC math; T1=5.25g – T2=6g −0.250 2.309
ASC math; T1=5.25g – T2=6.75g −0.484 2.309
ASC math; T1=5.5g – T2=6g −0.200 2.309
ASC math; T1=5.5g – T2=6.75g −0.414 2.309
ASC math; T1=6g – T2=6.75g −0.185 2.309
Duchesne et al., 2017; n=627
ASC general; T1=5g – T2=6g −0.179 22.496
Fenzel, 2000; n=116
ASC general; T1=5g – T2=6g 0.123 6.606
ASC general; T1=5g – T2=7g 0.175 6.606
ASC general; T1=6g – T2=7g 0.043 6.606
−2
−1
0
1
2
Effect Size


Forest Plot
Studies Effect Weight
Gniewosz et al., 2011; n=1953
ASC math; T1=6g − T2=6.5g 0.018 1.928
ASC math; T1=6g − T2=7g −0.044 1.928
ASC math; T1=6g − T2=7.5g −0.105 1.928
ASC math; T1=6.5g − T2=7g −0.062 1.928
ASC math; T1=6.5g − T2=7.5g −0.124 1.928
ASC math; T1=7g − T2=7.5g −0.064 1.928
ASC language; T1=6g − T2=6.5g 0.054 1.928
ASC language; T1=6g − T2=7g −0.080 1.928
ASC language; T1=6g − T2=7.5g −0.080 1.928
ASC language; T1=6.5g − T2=7g −0.130 1.928
ASC language; T1=6.5g − T2=7.5g −0.130 1.928
ASC language; T1=7g − T2=7.5g 0.000 1.928
Granleese & Joseph, 1994; n=24
ASC general; T1=3.25g − T2=6g −0.441 9.869
Helmke & Akem, 1995; n=697
ASC math; T1=2g − T2=3g −0.310 7.470
ASC math; T1=2g − T2=4g −0.390 7.470
ASC math; T1=3g − T2=4g −0.091 7.470
Kurtz−Costes & Schneider, 1994; n=46
ASC language; T1=2g − T2=4g −0.508 14.428
Marsh, Craven, & Debus_a, 1998; n=139
ASC general; T1=1g − T2=2g −0.014 6.244
ASC math; T1=1g − T2=2g 0.040 6.244
ASC language; T1=1g − T2=2g 0.121 6.244
Marsh, Craven, & Debus_b, 1998; n=130
ASC general; T1=2g − T2=3g −0.151 6.586
ASC math; T1=2g − T2=3g −0.065 6.586
ASC language; T1=2g − T2=3g −0.279 6.586
McGrath &Repetti, 2002; n=227
ASC general; T1=4g − T2=5g 0.350 7.059
ASC general; T1=4g − T2=6g 0.367 7.059
ASC general; T1=5g − T2=6g 0.020 7.059
Molloy, Ram, & Gest, 2011; n=541
ASC general; T1=3g − T2=3.5g −0.056 0.498
ASC general; T1=3g − T2=4g 0.028 0.498
ASC general; T1=3g − T2=4.5g 0.139 0.498
ASC general; T1=3g − T2=5g 0.028 0.498
ASC general; T1=3g − T2=5.5g 0.111 0.498
ASC general; T1=3g − T2=6g 0.125 0.498
ASC general; T1=3g − T2=6.5g 0.042 0.498
ASC general; T1=3g − T2=7g 0.056 0.498
ASC general; T1=3g − T2=7.5g −0.069 0.498
ASC general; T1=3.5g − T2=4g 0.082 0.498
ASC general; T1=3.5g − T2=4.5g 0.192 0.498
ASC general; T1=3.5g − T2=5g 0.082 0.498
ASC general; T1=3.5g − T2=5.5g 0.164 0.498
ASC general; T1=3.5g − T2=6g 0.178 0.498
ASC general; T1=3.5g − T2=6.5g 0.096 0.498
ASC general; T1=3.5g − T2=7g 0.110 0.498
ASC general; T1=3.5g − T2=7.5g −0.014 0.498
ASC general; T1=4g − T2=4.5g 0.113 0.498
ASC general; T1=4g − T2=5g 0.000 0.498
ASC general; T1=4g − T2=5.5g 0.085 0.498
ASC general; T1=4g − T2=6g 0.099 0.498
ASC general; T1=4g − T2=6.5g 0.014 0.498
ASC general; T1=4g − T2=7g 0.028 0.498
ASC general; T1=4g − T2=7.5g −0.099 0.498
−2 −1 0 1 2
Effect Size


Forest Plot
Studies
Effect
Weight
Molloy, Ram, & Gest, 2011; n=541
ASC general; T1=4.5g − T2=5g −0.108 0.498
ASC general; T1=4.5g − T2=5.5g −0.027 0.498
ASC general; T1=4.5g − T2=6g −0.014 0.498
ASC general; T1=4.5g − T2=6.5g −0.095 0.498
ASC general; T1=4.5g − T2=7g −0.081 0.498
ASC general; T1=4.5g − T2=7.5g −0.203 0.498
ASC general; T1=5g − T2=5.5g 0.081 0.498
ASC general; T1=5g − T2=6g 0.095 0.498
ASC general; T1=5g − T2=7g 0.014 0.498
ASC general; T1=5g − T2=7g 0.027 0.498
ASC general; T1=5g − T2=7.5g −0.095 0.498
ASC general; T1=5.5g − T2=6g 0.014 0.498
ASC general; T1=5.5g − T2=6.5g −0.068 0.498
ASC general; T1=5.5g − T2=7g −0.055 0.498
ASC general; T1=5.5g − T2=7.5g −0.178 0.498
ASC general; T1=6g − T2=6.5g −0.082 0.498
ASC general; T1=6g − T2=7g −0.068 0.498
ASC general; T1=6g − T2=7.5g −0.192 0.498
ASC general; T1=6.5g − T2=7g 0.013 0.498
ASC general; T1=6.5g − T2=7.5g −0.104 0.498
ASC general; T1=7g − T2=7.5g −0.123 0.498
Muldoon, 2000; n=110
ASC general; T1=3g − T2=6g 0.088 18.873
Musu−Gillette et al., 2015; n=421
ACS math; T1=4g −T2=5g −0.040 1.467
ACS math; T1=4g −T2=6g −0.110 1.467
ACS math; T1=4g −T2=10g −0.500 1.467
ACS math; T1=4g −T2=11g −0.600 1.467
ACS math; T1=4g −T2=12g −0.610 1.467
ACS math; T1=5g −T2=6g −0.070 1.467
ACS math; T1=5g −T2=10g −0.460 1.467
ACS math; T1=5g −T2=11g −0.560 1.467
ACS math; T1=5g −T2=12g −0.570 1.467
ACS math; T1=6g −T2=10g −0.340 1.467
ACS math; T1=6g −T2=11g −0.430 1.467
ACS math; T1=6g −T2=12g −0.440 1.467
ACS math; T1=10g −T2=11g −0.080 1.467
ACS math; T1=10g −T2=12g −0.090 1.467
ACS math; T1=11g −T2=12g −0.010 1.467
Nagy et al., 2010a; n=1333
ACS math; T1=7g −T2=8g −0.098 2.296
ACS math; T1=7g −T2=9g −0.081 2.296
ACS math; T1=7g −T2=10g −0.179 2.296
ACS math; T1=7g −T2=11g −0.171 2.296
ACS math; T1=8g −T2=9g 0.017 2.296
ACS math; T1=8g −T2=10g −0.083 2.296
ACS math; T1=8g −T2=11g −0.075 2.296
ACS math; T1=9g −T2=10g −0.108 2.296
ACS math; T1=9g −T2=11g −0.099 2.296
ACS math; T1=10g −T2=11g 0.010 2.296
Nagy et al., 2010b; n=2443
ACS math; T1=7g −T2=10g −0.331 7.703
ACS math; T1=7g −T2=12g −0.537 7.703
ACS math; T1=10g −T2=12g −0.187 7.703
Nagy et al., 2010c; n=4688
ACS math; T1=7g −T2=10g −0.160 7.746
ACS math; T1=7g −T2=12g −0.280 7.746
ACS math; T1=10g −T2=12g −0.120 7.746
−2 −1 0 1 2
Effect Size

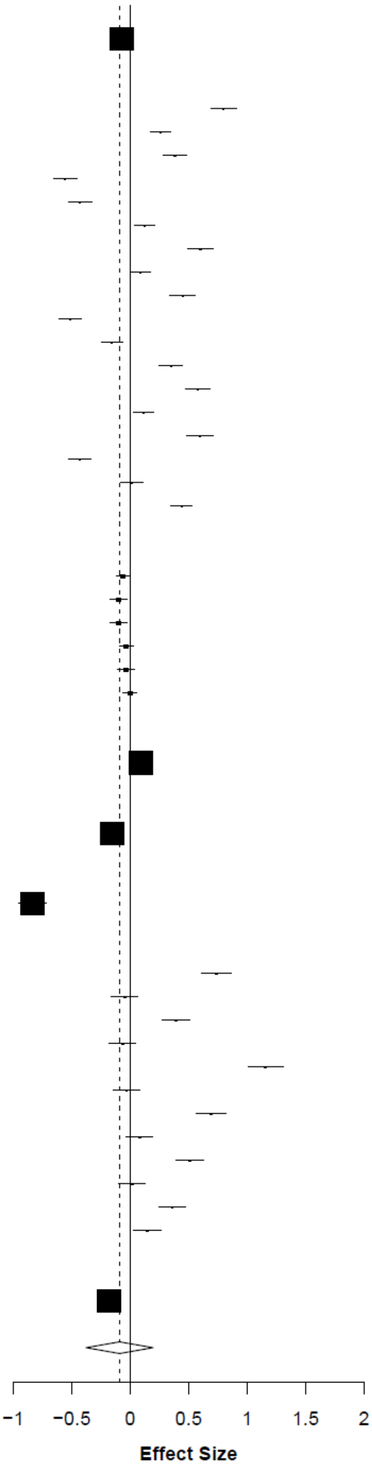

Forest Plot
Studies
Effect
Weight
Niepel, Brunner, & Preckel, 2014a; n=1529
ASC math; T1=5g −T2=5.5 −0.191 1.916
ASC math; T1=5g −T2=6.5 −0.511 1.916
ASC math; T1=5g −T2=8.5 −0.872 1.916
ASC math; T1=5.5g −T2=6.5 −0.297 1.916
ASC math; T1=5.5g −T2=8.5 −0.634 1.916
ASC math; T1=6.5g −T2=8.5 −0.301 1.916
ASC language; T1=5g −T2=5.5 −0.200 1.916
ASC language; T1=5g −T2=6.5 −0.358 1.916
ASC language; T1=5g −T2=8.5 −0.463 1.916
ASC language; T1=5.5g −T2=6.5 −0.153 1.916
ASC language; T1=5.5g −T2=8.5 −0.255 1.916
ASC language; T1=6.5g −T2=8.5 −0.099 1.916
Niepel, Brunner, & Preckel, 2014b; n=639
ASC math; T1=5g −T2=5.75 −0.296 1.873
ASC math; T1=5g −T2=6.75 −0.541 1.873
ASC math; T1=5g −T2=7.5 −0.531 1.873
ASC math; T1=5.75g −T2=6.75 −0.229 1.873
ASC math; T1=5.75g −T2=7.5 −0.219 1.873
ASC math; T1=6.75g −T2=7.5 0.009 1.873
ASC language; T1=5g −T2=5.75 −0.224 1.873
ASC language; T1=5g −T2=6.75 −0.355 1.873
ASC language; T1=5g −T2=7.5 −0.250 1.873
ASC language; T1=5.75g −T2=6.75 −0.119 1.873
ASC language; T1=5.75g −T2=7.5 −0.024 1.873
ASC language; T1=6.75g −T2=7.5 0.099 1.873
Obach, 2010; n=115
ASC general; T1=5g − T2=6g 0.241 3.225
ASC general; T1=5g − T2=7g 0.276 3.225
ASC general; T1=5g − T2=8g 0.414 3.225
ASC general; T1=6g − T2=7g 0.037 3.225
ASC general; T1=6g − T2=8g 0.186 3.225
ASC general; T1=7g − T2=8g 0.122 3.225
Ohannessian et al., 1999; n=75
ASC general; T1=6.5g − T2=7.5g −0.091 18.223
Pajares & Graham, 1999; n=273
ASC math; T1=6g − T2=6.5g −0.048 22.295
Parker, 2010; n=78
ASC general; T1=5.5g − T2=6g 1.100 3.022
ASC general; T1=5.5g − T2=6.5g 0.300 3.022
ASC general; T1=5.5g − T2=7g 0.200 3.022
ASC general; T1=6g − T2=6.5g −0.571 3.022
ASC general; T1=6g − T2=7g −0.643 3.022
ASC general; T1=6.5g − T2=7g −0.091 3.022
Pesu et al., 2016; n=231
ASC math; T1=7g − T2=9g −0.071 10.723
ASC language; T1=7g − T2=9g −0.130 10.723
Pinxten et al., 2013; n=4724
ASC math; T1=4.75g − T2=5.75g −0.115 3.875
ASC math; T1=4.75g − T2=6.75g −0.106 3.875
ASC math; T1=4.75g − T2=7.75g −0.265 3.875
ASC math; T1=5.75g − T2=6.75g 0.009 3.875
ASC math; T1=5.75g − T2=7.75g −0.148 3.875
ASC math; T1=6.75g − T2=7.75g −0.170 3.875
Reyes et al., 2000; n=107
ASC general; T1=8g − T2=9g 0.097 19.505
Shapka & Keating, 2005; n=518
ASC general; T1=9g − T2=11g −0.469 22.166
Simpkins, Davis−Kean, & Eccles, 2006; n=227
ASC math; T1=6.5g − T2=10.5g −0.401 21.079
−2
−1
0
1
2
Effect Size

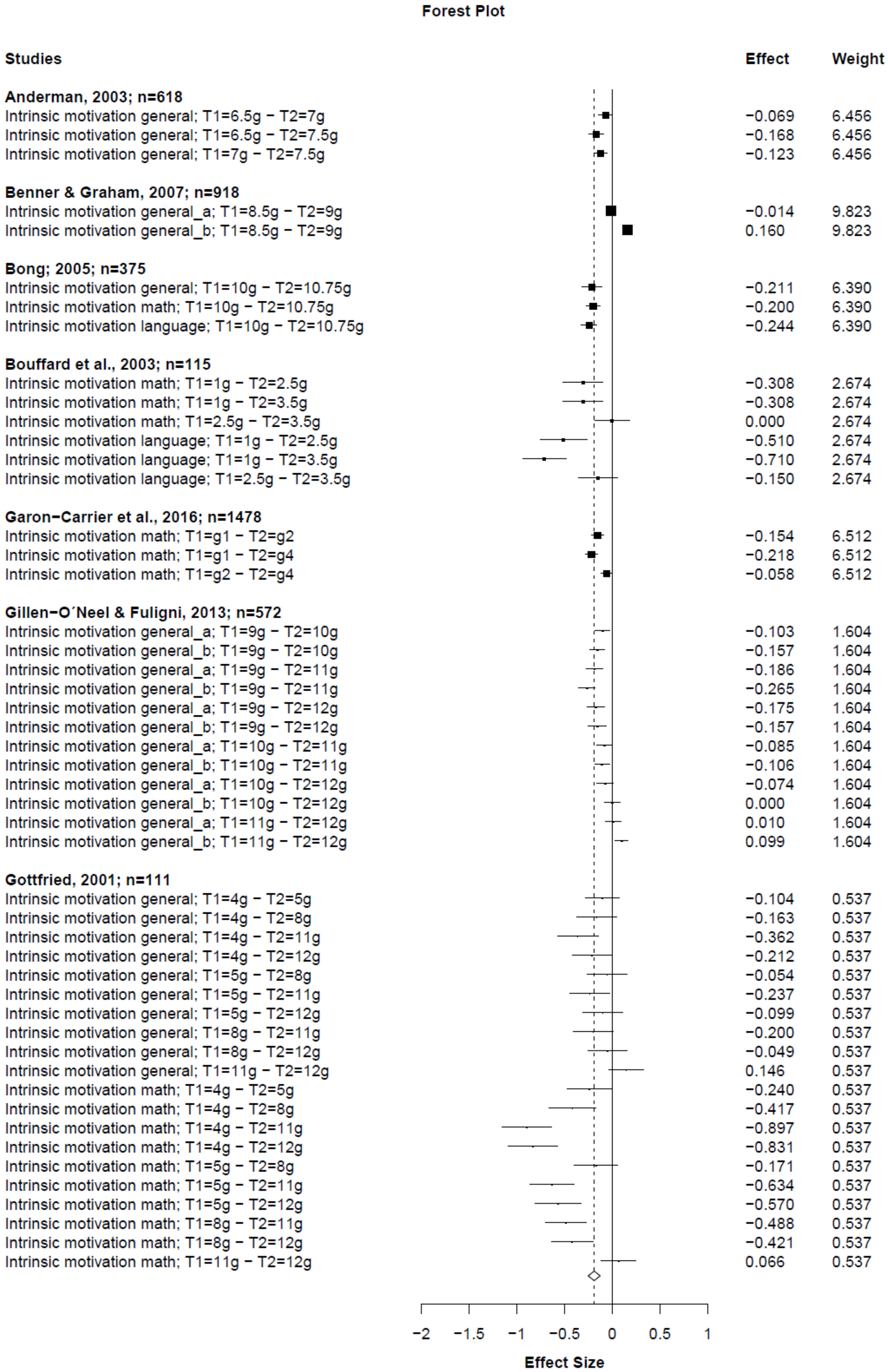

Forest Plot
Studies
Effect
Weight
Spinath & Spinath, 2005; n=789
ASC general; T1=1g − T2=1.5g −0.160 0.808
ASC general; T1=1g − T2=2g −0.380 0.808
ASC general; T1=1g − T2=2.5g −0.560 0.808
ASC general; T1=1g − T2=3g −0.760 0.808
ASC general; T1=1g − T2=3.5g −0.800 0.808
ASC general; T1=1g − T2=4g −0.880 0.808
ASC general; T1=1g − T2=4.5g −1.260 0.808
ASC general; T1=1.5g − T2=2g −0.196 0.808
ASC general; T1=1.5g − T2=2.5g −0.357 0.808
ASC general; T1=1.5g − T2=3g −0.536 0.808
ASC general; T1=1.5g − T2=3.5g −0.571 0.808
ASC general; T1=1.5g − T2=4g −0.643 0.808
ASC general; T1=1.5g − T2=4.5g −0.982 0.808
ASC general; T1=2g − T2=2.5g −0.158 0.808
ASC general; T1=2g − T2=3g −0.333 0.808
ASC general; T1=2g − T2=3.5g −0.368 0.808
ASC general; T1=2g − T2=4g −0.439 0.808
ASC general; T1=2g − T2=4.5g −0.772 0.808
ASC general; T1=2.5g − T2=3g −0.172 0.808
ASC general; T1=2.5g − T2=3.5g −0.207 0.808
ASC general; T1=2.5g − T2=4g −0.276 0.808
ASC general; T1=2.5g − T2=4.5g −0.603 0.808
ASC general; T1=3g − T2=3.5g −0.035 0.808
ASC general; T1=3g − T2=4g −0.105 0.808
ASC general; T1=3g − T2=4.5g −0.439 0.808
ASC general; T1=3.5g − T2=4g −0.073 0.808
ASC general; T1=3.5g − T2=4.5g −0.418 0.808
ASC general; T1=4g − T2=4.5g −0.358 0.808
Spinath & Steinmayr, 2008, n=670
ASC general; T1=2.75g − T2=3g −0.100 1.268
ASC general; T1=2.75g − T2=3.25g −0.160 1.268
ASC general; T1=2.75g − T2=3.5g −0.190 1.268
ASC general; T1=3g − T2=3.25g −0.070 1.268
ASC general; T1=3g − T2=3.5g −0.100 1.268
ASC general; T1=3.25g − T2=3.5g −0.030 1.268
ASC math; T1=2.75g − T2=3g −0.080 1.268
ASC math; T1=2.75g − T2=3.25g −0.100 1.268
ASC math; T1=2.75g − T2=3.5g −0.080 1.268
ASC math; T1=3g − T2=3.25g −0.020 1.268
ASC math; T1=3g − T2=3.5g 0.000 1.268
ASC math; T1=3.25g − T2=3.5g 0.020 1.268
ASC language; T1=2.75g − T2=3g −0.130 1.268
ASC language; T1=2.75g − T2=3.25g −0.160 1.268
ASC language; T1=2.75g − T2=3.5g −0.220 1.268
ASC language; T1=3g − T2=3.25g −0.030 1.268
ASC language; T1=3g − T2=3.5g −0.090 1.268
ASC language; T1=3.25g − T2=3.5g −0.060 1.268
Spinath & Steinmayr, 2012; n=348
ASC general; T1=11.5g − T2=12.5g 0.132 7.501
ASC math; T1=11.5g − T2=12.5g 0.150 7.501
ASC language; T1=11.5g − T2=12.5g 0.000 7.501
Steiger et al., 2014; n=1527
ASC general; T1=7g − T2=8g 0.059 2.299
ASC general; T1=7g − T2=9g 0.220 2.299
ASC general; T1=7g − T2=10g 0.253 2.299
ASC general; T1=7g − T2=11g 0.323 2.299
ASC general; T1=8g − T2=9g 0.160 2.299
ASC general; T1=8g − T2=10g 0.193 2.299
ASC general; T1=8g − T2=11g 0.262 2.299
ASC general; T1=9g − T2=10g 0.034 2.299
ASC general; T1=9g − T2=11g 0.107 2.299
ASC general; T1=10g − T2=11g 0.075 2.299
−2
−1
0
1
2
Effect Size

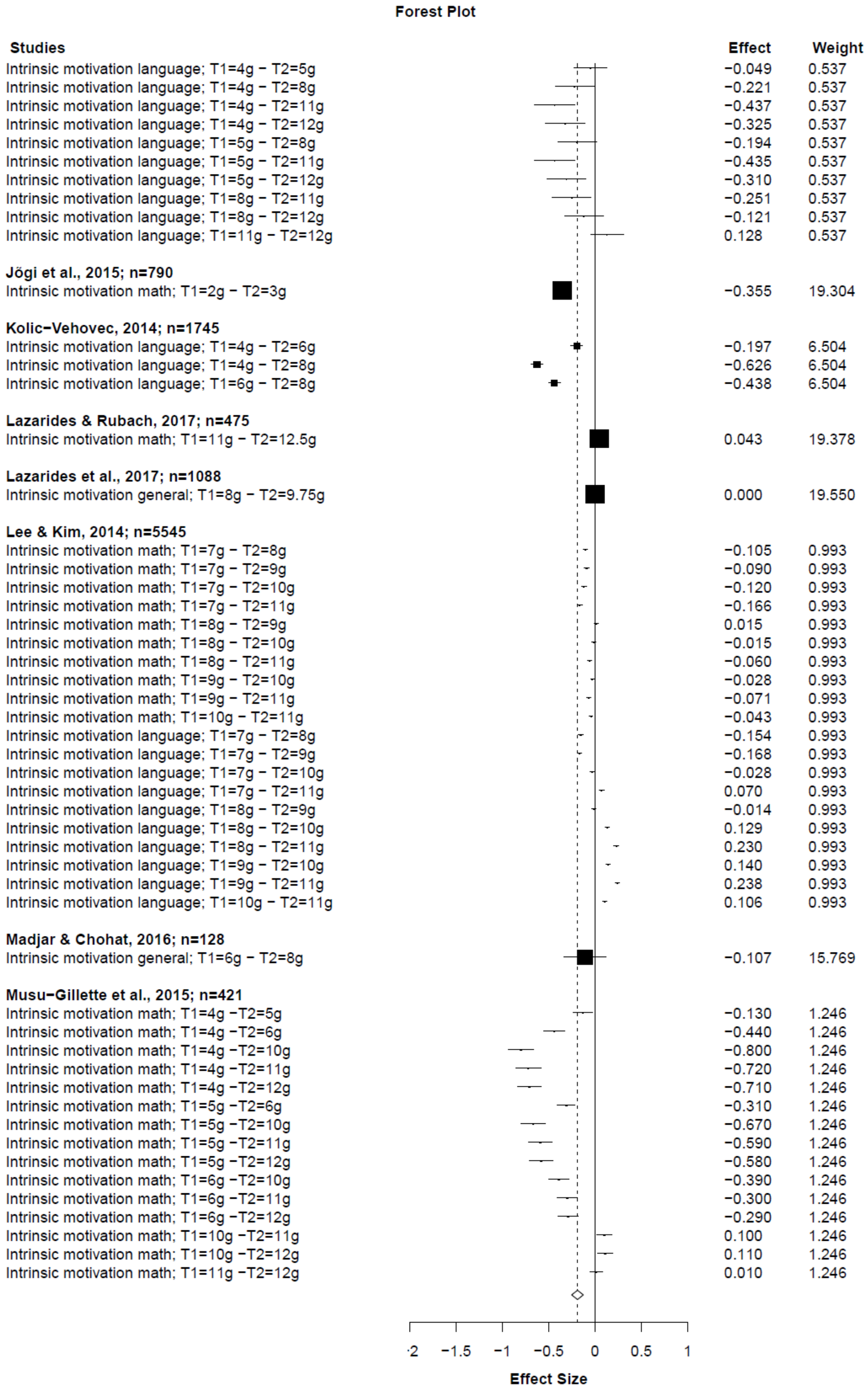

Forest Plot
Studies
Effect
Weight
Viljaranta et al., 2014; n=216
ASC math; T1=1g − T2=2g −0.070 3.482
ASC math; T1=1g − T2=4g −0.450 3.482
ASC math; T1=2g − T2=4g −0.430 3.482
ASC language; T1=1g − T2=2g −0.030 3.482
ASC language; T1=1g − T2=4g −0.480 3.482
ASC language; T1=2g − T2=4g −0.450 3.482
Wang, 2012; n=3048
ASC math; T1=7.5g − T2=10g −0.129 23.192
Weidinger, Steinmayr, & Spinath, 2017a/b; n=542
ASC math; T1=2.75g − T2=3.25g −0.088 1.072
ASC math; T1=2.75g − T2=3.5g −0.216 1.072
ASC math; T1=2.75g − T2=3.75g −0.262 1.072
ASC math; T1=2.75g − T2=4.25g −0.137 1.072
ASC math; T1=2.75g − T2=4.5g −0.261 1.072
ASC math; T1=2.75g − T2=4.75g −0.554 1.072
ASC math; T1=3.25g − T2=3.5g −0.129 1.072
ASC math; T1=3.25g − T2=3.75g −0.175 1.072
ASC math; T1=3.25g − T2=4.25g −0.049 1.072
ASC math; T1=3.25g − T2=4.5g −0.174 1.072
ASC math; T1=3.25g − T2=4.75g −0.469 1.072
ASC math; T1=3.5g − T2=3.75g −0.046 1.072
ASC math; T1=3.5g − T2=4.25g 0.078 1.072
ASC math; T1=3.5g − T2=4.5g −0.045 1.072
ASC math; T1=3.5g − T2=4.75g −0.337 1.072
ASC math; T1=3.75g − T2=4.25g 0.123 1.072
ASC math; T1=3.75g − T2=4.5g 0.001 1.072
ASC math; T1=3.75g − T2=4.75g −0.286 1.072
ASC math; T1=4.25g − T2=4.5g −0.127 1.072
ASC math; T1=4.25g − T2=4.75g −0.428 1.072
ASC math; T1=4.5g − T2=4.75g −0.287 1.072
Whitesell et al., 2009; n=1611
ASC general; T1=8g − T2=10.5g −0.026 23.004
−2
−1
0
1
2
Effect Size

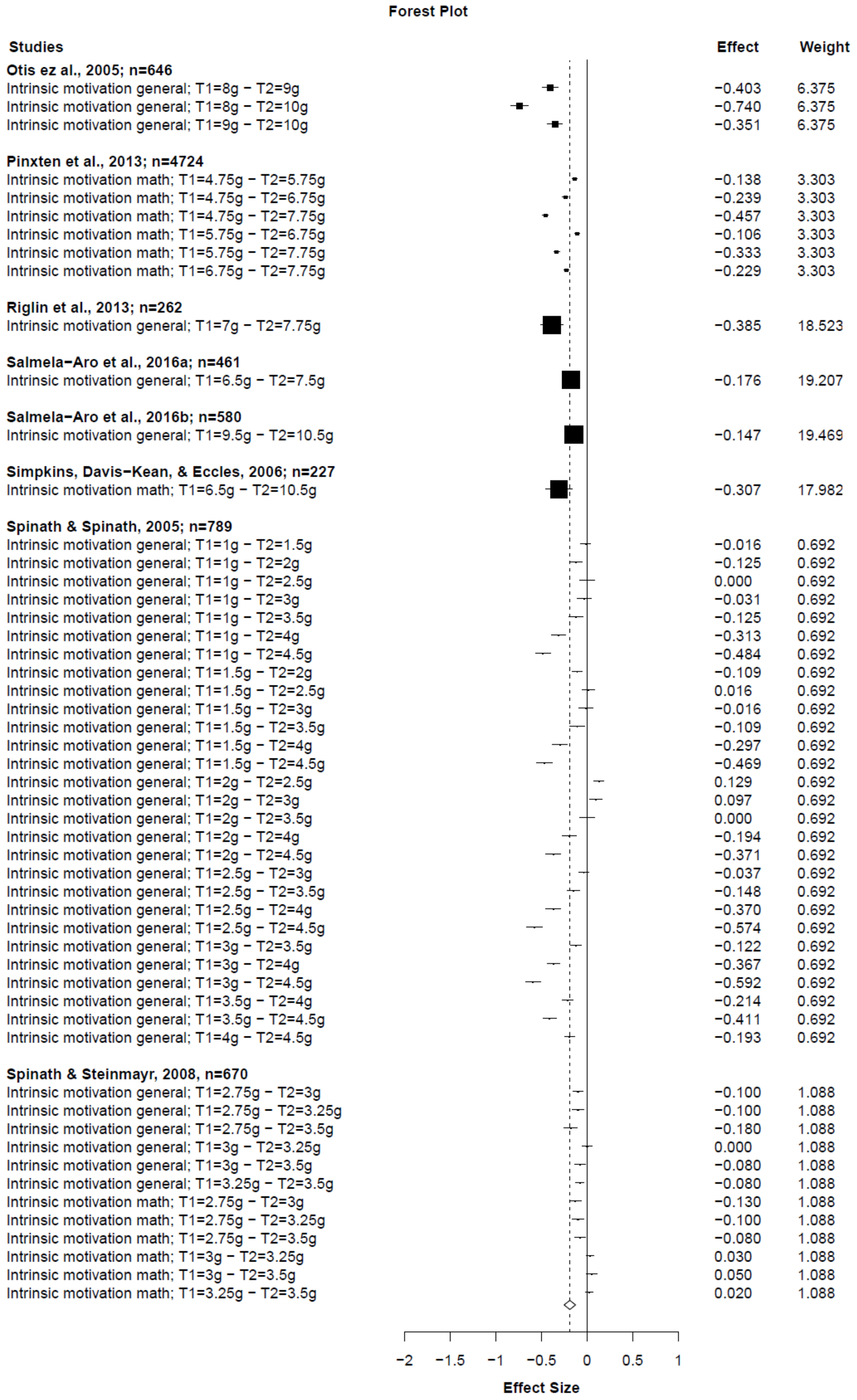
Forest Plot

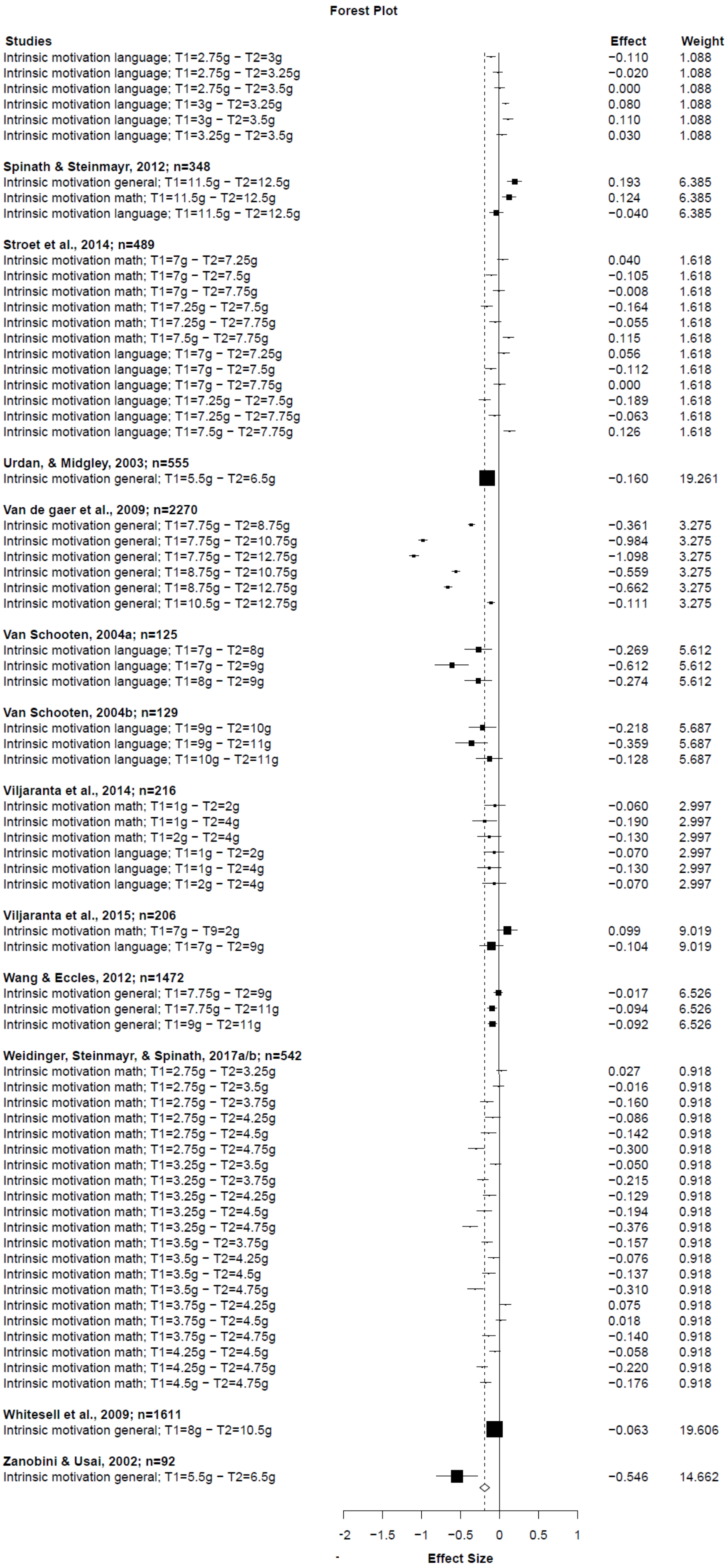
Studies	Effect	Weight
Arunkumur, Midgley, & Urdan, 1999; n=475		
Academic self−efficacy general; T1=5.5g − T2=6.5g	−0.075	8.690
Bong; 2005; n=375		
Academic self−efficacy general; T1=10g − T2=10.25g	0.800	0.483
Academic self−efficacy general; T1=10g − T2=10.5g	0.262	0.483
Academic self−efficacy general; T1=10g − T2=10.75g	0.385	0.483
Academic self−efficacy general; T1=10.25g − T2=10.	−0.556	0.483
Academic self−efficacy general; T1=10.25g − T2=10.	−0.429	0.483
Academic self−efficacy general; T1=10.5g − T2=10.7	0.123	0.483
Academic self−efficacy math; T1=10g − T2=10.25g	0.603	0.483
Academic self−efficacy math; T1=10g − T2=10.5g	0.090	0.483
Academic self−efficacy math; T1=10g − T2=10.75g	0.449	0.483
Academic self−efficacy math; T1=10.25g − T2=10.5g	−0.513	0.483
Academic self−efficacy math; T1=10.25g − T2=10.75g	−0.154	0.483
Academic self−efficacy math; T1=10.5g − T2=10.75g	0.350	0.483
Academic self−efficacy language; T1=10g − T2=10.25	0.580	0.483
Academic self−efficacy language; T1=10g − T2=10.5g	0.116	0.483
Academic self−efficacy language; T1=10g − T2=10.75	0.594	0.483
Academic self−efficacy language; T1=10.25g − T2=10	−0.432	0.483
Academic self−efficacy language; T1=10.25g − T2=10	0.014	0.483
Academic self−efficacy language; T1=10.5g − T2=10.	0.440	0.483
Friedel et al., 2010; n=929		
Academic self−efficacy math; T1=6g − T2=6.5g	−0.060	1.467
Academic self−efficacy math; T1=6g − T2=7g	−0.096	1.467
Academic self−efficacy math; T1=6g − T2=7.5g	−0.096	1.467
Academic self−efficacy math; T1=6.5g − T2=7g	−0.034	1.467
Academic self−efficacy math; T1=6.5g − T2=7.5g	−0.034	1.467
Academic self−efficacy math; T1=7g −T2=7.5g	0.000	1.467
Liew et al., 2008; n=733		
Academic self−efficacy general; T1=1g − T2=2g	0.093	8.752
Middleton, Kaplan, & Midgley, 2004; n=475		
Academic self−efficacy math; T1=6.5g − T2=7.5g	−0.151	8.698
Pajares & Graham, 1999; n=273		
Academic self−efficacy math; T1=6g − T2=6.5g	−0.833	8.624
Shell et al., 1995; n=364		
Academic self−efficacy language_a; T1=4g − T2=7g	0.740	0.717
Academic self−efficacy language_b; T1=4g − T2=7g	−0.045	0.717
Academic self−efficacy language_c; T1=4g − T2=7g	0.393	0.717
Academic self−efficacy language_d; T1=4g − T2=7g	−0.065	0.717
Academic self−efficacy language_a; T1=4g − T2=10g	1.160	0.717
Academic self−efficacy language_b; T1=4g − T2=10g	−0.030	0.717
Academic self−efficacy language_c; T1=4g − T2=10g	0.696	0.717
Academic self−efficacy language_d; T1=4g − T2=10g	0.081	0.717
Academic self−efficacy language_a; T1=7g − T2=10g	0.512	0.717
Academic self−efficacy language_b; T1=7g − T2=10g	0.018	0.717
Academic self−efficacy language_c; T1=7g − T2=10g	0.362	0.717
Academic self−efficacy language_d; T1=7g − T2=10g	0.148	0.717
Urdan, & Midgley, 2003; n=555		
Academic self−efficacy general; T1=5.5g − T2=6.5g	−0.182	8.715

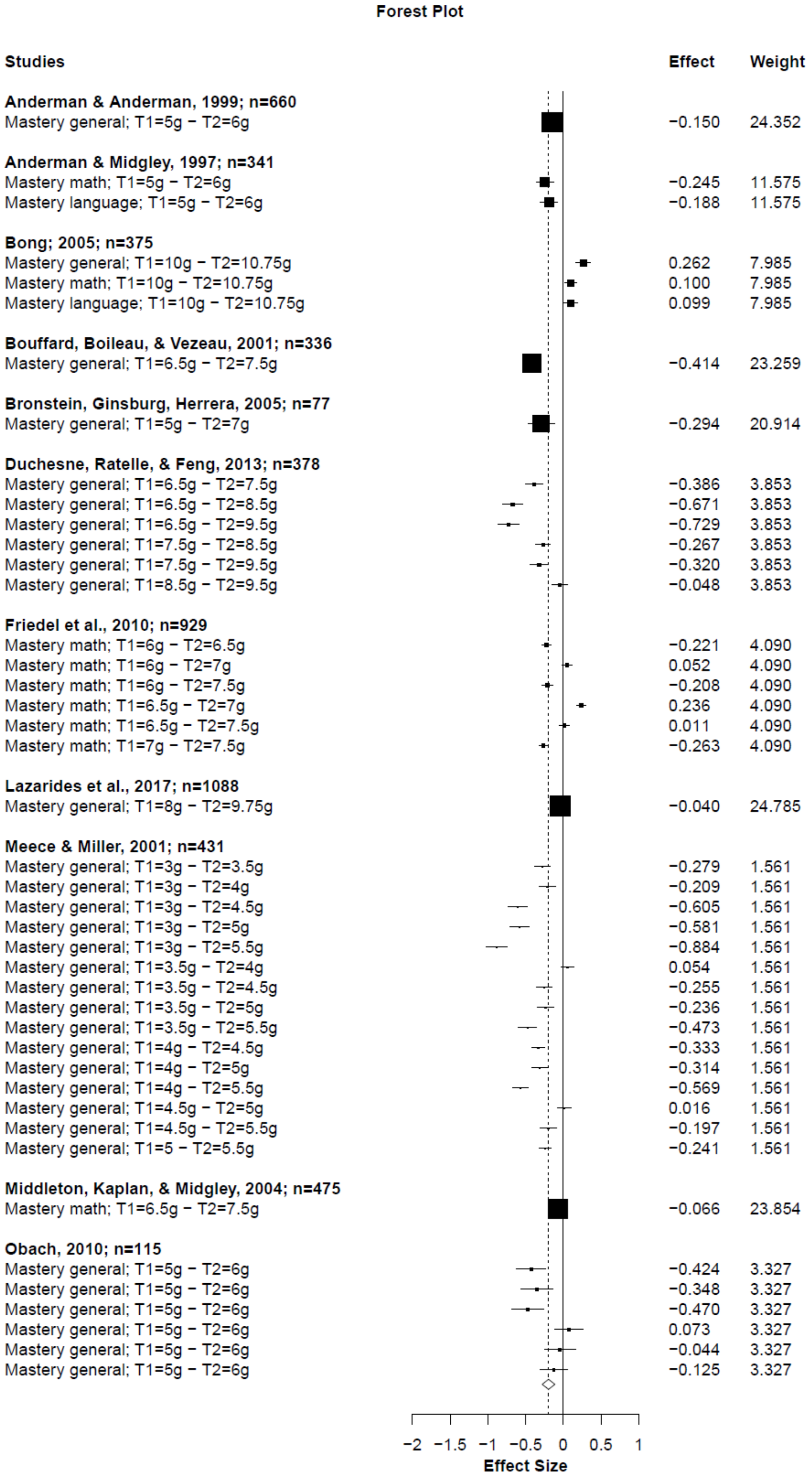


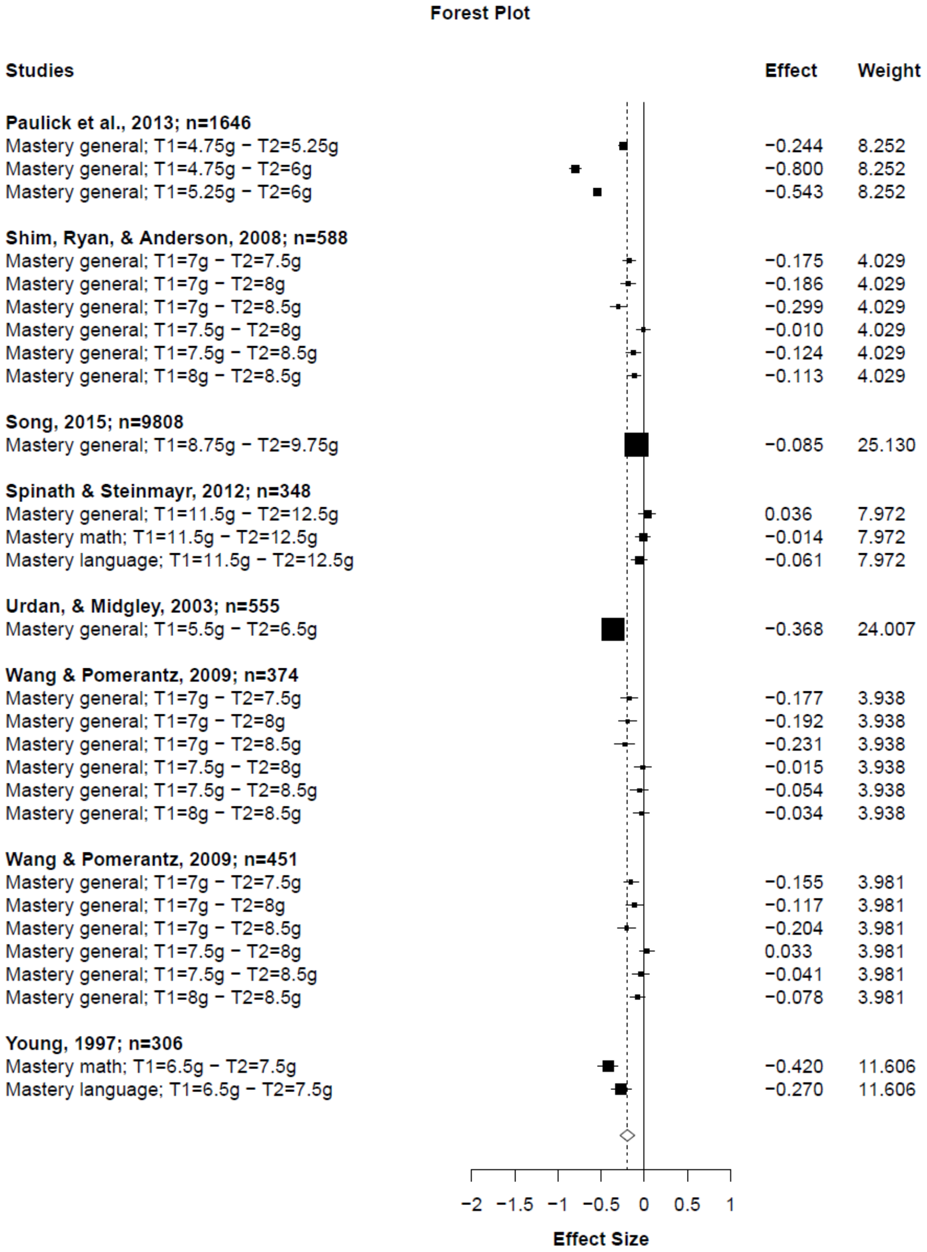

Forest Plot
Studies
Effect
Weight
Anderman, 2003; n=618
Intrinsic motivation general; T1=6.5g − T2=7g −0.069 6.456
Intrinsic motivation general; T1=6.5g − T2=7.5g −0.168 6.456
Intrinsic motivation general; T1=7g − T2=7.5g −0.123 6.456
Benner & Graham, 2007; n=918
Intrinsic motivation general_a; T1=8.5g − T2=9g −0.014 9.823
Intrinsic motivation general_b; T1=8.5g − T2=9g 0.160 9.823
Bong; 2005; n=375
Intrinsic motivation general; T1=10g − T2=10.75g −0.211 6.390
Intrinsic motivation math; T1=10g − T2=10.75g −0.200 6.390
Intrinsic motivation language; T1=10g − T2=10.75g −0.244 6.390
Bouffard et al., 2003; n=115
Intrinsic motivation math; T1=1g − T2=2.5g −0.308 2.674
Intrinsic motivation math; T1=1g − T2=3.5g −0.308 2.674
Intrinsic motivation math; T1=2.5g − T2=3.5g 0.000 2.674
Intrinsic motivation language; T1=1g − T2=2.5g −0.510 2.674
Intrinsic motivation language; T1=1g − T2=3.5g −0.710 2.674
Intrinsic motivation language; T1=2.5g − T2=3.5g −0.150 2.674
Garon−Carrier et al., 2016; n=1478
Intrinsic motivation math; T1=g1 − T2=g2 −0.154 6.512
Intrinsic motivation math; T1=g1 − T2=g4 −0.218 6.512
Intrinsic motivation math; T1=g2 − T2=g4 −0.058 6.512
Gillen−O´Neel & Fuligni, 2013; n=572
Intrinsic motivation general_a; T1=9g − T2=10g −0.103 1.604
Intrinsic motivation general_b; T1=9g − T2=10g −0.157 1.604
Intrinsic motivation general_a; T1=9g − T2=11g −0.186 1.604
Intrinsic motivation general_b; T1=9g − T2=11g −0.265 1.604
Intrinsic motivation general_a; T1=9g − T2=12g −0.175 1.604
Intrinsic motivation general_b; T1=9g − T2=12g −0.157 1.604
Intrinsic motivation general_a; T1=10g − T2=11g −0.085 1.604
Intrinsic motivation general_b; T1=10g − T2=11g −0.106 1.604
Intrinsic motivation general_a; T1=10g − T2=12g −0.074 1.604
Intrinsic motivation general_b; T1=10g − T2=12g 0.000 1.604
Intrinsic motivation general_a; T1=11g − T2=12g 0.010 1.604
Intrinsic motivation general_b; T1=11g − T2=12g 0.099 1.604
Gottfried, 2001; n=111
Intrinsic motivation general; T1=4g − T2=5g −0.104 0.537
Intrinsic motivation general; T1=4g − T2=8g −0.163 0.537
Intrinsic motivation general; T1=4g − T2=11g −0.362 0.537
Intrinsic motivation general; T1=4g − T2=12g −0.212 0.537
Intrinsic motivation general; T1=5g − T2=8g −0.054 0.537
Intrinsic motivation general; T1=5g − T2=11g −0.237 0.537
Intrinsic motivation general; T1=5g − T2=12g −0.099 0.537
Intrinsic motivation general; T1=8g − T2=11g −0.200 0.537
Intrinsic motivation general; T1=8g − T2=12g −0.049 0.537
Intrinsic motivation general; T1=11g − T2=12g 0.146 0.537
Intrinsic motivation math; T1=4g − T2=5g −0.240 0.537
Intrinsic motivation math; T1=4g − T2=8g −0.417 0.537
Intrinsic motivation math; T1=4g − T2=11g −0.897 0.537
Intrinsic motivation math; T1=4g − T2=12g −0.831 0.537
Intrinsic motivation math; T1=5g − T2=8g −0.171 0.537
Intrinsic motivation math; T1=5g − T2=11g −0.634 0.537
Intrinsic motivation math; T1=5g − T2=12g −0.570 0.537
Intrinsic motivation math; T1=8g − T2=11g −0.488 0.537
Intrinsic motivation math; T1=8g − T2=12g −0.421 0.537
Intrinsic motivation math; T1=11g − T2=12g 0.066 0.537
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size

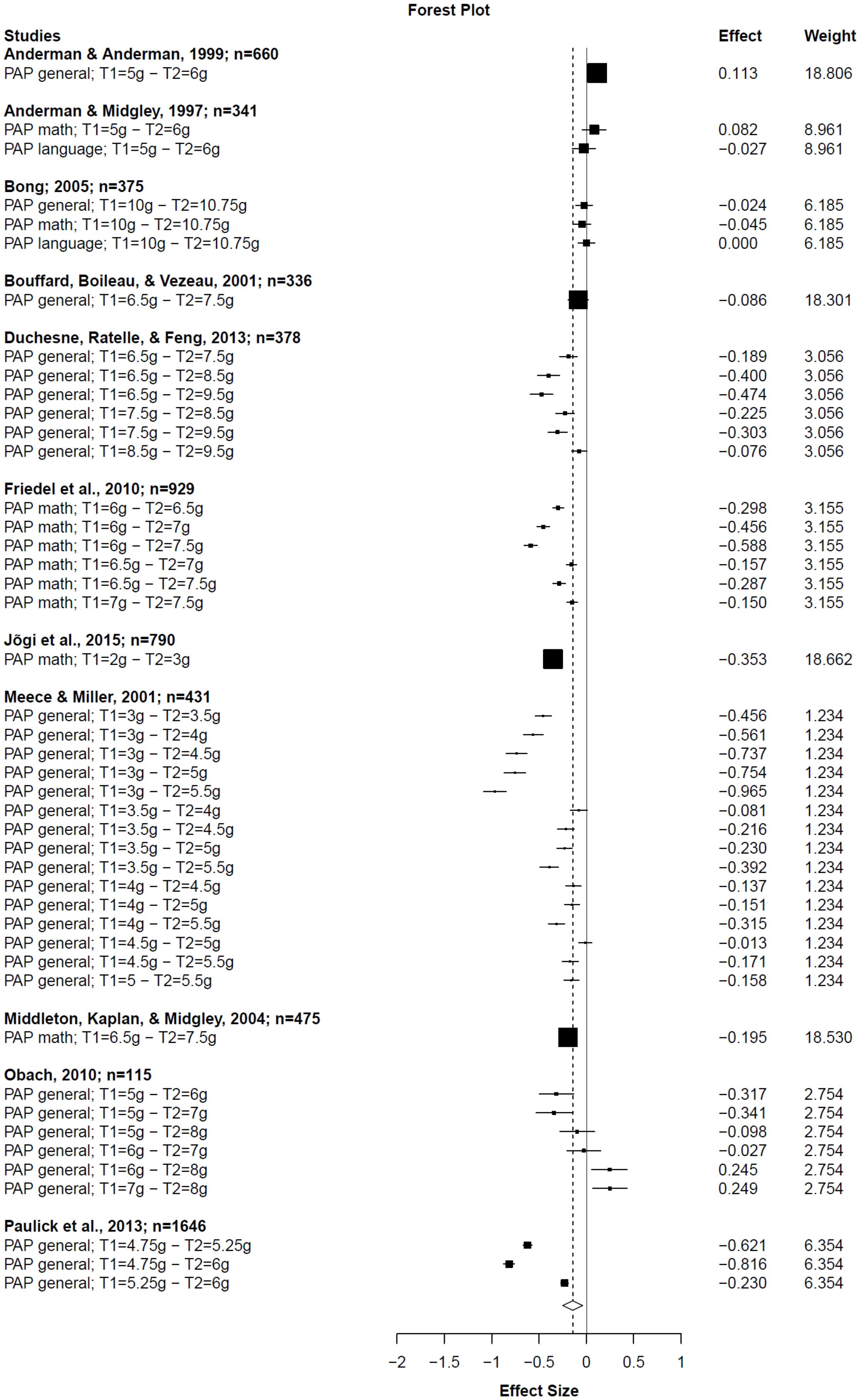

Forest Plot
Studies
Effect
Weight
Intrinsic motivation language; T1=4g − T2=5g −0.049 0.537
Intrinsic motivation language; T1=4g − T2=8g −0.221 0.537
Intrinsic motivation language; T1=4g − T2=11g −0.437 0.537
Intrinsic motivation language; T1=4g − T2=12g −0.325 0.537
Intrinsic motivation language; T1=5g − T2=8g −0.194 0.537
Intrinsic motivation language; T1=5g − T2=11g −0.435 0.537
Intrinsic motivation language; T1=5g − T2=12g −0.310 0.537
Intrinsic motivation language; T1=8g − T2=11g −0.251 0.537
Intrinsic motivation language; T1=8g − T2=12g −0.121 0.537
Intrinsic motivation language; T1=11g − T2=12g 0.128 0.537
Jõgi et al., 2015; n=790
Intrinsic motivation math; T1=2g − T2=3g −0.355 19.304
Kolic−Vehovec, 2014; n=1745
Intrinsic motivation language; T1=4g − T2=6g −0.197 6.504
Intrinsic motivation language; T1=4g − T2=8g −0.626 6.504
Intrinsic motivation language; T1=6g − T2=8g −0.438 6.504
Lazarides & Rubach, 2017; n=475
Intrinsic motivation math; T1=11g − T2=12.5g 0.043 19.378
Lazarides et al., 2017; n=1088
Intrinsic motivation general; T1=8g − T2=9.75g 0.000 19.550
Lee & Kim, 2014; n=5545
Intrinsic motivation math; T1=7g − T2=8g −0.105 0.993
Intrinsic motivation math; T1=7g − T2=9g −0.090 0.993
Intrinsic motivation math; T1=7g − T2=10g −0.120 0.993
Intrinsic motivation math; T1=7g − T2=11g −0.166 0.993
Intrinsic motivation math; T1=8g − T2=9g 0.015 0.993
Intrinsic motivation math; T1=8g − T2=10g −0.015 0.993
Intrinsic motivation math; T1=8g − T2=11g −0.060 0.993
Intrinsic motivation math; T1=9g − T2=10g −0.028 0.993
Intrinsic motivation math; T1=9g − T2=11g −0.071 0.993
Intrinsic motivation math; T1=10g − T2=11g −0.043 0.993
Intrinsic motivation language; T1=7g − T2=8g −0.154 0.993
Intrinsic motivation language; T1=7g − T2=9g −0.168 0.993
Intrinsic motivation language; T1=7g − T2=10g −0.028 0.993
Intrinsic motivation language; T1=7g − T2=11g 0.070 0.993
Intrinsic motivation language; T1=8g − T2=9g −0.014 0.993
Intrinsic motivation language; T1=8g − T2=10g 0.129 0.993
Intrinsic motivation language; T1=8g − T2=11g 0.230 0.993
Intrinsic motivation language; T1=9g − T2=10g 0.140 0.993
Intrinsic motivation language; T1=9g − T2=11g 0.238 0.993
Intrinsic motivation language; T1=10g − T2=11g 0.106 0.993
Madjar & Chohat, 2016; n=128
Intrinsic motivation general; T1=6g − T2=8g −0.107 15.769
Musu−Gillette et al., 2015; n=421
Intrinsic motivation math; T1=4g −T2=5g −0.130 1.246
Intrinsic motivation math; T1=4g −T2=6g −0.440 1.246
Intrinsic motivation math; T1=4g −T2=10g −0.800 1.246
Intrinsic motivation math; T1=4g −T2=11g −0.720 1.246
Intrinsic motivation math; T1=4g −T2=12g −0.710 1.246
Intrinsic motivation math; T1=5g −T2=6g −0.310 1.246
Intrinsic motivation math; T1=5g −T2=10g −0.670 1.246
Intrinsic motivation math; T1=5g −T2=11g −0.590 1.246
Intrinsic motivation math; T1=5g −T2=12g −0.580 1.246
Intrinsic motivation math; T1=6g −T2=10g −0.390 1.246
Intrinsic motivation math; T1=6g −T2=11g −0.300 1.246
Intrinsic motivation math; T1=6g −T2=12g −0.290 1.246
Intrinsic motivation math; T1=10g −T2=11g 0.100 1.246
Intrinsic motivation math; T1=10g −T2=12g 0.110 1.246
Intrinsic motivation math; T1=11g −T2=12g 0.010 1.246
-2 −1.5 −1 −0.5 0 0.5 1
Effect Size

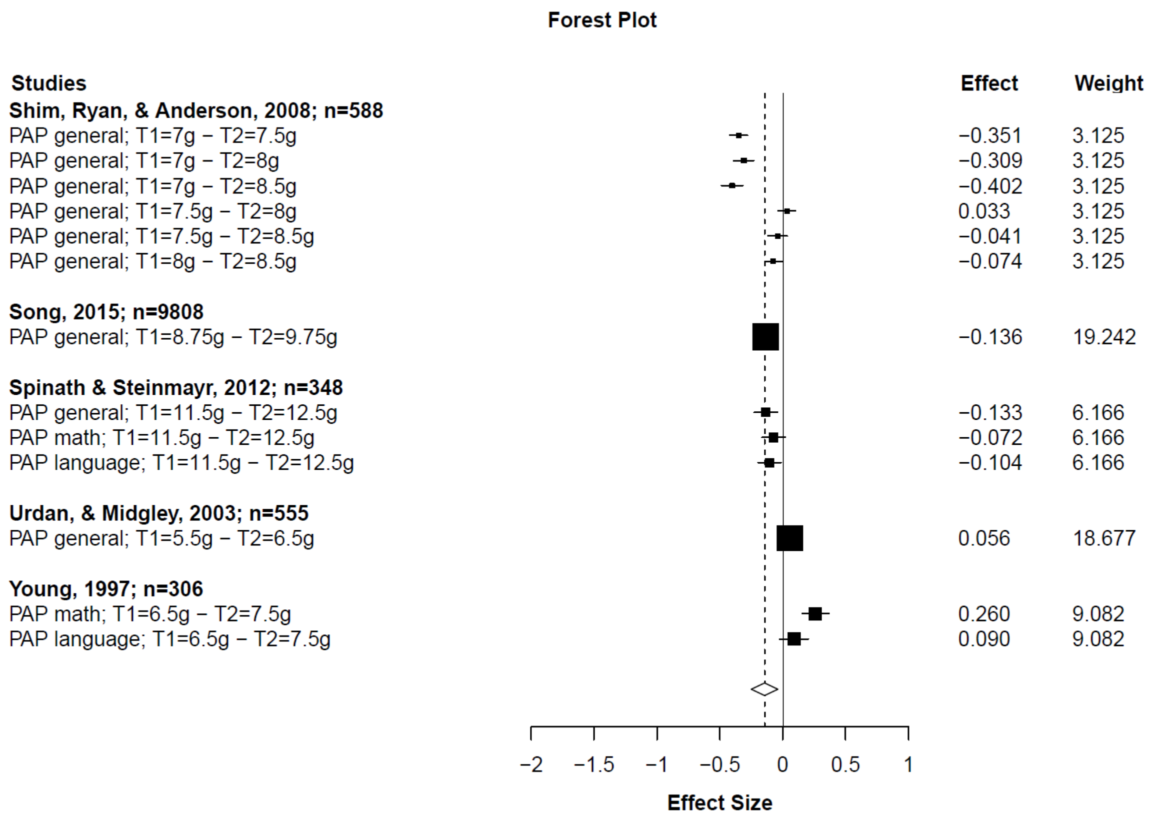

Forest Plot
Studies
Effect
Weight
Otis ez al., 2005; n=646
Intrinsic motivation general; T1=8g − T2=9g −0.403 6.375
Intrinsic motivation general; T1=8g − T2=10g −0.740 6.375
Intrinsic motivation general; T1=9g − T2=10g −0.351 6.375
Pinxten et al., 2013; n=4724
Intrinsic motivation math; T1=4.75g − T2=5.75g −0.138 3.303
Intrinsic motivation math; T1=4.75g − T2=6.75g −0.239 3.303
Intrinsic motivation math; T1=4.75g − T2=7.75g −0.457 3.303
Intrinsic motivation math; T1=5.75g − T2=6.75g −0.106 3.303
Intrinsic motivation math; T1=5.75g − T2=7.75g −0.333 3.303
Intrinsic motivation math; T1=6.75g − T2=7.75g −0.229 3.303
Riglin et al., 2013; n=262
Intrinsic motivation general; T1=7g − T2=7.75g −0.385 18.523
Salmela−Aro et al., 2016a; n=461
Intrinsic motivation general; T1=6.5g − T2=7.5g −0.176 19.207
Salmela−Aro et al., 2016b; n=580
Intrinsic motivation general; T1=9.5g − T2=10.5g −0.147 19.469
Simpkins, Davis−Kean, & Eccles, 2006; n=227
Intrinsic motivation math; T1=6.5g − T2=10.5g −0.307 17.982
Spinath & Spinath, 2005; n=789
Intrinsic motivation general; T1=1g − T2=1.5g −0.016 0.692
Intrinsic motivation general; T1=1g − T2=2g −0.125 0.692
Intrinsic motivation general; T1=1g − T2=2.5g 0.000 0.692
Intrinsic motivation general; T1=1g − T2=3g −0.031 0.692
Intrinsic motivation general; T1=1g − T2=3.5g −0.125 0.692
Intrinsic motivation general; T1=1g − T2=4g −0.313 0.692
Intrinsic motivation general; T1=1g − T2=4.5g −0.484 0.692
Intrinsic motivation general; T1=1.5g − T2=2g −0.109 0.692
Intrinsic motivation general; T1=1.5g − T2=2.5g 0.016 0.692
Intrinsic motivation general; T1=1.5g − T2=3g −0.016 0.692
Intrinsic motivation general; T1=1.5g − T2=3.5g −0.109 0.692
Intrinsic motivation general; T1=1.5g − T2=4g −0.297 0.692
Intrinsic motivation general; T1=1.5g − T2=4.5g −0.469 0.692
Intrinsic motivation general; T1=2g − T2=2.5g 0.129 0.692
Intrinsic motivation general; T1=2g − T2=3g 0.097 0.692
Intrinsic motivation general; T1=2g − T2=3.5g 0.000 0.692
Intrinsic motivation general; T1=2g − T2=4g −0.194 0.692
Intrinsic motivation general; T1=2g − T2=4.5g −0.371 0.692
Intrinsic motivation general; T1=2.5g − T2=3g −0.037 0.692
Intrinsic motivation general; T1=2.5g − T2=3.5g −0.148 0.692
Intrinsic motivation general; T1=2.5g − T2=4g −0.370 0.692
Intrinsic motivation general; T1=2.5g − T2=4.5g −0.574 0.692
Intrinsic motivation general; T1=3g − T2=3.5g −0.122 0.692
Intrinsic motivation general; T1=3g − T2=4g −0.367 0.692
Intrinsic motivation general; T1=3g − T2=4.5g −0.592 0.692
Intrinsic motivation general; T1=3.5g − T2=4g −0.214 0.692
Intrinsic motivation general; T1=3.5g − T2=4.5g −0.411 0.692
Intrinsic motivation general; T1=4g − T2=4.5g −0.193 0.692
Spinath & Steinmayr, 2008, n=670
Intrinsic motivation general; T1=2.75g − T2=3g −0.100 1.088
Intrinsic motivation general; T1=2.75g − T2=3.25g −0.100 1.088
Intrinsic motivation general; T1=2.75g − T2=3.5g −0.180 1.088
Intrinsic motivation general; T1=3g − T2=3.25g 0.000 1.088
Intrinsic motivation general; T1=3g − T2=3.5g −0.080 1.088
Intrinsic motivation general; T1=3.25g − T2=3.5g −0.080 1.088
Intrinsic motivation math; T1=2.75g − T2=3g −0.130 1.088
Intrinsic motivation math; T1=2.75g − T2=3.25g −0.100 1.088
Intrinsic motivation math; T1=2.75g − T2=3.5g −0.080 1.088
Intrinsic motivation math; T1=3g − T2=3.25g 0.030 1.088
Intrinsic motivation math; T1=3g − T2=3.5g 0.050 1.088
Intrinsic motivation math; T1=3.25g − T2=3.5g 0.020 1.088
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size


Forest Plot
Studies
Effect
Weight
Intrinsic motivation language; T1=2.75g − T2=3g −0.110 1.088
Intrinsic motivation language; T1=2.75g − T2=3.25g −0.020 1.088
Intrinsic motivation language; T1=2.75g − T2=3.5g 0.000 1.088
Intrinsic motivation language; T1=3g − T2=3.25g 0.080 1.088
Intrinsic motivation language; T1=3g − T2=3.5g 0.110 1.088
Intrinsic motivation language; T1=3.25g − T2=3.5g 0.030 1.088
Spinath & Steinmayr, 2012; n=348
Intrinsic motivation general; T1=11.5g − T2=12.5g 0.193 6.385
Intrinsic motivation math; T1=11.5g − T2=12.5g 0.124 6.385
Intrinsic motivation language; T1=11.5g − T2=12.5g −0.040 6.385
Stroet et al., 2014; n=489
Intrinsic motivation math; T1=7g − T2=7.25g 0.040 1.618
Intrinsic motivation math; T1=7g − T2=7.5g −0.105 1.618
Intrinsic motivation math; T1=7g − T2=7.75g −0.008 1.618
Intrinsic motivation math; T1=7.25g − T2=7.5g −0.164 1.618
Intrinsic motivation math; T1=7.25g − T2=7.75g −0.055 1.618
Intrinsic motivation math; T1=7.5g − T2=7.75g 0.115 1.618
Intrinsic motivation language; T1=7g − T2=7.25g 0.056 1.618
Intrinsic motivation language; T1=7g − T2=7.5g −0.112 1.618
Intrinsic motivation language; T1=7g − T2=7.75g 0.000 1.618
Intrinsic motivation language; T1=7.25g − T2=7.5g −0.189 1.618
Intrinsic motivation language; T1=7.25g − T2=7.75g −0.063 1.618
Intrinsic motivation language; T1=7.5g − T2=7.75g 0.126 1.618
Urdan, & Midgley, 2003; n=555
Intrinsic motivation general; T1=5.5g − T2=6.5g −0.160 19.261
Van de gaer et al., 2009; n=2270
Intrinsic motivation general; T1=7.75g − T2=8.75g −0.361 3.275
Intrinsic motivation general; T1=7.75g − T2=10.75g −0.984 3.275
Intrinsic motivation general; T1=7.75g − T2=12.75g −1.098 3.275
Intrinsic motivation general; T1=8.75g − T2=10.75g −0.559 3.275
Intrinsic motivation general; T1=8.75g − T2=12.75g −0.662 3.275
Intrinsic motivation general; T1=10.5g − T2=12.75g −0.111 3.275
Van Schooten, 2004a; n=125
Intrinsic motivation language; T1=7g − T2=8g −0.269 5.612
Intrinsic motivation language; T1=7g − T2=9g −0.612 5.612
Intrinsic motivation language; T1=8g − T2=9g −0.274 5.612
Van Schooten, 2004b; n=129
Intrinsic motivation language; T1=9g − T2=10g −0.218 5.687
Intrinsic motivation language; T1=9g − T2=11g −0.359 5.687
Intrinsic motivation language; T1=10g − T2=11g −0.128 5.687
Viljaranta et al., 2014; n=216
Intrinsic motivation math; T1=1g − T2=2g −0.060 2.997
Intrinsic motivation math; T1=1g − T2=4g −0.190 2.997
Intrinsic motivation math; T1=2g − T2=4g −0.130 2.997
Intrinsic motivation language; T1=1g − T2=2g −0.070 2.997
Intrinsic motivation language; T1=1g − T2=4g −0.130 2.997
Intrinsic motivation language; T1=2g − T2=4g −0.070 2.997
Viljaranta et al., 2015; n=206
Intrinsic motivation math; T1=7g − T9=2g 0.099 9.019
Intrinsic motivation language; T1=7g − T2=9g −0.104 9.019
Wang & Eccles, 2012; n=1472
Intrinsic motivation general; T1=7.75g − T2=9g −0.017 6.526
Intrinsic motivation general; T1=7.75g − T2=11g −0.094 6.526
Intrinsic motivation general; T1=9g − T2=11g −0.092 6.526
Weidinger, Steinmayr, & Spinath, 2017a/b; n=542
Intrinsic motivation math; T1=2.75g − T2=3.25g 0.027 0.918
Intrinsic motivation math; T1=2.75g − T2=3.5g −0.016 0.918
Intrinsic motivation math; T1=2.75g − T2=3.75g −0.160 0.918
Intrinsic motivation math; T1=2.75g − T2=4.25g −0.086 0.918
Intrinsic motivation math; T1=2.75g − T2=4.5g −0.142 0.918
Intrinsic motivation math; T1=2.75g − T2=4.75g −0.300 0.918
Intrinsic motivation math; T1=3.25g − T2=3.5g −0.050 0.918
Intrinsic motivation math; T1=3.25g − T2=3.75g −0.215 0.918
Intrinsic motivation math; T1=3.25g − T2=4.25g −0.129 0.918
Intrinsic motivation math; T1=3.25g − T2=4.5g −0.194 0.918
Intrinsic motivation math; T1=3.25g − T2=4.75g −0.376 0.918
Intrinsic motivation math; T1=3.5g − T2=3.75g −0.157 0.918
Intrinsic motivation math; T1=3.5g − T2=4.25g −0.076 0.918
Intrinsic motivation math; T1=3.5g − T2=4.5g −0.137 0.918
Intrinsic motivation math; T1=3.5g − T2=4.75g −0.310 0.918
Intrinsic motivation math; T1=3.75g − T2=4.25g 0.075 0.918
Intrinsic motivation math; T1=3.75g − T2=4.5g 0.018 0.918
Intrinsic motivation math; T1=3.75g − T2=4.75g −0.140 0.918
Intrinsic motivation math; T1=4.25g − T2=4.5g −0.058 0.918
Intrinsic motivation math; T1=4.25g − T2=4.75g −0.220 0.918
Intrinsic motivation math; T1=4.5g − T2=4.75g −0.176 0.918
Whitesell et al., 2009; n=1611
Intrinsic motivation general; T1=8g − T2=10.5g −0.063 19.606
Zanobini & Usai, 2002; n=92
Intrinsic motivation general; T1=5.5g − T2=6.5g −0.546 14.662
-2 −1.5 −1 −0.5 0 0.5 1
Effect Size


Forest Plot
Studies
Effect
Weight
Anderman & Anderman, 1999; n=660
Mastery general; T1=5g − T2=6g −0.150 24.352
Anderman & Midgley, 1997; n=341
Mastery math; T1=5g − T2=6g −0.245 11.575
Mastery language; T1=5g − T2=6g −0.188 11.575
Bong; 2005; n=375
Mastery general; T1=10g − T2=10.75g 0.262 7.985
Mastery math; T1=10g − T2=10.75g 0.100 7.985
Mastery language; T1=10g − T2=10.75g 0.099 7.985
Bouffard, Boileau, & Vezeau, 2001; n=336
Mastery general; T1=6.5g − T2=7.5g −0.414 23.259
Bronstein, Ginsburg, Herrera, 2005; n=77
Mastery general; T1=5g − T2=7g −0.294 20.914
Duchesne, Ratelle, & Feng, 2013; n=378
Mastery general; T1=6.5g − T2=7.5g −0.386 3.853
Mastery general; T1=6.5g − T2=8.5g −0.671 3.853
Mastery general; T1=6.5g − T2=9.5g −0.729 3.853
Mastery general; T1=7.5g − T2=8.5g −0.267 3.853
Mastery general; T1=7.5g − T2=9.5g −0.320 3.853
Mastery general; T1=8.5g − T2=9.5g −0.048 3.853
Friedel et al., 2010; n=929
Mastery math; T1=6g − T2=6.5g −0.221 4.090
Mastery math; T1=6g − T2=7g 0.052 4.090
Mastery math; T1=6g − T2=7.5g −0.208 4.090
Mastery math; T1=6.5g − T2=7g 0.236 4.090
Mastery math; T1=6.5g − T2=7.5g 0.011 4.090
Mastery math; T1=7g − T2=7.5g −0.263 4.090
Lazarides et al., 2017; n=1088
Mastery general; T1=8g − T2=9.75g −0.040 24.785
Meece & Miller, 2001; n=431
Mastery general; T1=3g − T2=3.5g −0.279 1.561
Mastery general; T1=3g − T2=4g −0.209 1.561
Mastery general; T1=3g − T2=4.5g −0.605 1.561
Mastery general; T1=3g − T2=5g −0.581 1.561
Mastery general; T1=3g − T2=5.5g −0.884 1.561
Mastery general; T1=3.5g − T2=4g 0.054 1.561
Mastery general; T1=3.5g − T2=4.5g −0.255 1.561
Mastery general; T1=3.5g − T2=5g −0.236 1.561
Mastery general; T1=3.5g − T2=5.5g −0.473 1.561
Mastery general; T1=4g − T2=4.5g −0.333 1.561
Mastery general; T1=4g − T2=5g −0.314 1.561
Mastery general; T1=4g − T2=5.5g −0.569 1.561
Mastery general; T1=4.5g − T2=5g 0.016 1.561
Mastery general; T1=4.5g − T2=5.5g −0.197 1.561
Mastery general; T1=5 − T2=5.5g −0.241 1.561
Middleton, Kaplan, & Midgley, 2004; n=475
Mastery math; T1=6.5g − T2=7.5g −0.066 23.854
Obach, 2010; n=115
Mastery general; T1=5g − T2=6g −0.424 3.327
Mastery general; T1=5g − T2=6g −0.348 3.327
Mastery general; T1=5g − T2=6g −0.470 3.327
Mastery general; T1=5g − T2=6g 0.073 3.327
Mastery general; T1=5g − T2=6g −0.044 3.327
Mastery general; T1=5g − T2=6g −0.125 3.327
−2 −1.5 −1 −0.5 0 0.5 1
Effect Size


Forest Plot
Studies
Effect
Weight
Paulick et al., 2013; n=1646
Mastery general; T1=4.75g − T2=5.25g
−0.244
8.252
Mastery general; T1=4.75g − T2=6g
−0.800
8.252
Mastery general; T1=5.25g − T2=6g
−0.543
8.252
Shim, Ryan, & Anderson, 2008; n=588
Mastery general; T1=7g − T2=7.5g
−0.175
4.029
Mastery general; T1=7g − T2=8g
−0.186
4.029
Mastery general; T1=7g − T2=8.5g
−0.299
4.029
Mastery general; T1=7.5g − T2=8g
−0.010
4.029
Mastery general; T1=7.5g − T2=8.5g
−0.124
4.029
Mastery general; T1=8g − T2=8.5g
−0.113
4.029
Song, 2015; n=9808
Mastery general; T1=8.75g − T2=9.75g
−0.085
25.130
Spinath & Steinmayr, 2012; n=348
Mastery general; T1=11.5g − T2=12.5g
0.036
7.972
Mastery math; T1=11.5g − T2=12.5g
−0.014
7.972
Mastery language; T1=11.5g − T2=12.5g
−0.061
7.972
Urdan, & Midgley, 2003; n=555
Mastery general; T1=5.5g − T2=6.5g
−0.368
24.007
Wang & Pomerantz, 2009; n=374
Mastery general; T1=7g − T2=7.5g
−0.177
3.938
Mastery general; T1=7g − T2=8g
−0.192
3.938
Mastery general; T1=7g − T2=8.5g
−0.231
3.938
Mastery general; T1=7.5g − T2=8g
−0.015
3.938
Mastery general; T1=7.5g − T2=8.5g
−0.054
3.938
Mastery general; T1=8g − T2=8.5g
−0.034
3.938
Wang & Pomerantz, 2009; n=451
Mastery general; T1=7g − T2=7.5g
−0.155
3.981
Mastery general; T1=7g − T2=8g
−0.117
3.981
Mastery general; T1=7g − T2=8.5g
−0.204
3.981
Mastery general; T1=7.5g − T2=8g
0.033
3.981
Mastery general; T1=7.5g − T2=8.5g
−0.041
3.981
Mastery general; T1=8g − T2=8.5g
−0.078
3.981
Young, 1997; n=306
Mastery math; T1=6.5g − T2=7.5g
−0.420
11.606
Mastery language; T1=6.5g − T2=7.5g
−0.270
11.606
−2
−1.5
−1
−0.5
0
0.5
1
Effect Size


Forest Plot
Studies
Effect
Weight
Anderman & Anderman, 1999; n=660
PAP general; T1=5g – T2=6g 0.113 18.806
Anderman & Midgley, 1997; n=341
PAP math; T1=5g – T2=6g 0.082 8.961
PAP language; T1=5g – T2=6g −0.027 8.961
Bong; 2005; n=375
PAP general; T1=10g – T2=10.75g −0.024 6.185
PAP math; T1=10g – T2=10.75g −0.045 6.185
PAP language; T1=10g – T2=10.75g 0.000 6.185
Bouffard, Boileau, & Vezeau, 2001; n=336
PAP general; T1=6.5g – T2=7.5g −0.086 18.301
Duchesne, Ratelle, & Feng, 2013; n=378
PAP general; T1=6.5g – T2=7.5g −0.189 3.056
PAP general; T1=6.5g – T2=8.5g −0.400 3.056
PAP general; T1=6.5g – T2=9.5g −0.474 3.056
PAP general; T1=7.5g – T2=8.5g −0.225 3.056
PAP general; T1=7.5g – T2=9.5g −0.303 3.056
PAP general; T1=8.5g – T2=9.5g −0.076 3.056
Friedel et al., 2010; n=929
PAP math; T1=6g – T2=6.5g −0.298 3.155
PAP math; T1=6g – T2=7g −0.456 3.155
PAP math; T1=6g – T2=7.5g −0.588 3.155
PAP math; T1=6.5g – T2=7g −0.157 3.155
PAP math; T1=6.5g – T2=7.5g −0.287 3.155
PAP math; T1=7g – T2=7.5g −0.150 3.155
Jõgi et al., 2015; n=790
PAP math; T1=2g – T2=3g −0.353 18.662
Meece & Miller, 2001; n=431
PAP general; T1=3g – T2=3.5g −0.456 1.234
PAP general; T1=3g – T2=4g −0.561 1.234
PAP general; T1=3g – T2=4.5g −0.737 1.234
PAP general; T1=3g – T2=5g −0.754 1.234
PAP general; T1=3g – T2=5.5g −0.965 1.234
PAP general; T1=3.5g – T2=4g −0.081 1.234
PAP general; T1=3.5g – T2=4.5g −0.216 1.234
PAP general; T1=3.5g – T2=5g −0.230 1.234
PAP general; T1=3.5g – T2=5.5g −0.392 1.234
PAP general; T1=4g – T2=4.5g −0.137 1.234
PAP general; T1=4g – T2=5g −0.151 1.234
PAP general; T1=4g – T2=5.5g −0.315 1.234
PAP general; T1=4.5g – T2=5g −0.013 1.234
PAP general; T1=4.5g – T2=5.5g −0.171 1.234
PAP general; T1=5 – T2=5.5g −0.158 1.234
Middleton, Kaplan, & Midgley, 2004; n=475
PAP math; T1=6.5g – T2=7.5g −0.195 18.530
Obach, 2010; n=115
PAP general; T1=5g – T2=6g −0.317 2.754
PAP general; T1=5g – T2=7g −0.341 2.754
PAP general; T1=5g – T2=8g −0.098 2.754
PAP general; T1=6g – T2=7g −0.027 2.754
PAP general; T1=6g – T2=8g 0.245 2.754
PAP general; T1=7g – T2=8g 0.249 2.754
Paulick et al., 2013; n=1646
PAP general; T1=4.75g – T2=5.25g −0.621 6.354
PAP general; T1=4.75g – T2=6g −0.816 6.354
PAP general; T1=5.25g – T2=6g −0.230 6.354
−2
−1.5
−1
−0.5
0
0.5
1
Effect Size


Forest Plot
Studies
Effect
Weight
Shim, Ryan, & Anderson, 2008; n=588
PAP general; T1=7g – T2=7.5g −0.351 3.125
PAP general; T1=7g – T2=8g −0.309 3.125
PAP general; T1=7g – T2=8.5g −0.402 3.125
PAP general; T1=7.5g – T2=8g 0.033 3.125
PAP general; T1=7.5g – T2=8.5g −0.041 3.125
PAP general; T1=8g – T2=8.5g −0.074 3.125
Song, 2015; n=9808
PAP general; T1=8.75g – T2=9.75g −0.136 19.242
Spinath & Steinmayr, 2012; n=348
PAP general; T1=11.5g – T2=12.5g −0.133 6.166
PAP math; T1=11.5g – T2=12.5g −0.072 6.166
PAP language; T1=11.5g – T2=12.5g −0.104 6.166
Urdan, & Midgley, 2003; n=555
PAP general; T1=5.5g – T2=6.5g 0.056 18.677
Young, 1997; n=306
PAP math; T1=6.5g – T2=7.5g 0.260 9.082
PAP language; T1=6.5g – T2=7.5g 0.090 9.082
−2
−1.5
−1
−0.5
0
0.5
1
Effect Size

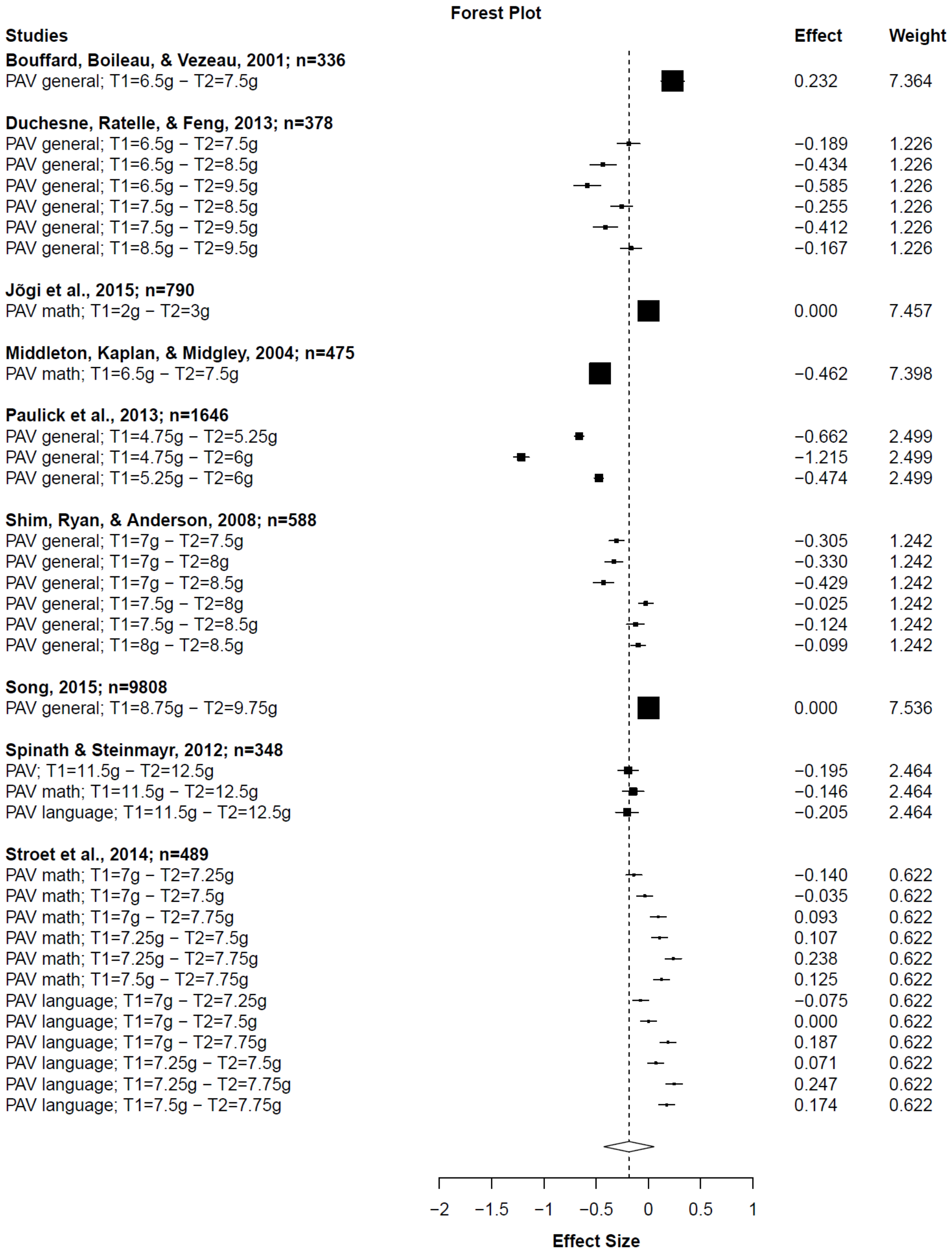


Figure A2. Forest plots of effect sizes in the dataset separated by motivational construct. ASC = academic self-concept. PAP = performance-approach achievement goals. PAV = performance-avoidance achievement goals.

References

References marked with an asterisk indicate articles included in the meta-analysis.

*Anderman, E. M., & Midgley, C. (1997). Changes in Achievement Goal Orientations, Perceived Academic Competence, and Grades across the Transition to Middle-Level Schools. *Contemporary Educational Psychology*, 22, 269–298. doi:10.1006/ceps.1996.0926

*Anderman, L. H. (2003). Academic and social perceptions as predictors of change in middle school students' sense of school belonging. *The Journal of Experimental Education*, *72*, 5-22. doi:10.1080/00220970309600877

*Anderman, L. H., & Anderman, E. M. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology, 24*, 21-37. doi:10.1006/ceps.1998.0978

Antunes, C., & Fontaine, A. (2007). Gender differences in the causal relation between adolescents' maths self-concept and scholastic performance. *Psychologica Belgica*, *47*. doi:10.5334/pb-47-1-71

*Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, 109, 621–634. doi:10.1037/edu0000163

*Arunkumar, R., Midgley, C., & Urdan, T. (1999). Perceiving High or Low Home-School Dissonance: Longitudinal Effects on Adolescent Emotional and Academic Well-Being. *Journal of Research on Adolescence*, 9, 441–466. doi:10.1207/s15327795jra0904_4

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology*, *35,* 365–380. doi:10.1037/0022-3514.35.6.365

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted Impact of Self-Efficacy Beliefs on Academic Functioning. *Child Development, 67*, 1206–1222. doi:10.1111/j.1467-8624.1996.tb01791.x

Barber, B. K., & Olsen, J. A. (2004). Assessing the Transitions to Middle and High School. *Journal of Adolescent Research, 19,* 3–30. doi:10.1177/0743558403258113

Bartle-Haring, S., Brucker, P., & Hock, E. (2002). The Impact of Parental Separation Anxiety on Identity Development in Late Adolescence and Early Adulthood. *Journal of Adolescent Research, 17*, 439–450. doi:10.1177/0743558402175001

Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology, 102,* 773–785. doi:10.1037/a0020084

*Benner, A. D., & Graham, S. (2007). Navigating the transition to multi-ethnic urban high schools: Changing ethnic congruence and adolescents' school-related affect. *Journal of Research on Adolescence, 17*, 207-220. doi:10.1111/j.1532-7795.2007.00519.x

*Birkeland, M. S., Melkevik, O., Holsen, I., & Wold, B. (2012). Trajectories of global self-esteem development during adolescence. *Journal of Adolescence*, 35, 43–54. doi:10.1016/j.adolescence.2011.06.006

*Bong, M. (2005). Within-grade changes in Korean girls' motivation and perceptions of the learning environment across domains and achievement levels. *Journal of Educational Psychology, 97,* 656. doi:10.1037/0022-0663.97.4.656

*Bouffard, T., Boileau, L., & Vezeau, C. (2001). Students' transition from elementary to high school and changes of the relationship between motivation and academic performance. *European Journal of Psychology of Education*, 16, 589–604. doi:10.1007/bf03173199

*Bouffard, T., Marcoux, M.-F., Vezeau, C., & Bordeleau, L. (2003). Changes in self-perceptions of competence and intrinsic motivation among elementary schoolchildren. *British Journal of Educational Psychology*, 73, 171–186. doi:10.1348/00070990360626921

Bouffard, T., Vezeau, C., Romano, G., Chouinard, R., Bordeleau, L., & Filion, C. (1998). Élaboration et validation d'un questionnaire d'évaluation des buts d'apprentissage des élèves. *Revue canadienne des sciences du comportement*, *30*, 203-206.

*Bronstein, P., Ginsburg, G. S., & Herrera, I. S. (2005). Parental Predictors of Motivational Orientation in Early Adolescence: A Longitudinal Study. *Journal of Youth and Adolescence*, 34, 559–575. doi:10.1007/s10964-005-8946-0

*Brown, K. M., McMahon, R. P., Biro, F. M., Crawford, P., Schreiber, G. B., Similo, S. L., … Striegel-Moore, R. (1998). Changes in self-esteem in black and white girls between the ages of 9 and 14 years. *Journal of Adolescent Health*, 23, 7–19. doi:10.1016/s1054-139x(97)00238-3

Chan, L. K. S., & Moore, P. J. (2006). Development of Attributional Beliefs and Strategic Knowledge in Years 5–9: A longitudinal analysis. *Educational Psychology, 26*, 161–185. doi:10.1080/01443410500344209

Chaplin, T. M., Freiburger, M. B., Mayes, L. C., & Sinha, R. (2010). Prenatal cocaine exposure, gender, and adolescent stress response: A prospective longitudinal study. *Neurotoxicology and Teratology, 32*, 595–604. doi:10.1016/j.ntt.2010.08.007

*Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology*, 67, 279–291. doi:10.1111/j.2044-8279.1997.tb01244.x

*Chubb, N. H., Fertman, C. I., & Ross, J. L. (1997). Adolescent self-esteem and locus of control: A logitudinal study of gender and age differences. *Adolescence*, *32*(125), 113.

*Chung, H., Elias, M., & Schneider, K. (1998). Patterns of Individual Adjustment Changes During Middle School Transition. *Journal of School Psychology*, 36, 83–101. doi:10.1016/s0022-4405(97)00051-4

Coopersmith, S. (1975). Coopersmith Self-Esteem Inventory, technical manual. Palo Alto, CA: Consulting Psychologists Press.

*Damrongpanit, S., Reungtragul, A., & Pittayanon, T. (2009). An Investigation of the Effects between Academic Self-concept, Nonacademic Self-concept, and Academic Achievement: Causal Ordering Models. *Research in Higher Education Journal*, *2*, 1.

*Daniels, E., & Leaper, C. (2006). A Longitudinal Investigation of Sport Participation, Peer Acceptance, and Self-esteem among Adolescent Girls and Boys. *Sex Roles*, 55(11-12), 875–880. doi:10.1007/s11199-006-9138-4

*DesRoches, A., & Willoughby, T. (2013). Bidirectional Associations Between Valued Activities and Adolescent Positive Adjustment in a Longitudinal Study: Positive Mood as a Mediator. *Journal of Youth and Adolescence*, 43, 208–220. doi:10.1007/s10964-013-9957-x

*Dickhäuser, O., Janke, S., Praetorius, A. K., & Dresel, M. (2017). The Effects of Teachers' Reference Norm Orientations on Students' Implicit Theories and Academic Self-Concepts. *Zeitschrift für Pädagogische Psychologie*.

DuBois, D. L., Felner, R. D., Brand, S., Phillips, R. S., & Lease, A. M. (1996). Early adolescent self-esteem: A developmental-ecological framework and assessment strategy. *Journal of research on adolescence*, *6*, 543-579.

*Duchesne, S., Ratelle, C. F., & Feng, B. (2013). Developmental Trajectories of Achievement Goal Orientations During the Middle School Transition. *The Journal of Early Adolescence*, 34, 486–517. doi:10.1177/0272431613495447

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

Elliott, M. (1996). Impact of work, family, and welfare receipt on women's self-esteem in young adulthood. *Social Psychology Quarterly, 80-95*. doi:10.2307/2787120

*Fenzel, L. M. (2000). Prospective Study of Changes in Global Self-Worth and Strain During the Transition to Middle School. *The Journal of Early Adolescence*, 20, 93–116. doi:10.1177/0272431600020001005

*Friedel, J. M., Cortina, K. S., Turner, J. C., & Midgley, C. (2010). Changes in efficacy beliefs in mathematics across the transition to middle school: Examining the effects of perceived teacher and parent goal emphases. *Journal of Educational Psychology*, 102, 102–114. doi:10.1037/a0017590

*Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J.-P., … Tremblay, R. E. (2015). Intrinsic Motivation and Achievement in Mathematics in Elementary School: A Longitudinal Investigation of Their Association. *Child Development*, 87, 165–175. doi:10.1111/cdev.12458

Getzoff, E. A. (2004). *Emotional well-being in young adults with sickle cell disease and matched comparison peers: A longitudinal study*. Unpublished Doctoral Dissertation of University of Cincinnati.

*Gillen-O'Neel, C. G.-O., & Fuligni, A. (2012). A Longitudinal Study of School Belonging and Academic Motivation Across High School. *Child Development*, 84, 678–692. doi:10.1111/j.1467-8624.2012.01862.x

*Gniewosz, B., Eccles, J. S., & Noack, P. (2011). Secondary School Transition and the Use of Different Sources of Information for the Construction of the Academic Self-concept. *Social Development*, 21, 537–557. doi:10.1111/j.1467-9507.2011.00635.x

Goldberg, M. D., & Cornell, D. G. (1998). The Influence of Intrinsic Motivation and Self-concept on Academic Achievement in Second- and Third-Grade Students. *Journal for the Education of the Gifted*, *21*, 179–205. doi:10.1177/016235329802100204

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology*, *77*, 631–645. doi:10.1037/0022-0663.77.6.631

*Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology*, 93, 3–13. doi:10.1037/0022-0663.93.1.3

*Granleese, J., & Joseph, S. (1994). Reliability of the Harter Self-Perception Profile for Children and Predictors of Global Self-Worth. *The Journal of Genetic Psychology*, 155, 487–492. doi:10.1080/00221325.1994.9914796

*Greene, M. L., Way, N., & Pahl, K. (2006). Trajectories of perceived adult and peer discrimination among Black, Latino, and Asian American adolescents: Patterns and psychological correlates. *Developmental Psychology*, 42, 218–236. doi:10.1037/0012-1649.42.2.218

*Gruenenfelder-Steiger, A. E., Harris, M. A., & Fend, H. A. (2016). Subjective and objective peer approval evaluations and self-esteem development: A test of reciprocal, prospective, and long-term effects. *Developmental Psychology*, 52, 1563–1577. doi:10.1037/dev0000147

Guay, F., Boivin, M., & Hodges, E. V. E. (1999). Predicting change in academic achievement: A model of peer experiences and self-system processes. *Journal of Educational Psychology, 91*, 105–115. doi:10.1037/0022-0663.91.1.105

*Gutman, L. M., & Eccles, J. S. (2007). Stage-environment fit during adolescence: Trajectories of family relations and adolescent outcomes. *Developmental Psychology*, 43, 522–537. doi:10.1037/0012-1649.43.2.522

*Harris, M. A., Gruenenfelder-Steiger, A. E., Ferrer, E., Donnellan, M. B., Allemand, M., Fend, H., … Trzesniewski, K. H. (2015). Do Parents Foster Self-Esteem? Testing the Prospective Impact of Parent Closeness on Adolescent Self-Esteem. *Child Development*, 86, 995–1013. doi:10.1111/cdev.12356

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17*, 300–312. doi:10.1037/0012-1649.17.3.300

Harter, S. (1985). Competence as a dimension of self-evaluation. Toward a comprehensive model of self-worth. In R. Leahy (Ed.), *The development of the self* (pp. 55–121). New York: Academic Press.

*Heinonen, K., Räikkönen, K., Keskivaara, P., & Keltikangas-Järvinen, L. (2002). Difficult temperament predicts self-esteem in adolescence. *European Journal of Personality*, 16, 439–455. doi:10.1002/per.464

*Helmke, A., & van Aken, M. A. G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology*, 87, 624–637. doi:10.1037/0022-0663.87.4.624

Heo, G., & Kim, T. (2015). Autoregressive Cross-Lagged Modeling of the Reciprocal Longitudinal Relationship Between Self-Esteem and Career Maturity. *Journal of Career Development*, 43, 273–288. doi:10.1177/0894845315598002

Hoglund, C. L. (1995). *Longitudinal study of gender differences in global and domain-specific self-esteem in children.* Unpublished Doctoral Dissertation of California School of Professional Psychology - Los Angeles.

*Impett, E. A., Henson, J. M., Breines, J. G., Schooler, D., & Tolman, D. L. (2011). Embodiment Feels Better. *Psychology of Women Quarterly*, 35, 46–58. doi:10.1177/0361684310391641

*Impett, E. A., Sorsoli, L., Schooler, D., Henson, J. M., & Tolman, D. L. (2008). Girls' relationship authenticity and self-esteem across adolescence. *Developmental Psychology*, *44*, 722. doi:10.1037/0012-1649.44.3.722

Jiang, X., Huebner, E. S., & Siddall, J. (2012). A Short-Term Longitudinal Study of Differential Sources of School-Related Social Support and Adolescents' School Satisfaction. *Social Indicators Research, 114,* 1073–1086. doi:10.1007/s11205-012-0190-x

*Jõgi, A.-L., Kikas, E., Lerkkanen, M.-K., & Mägi, K. (2015). Cross-lagged relations between math-related interest, performance goals and skills in groups of children with different general abilities. *Learning and Individual Differences,* 39, 105–113. doi:10.1016/j.lindif.2015.03.018

*Juang, L. P., Syed, M., & Cookston, J. T. (2012). Acculturation-based and everyday parent–adolescent conflict among Chinese American adolescents: Longitudinal trajectories and implications for mental health. Journal of Family Psychology, 26, 916–926. doi:10.1037/a0030057

Keltikangas-Järvinen, L. (1990). The Stability of Self-Concept During Adolescence and Early Adulthood: A Six-Year Follow-Up Study. *The Journal of General Psychology, 117*, 361–368. doi:10.1080/00221309.1990.9921142

*Kolić-Vehovec, S., Rončević Zubković, B., & Pahljina-Reinić, R. (2014). Development of metacognitive knowledge of reading strategies and attitudes toward reading in early adolescence: The effect on reading comprehension. *Psihologijske teme*, *23*(1), 77-98.

Köller, O., & Baumert, J. (1998). Ein deutsches Instrument zur Erfassung von Zielorientierungen bei Schülerinnen und Schülern [A German questionnaire on pupils' goal orientations]. *Diagnostica*, *44*, 173–181.

*Kurtz-Costes, B. E., & Schneider, W. (1994). Self-Concept, Attributional Beliefs, and School Achievement: A Longitudinal Analysis. Contemporary Educational Psychology, 19(2), 199–216. doi:10.1006/ceps.1994.1017

*Lazarides, R., & Raufelder, D. (2017). Longitudinal Effects of Student-Perceived Classroom Support on Motivation – A Latent Change Model. *Frontiers in Psychology*, 8. doi:10.3389/fpsyg.2017.00417

*Lazarides, R., Rubach, C., & Ittel, A. (2017). Adolescents' perceptions of socializers' beliefs, career-related conversations, and motivation in mathematics. *Developmental Psychology*, 53, 525–539. doi:10.1037/dev0000270

*Lee, H., & Kim, Y. (2014). Korean adolescents' longitudinal change of intrinsic motivation in learning English and mathematics during secondary school years: Focusing on gender difference and school characteristics. *Learning and Individual Differences*, 36, 131–139. doi:10.1016/j.lindif.2014.07.018

*Liew, J., McTigue, E. M., Barrois, L., & Hughes, J. N. (2008). Adaptive and effortful control and academic self-efficacy beliefs on achievement: A longitudinal study of 1st through 3rd graders. *Early Childhood Research Quarterly*, 23, 515–526. doi:10.1016/j.ecresq.2008.07.003

*Lintunen, T., Leskinen, E., Oinonen, M., Salinto, M., & Rahkila, P. (1995). Change, Reliability, and Stability in Self-perceptions in Early Adolescence: A Four-year Follow-up Study. *International Journal of Behavioral Development*, 18, 351–364. doi:10.1177/016502549501800210

*Madjar, N., & Chohat, R. (2016). Will I succeed in middle school? A longitudinal analysis of self-efficacy in school transitions in relation to goal structures and engagement. *Educational Psychology*, 37, 680–694. doi:10.1080/01443410.2016.1179265

*Maltais, C., Duchesne, S., Ratelle, C. F., & Feng, B. (2017). Learning climate, academic competence, and anxiety during the transition to middle school: Parental attachment as a protective factor. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *67*(2), 103-112.

*Marsh, H. W., & Grayson, D. (1994). Longitudinal stability of latent means and individual differences: A unified approach. Structural Equation Modeling: *A Multidisciplinary Journal*, 1, 317–359. doi:10.1080/10705519409539984

*Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child development*, *69*, 1030-1053. doi:10.2307/1132361

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and Social Psychology*, *45*, 173–187. doi:10.1037/0022-3514.45.1.173

*McGrath, E. P., & Repetti, R. L. (2002). A longitudinal study of children's depressive symptoms, self-perceptions, and cognitive distortions about the self. *Journal of abnormal psychology*, *111*, 77. doi:0021-843x.111.1.77

*Meece, J. L., & Miller, S. D. (2001). A longitudinal analysis of elementary school students' achievement goals in literacy activities. *Contemporary Educational Psychology*, *26*, 454-480. doi:10.1006/ceps.2000.1071

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of educational psychology*, *80*, 514. doi:10.1037//0022-0663.80.4.514

*Middleton, M. J., Kaplan, A., & Midgley, C. (2004). The Change in Middle School Students' Achievement Goals in Mathematics Over Time. *Social Psychology of Education*, 7, 289–311. doi:10.1023/b:spoe.0000037484.86850.fa

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., … Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor: University of Michigan. Retrieved from http://www.umich.edu/~pals/PALS%202000_V12Word97.pdf

*Molloy, L. E., Ram, N., & Gest, S. D. (2011). The storm and stress (or calm) of early adolescent self-concepts: Within- and between-subjects variability. *Developmental Psychology*, 47, 1589–1607. doi:10.1037/a0025413

*Muldoon, O. T. (2000). Social group membership and self-perceptions in Northern Irish children: A longitudinal study. *British Journal of Developmental Psychology*, 18, 65–80. doi:10.1348/026151000165571

*Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, 21, 343–370. doi:10.1080/13803611.2015.1057161

*Nagy, G., Watt, H. M., 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,* 482-506. doi:10.1111/j.1532-7795.2010.00644.x

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

*Obach, M. S. (2003). A Longitudinal-Sequential Study of Perceived Academic Competence and Motivational Beliefs for Learning Among Children in Middle School. *Educational Psychology*, 23, 323–338. doi:10.1080/0144341032000060156

*Ohannessian, C. M., LERNER, R. M., LERNER, J. V., & VON EYE, A. (1999). Does self-competence predict gender differences in adolescent depression and anxiety? *Journal of Adolescence*, 22, 397–411. doi:10.1006/jado.1999.0231

*Otis, N., Grouzet, F. M. E., & Pelletier, L. G. (2005). Latent Motivational Change in an Academic Setting: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 97, 170–183. doi:10.1037/0022-0663.97.2.170

*Pajares, F., & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology,* 24, 124–139. doi:10.1006/ceps.1998.0991

*Parker, A. K. (2010). A Longitudinal Investigation of Young Adolescents' Self-Concepts in the Middle Grades. *RMLE Online,* 33, 1–13. doi:10.1080/19404476.2010.11462073

*Paulick, I., Watermann, R., & Nückles, M. (2013). Achievement goals and school achievement: The transition to different school tracks in secondary school. Contemporary Educational Psychology, 38, 75–86. doi:10.1016/j.cedpsych.2012.10.003

*Pesu, L. A., Aunola , K., Viljaranta, J., & Nurmi, J.-E. (2016). The development of adolescents' self-concept of ability through grades 7-9 and the role of parental beliefs. *Frontline Learning Research,* 4, 92–109. doi:10.14786/flr.v4i3.249

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33–40. doi:10.1037/0022-0663.82.1.33

*Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2013). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology,* 84, 152–174. doi:10.1111/bjep.12028

Pullmann, H., & Allik, J. (2000). The Rosenberg Self-Esteem Scale: its dimensionality, stability and personality correlates in Estonian. *Personality and Individual Differences, 28,* 701–715. doi:10.1016/s0191-8869(99)00132-4

*Reddy, R., RHODES, J. E., & MULHALL, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology,* 15, 119–138. doi:10.1017/s0954579403000075

*Reyes, O., Gillock, K. L., Kobus, K., & Sanchez, B. (2000). A Longitudinal Examination of the Transition into Senior High School for Adolescents from Urban, Low-Income Status, and Predominantly Minority Backgrounds. *American Journal of Community Psychology,* 28, 519–544. doi:10.1023/a:1005140631988

*Riglin, L., Frederickson, N., Shelton, K. H., & Rice, F. (2013). A longitudinal study of psychological functioning and academic attainment at the transition to secondary school. *Journal of Adolescence,* 36, 507–517. doi:10.1016/j.adolescence.2013.03.002

Roberts, B. W., & Robins, R. W. (2004). Person-Environment Fit and Its Implications for Personality Development: A Longitudinal Study. *Journal of Personality, 72,* 89–110. doi:10.1111/j.0022-3506.2004.00257.x

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of educational psychology, 88,* 408–422. doi: 10.1037//0022-0663.88.3.408

Rosenberg, M. (1989). *Society and the adolescent self-im*. (Revised edition.). Middletown, CT: Wesleyen University Press.

*Salmela-Aro, K., Upadyaya, K., Hakkarainen, K., Lonka, K., & Alho, K. (2016). The Dark Side of Internet Use: Two Longitudinal Studies of Excessive Internet Use, Depressive Symptoms, School Burnout and Engagement Among Finnish Early and Late Adolescents. *Journal of Youth and Adolescence,* 46, 343–357. doi:10.1007/s10964-016-0494-2

*Schmalz, D. L., Deane, G. D., Birch, L. L., & Davison, K. K. (2007). A Longitudinal Assessment of the Links Between Physical Activity and Self-Esteem in Early Adolescent Non-Hispanic Females. *Journal of Adolescent Health*, 41, 559–565. doi:10.1016/j.jadohealth.2007.07.001

Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35-37). Windsor, UK: NFER-NELSON.

Schwinger, M., & Wild, E. (2006). Die Entwicklung von Zielorientierungen im Fach Mathematik von der 3. bis 5. Jahrgangsstufe [The development of goal orientations in mathematics in 3rd to 5th graders]. *Zeitschrift für Pädagogische Psychologi*, *20*, 269–278.

Seegers, G., Van Putten, C. M., & De Brabander, C. J. (2002). Goal orientation, perceived task outcome and task demands in mathematics tasks: Effects on students' attitude in actual task settings. *British Journal of Educational Psychology, 72,* 365–384.doi:10.1348/000709902320634366

Settles, R. E., Zapolski, T. C. B., & Smith, G. T. (2014). Longitudinal test of a developmental model of the transition to early drinking. *Journal of Abnormal Psychology, 123,* 141–151. doi:10.1037/a0035670

*Shapka, J. D., & Keating, D. P. (2005). Structure and Change in Self-Concept During Adolescence. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *37*(2), 83.

*Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology*, 87, 386–398. doi:10.1037/0022-0663.87.3.386

*Shim, S. S., Ryan, A. M., & Anderson, C. J. (2008). Achievement goals and achievement during early adolescence: Examining time-varying predictor and outcome variables in growth-curve analysis. *Journal of Educational Psychology*, 100, 655–671. doi:10.1037/0022-0663.100.3.655

*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, 70–83. doi:10.1037/0012-1649.42.1.70

*Song, J., Bong, M., Lee, K., & Kim, S. (2015). Longitudinal investigation into the role of perceived social support in adolescents' academic motivation and achievement. *Journal of Educational Psychology*, 107, 821–841. doi:10.1037/edu0000016

*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, 87–102. doi: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, 1555–1569. doi:10.1111/j.1467-8624.2008.01205.x

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

Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2002). *Skalen zur Erfassung der Lern-und Leistungsmotivation: SELLMO*. Göttingen: Hogrefe.

*Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology*, 106, 325–338. doi:10.1037/a0035133

Stein, J. A., Newcomb, M. D., & Bentler, P. . (1986). Stability and change in personality: A longitudinal study from early adolescence to young adulthood. *Journal of Research in Personality*, *20*, 276–291. doi:10.1016/0092-6566(86)90135-2

*Stroet, K., Opdenakker, M.-C., & Minnaert, A. (2014). Fostering early adolescents' motivation: a longitudinal study into the effectiveness of social constructivist, traditional and combined schools for prevocational education. *Educational Psychology*, 36, 1–25. doi:10.1080/01443410.2014.893561

*Tetzner, J., & Becker, M. (2017). Think Positive? Examining the Impact of Optimism on Academic Achievement in Early Adolescents. *Journal of Personality*, 86, 283–295. doi:10.1111/jopy.12312

*Urdan, T., & Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology*, 28, 524–551. doi:10.1016/s0361-476x(02)00060-7

*Van de gaer, Heidi Pustjens, Jan Van Damme, & Agnes De Munter. (2009). School Engagement and Language Achievement: A Longitudinal Study of Gender Differences across Secondary School. *Merrill-Palmer Quarterly*, 55, 373–405. doi:10.1353/mpq.0.0034

Van der Werf, G., Opdenakker, M.-C., & Kuyper, H. (2008). Testing a dynamic model of student and school effectiveness with a multivariate multilevel latent growth curve approach. *School Effectiveness and School Improvement, 19*, 447–462. doi:10.1080/09243450802535216

*Van Schooten, E., de Glopper, K., & Stoel, R. D. (2004). Development of attitude toward reading adolescent literature and literary reading behavior. *Poetics*, 32, 343–386. doi:10.1016/j.poetic.2004.07.001

*Vasalampi, K., Salmela-Aro, K., & Nurmi, J.-E. (2010). Education-related goal appraisals and self-esteem during the transition to secondary education: A longitudinal study. *International Journal of Behavioral Development*, 34, 481–490. doi:10.1177/0165025409359888

*Viljaranta, J., Lazarides, R., Aunola, K., Räikkönen, E., & Nurmi, J. E. (2015). The different role of mothers' and fathers' beliefs in the development of adolescents' mathematics and literacy task values. *International Journal of Gender, Science and Technology*, *7*(2).

*Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J.-E. (2014). The Developmental Dynamics between Interest, Self-concept of Ability, and Academic Performance. *Scandinavian Journal of Educational Research*, 58, 734–756. doi:10.1080/00313831.2014.904419

Wade, T. J., Thompson, V., Tashakkori, A., & Valente, E. (1989). A longitudinal analysis of sex by race differences in predictors of adolescent self-esteem. *Personality and Individual Differences, 10*, 717–729. doi:10.1016/0191-8869(89)90118-9

*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, 1643–1657. doi:10.1037/a0027247

*Wang, M.-T., & Eccles, J. S. (2012). Social Support Matters: Longitudinal Effects of Social Support on Three Dimensions of School Engagement From Middle to High School. *Child Development*, 83, 877–895. doi:10.1111/j.1467-8624.2012.01745.x

*Wang, Q., & Pomerantz, E. M. (2009). The Motivational Landscape of Early Adolescence in the United States and China: A Longitudinal Investigation. *Child Development*, 80, 1272–1287. doi:10.1111/j.1467-8624.2009.01331.x

*Way, N., & Robinson, M. G. (2003). A Longitudinal Study of the Effects of Family, Friends, and School Experiences on the Psychological Adjustment of Ethnic Minority, Low-SES Adolescents. *Journal of Adolescent Research*, 18, 324–346. doi:10.1177/0743558403018004001

*Way, N., Reddy, R., & Rhodes, J. (2007). Students' Perceptions of School Climate During the Middle School Years: Associations with Trajectories of Psychological and Behavioral Adjustment. *American Journal of Community Psychology*, 40, 194–213. doi:10.1007/s10464-007-9143-y

*Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: A longitudinal investigation of their association. *British Journal of Educational Psychology*, 87, 187–204. doi:10.1111/bjep.12143

Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology, 89*, 411–419. doi:10.1037/0022-0663.89.3.411

*Whitesell, N. R., Mitchell, C. M., & Spicer, P. (2009). A longitudinal study of self-esteem, cultural identity, and academic success among American Indian adolescents. Cultural Diversity and Ethnic Minority Psychology, 15(1), 38–50. doi:10.1037/a0013456

Whitesell, N. R., Mitchell, C. M., & Spicer, P. (2009). A longitudinal study of self-esteem, cultural identity, and academic success among American Indian adolescents. *Cultural Diversity and Ethnic Minority Psychology, 15*, 38–50. doi:10.1037/a0013456

*Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem: Change across elementary and middle school. *The Journal of Early Adolescence*, *14*(2), 107-138.

*Young, A. J. (1997). I think, therefore I'm motivated: The relations among cognitive strategy use, motivational orientation and classroom perceptions over time. *Learning and Individual Differences*, *9*, 249-283. doi:10.1016/s1041-6080(97)90009-1

*Zanobini, M., & Usai, M. C. (2002). Domain-specific Self-concept and Achievement Motivation in the Transition from Primary to Low Middle School. *Educational Psychology*, 22, 203–217. doi:10.1080/01443410120115265

*Zeiders, K. H., Umaña-Taylor, A. J., & Derlan, C. L. (2013). Trajectories of depressive symptoms and self-esteem in Latino youths: Examining the role of gender and perceived discrimination. *Developmental Psychology*, 49, 951–963. doi:10.1037/a0028866

*Zimmerman, M. A., Copeland, L. A., Shope, J. T., & Dielman, T. E. (1997). A longitudinal study of self-esteem: Implications for adolescent development. *Journal of youth and Adolescence*, *26*, 117-141

Appendix B

Table B1. *Cronbach's Alpha* α *and McDonald's Omega* ω *by Measurement Point as well as Item Wording of the Scales in Study 1*

Math Scales		T1	T2	T3	T4
Mastery-approach goals	*α*=	.718	.781	.810	.814
Item 1: "Mein Ziel in Mathematik ist es, den Stoff möglichst vollständig zu beherrschen." ("I desire to completely master the material presented in mathematics.")	ω=	.751	.801	.822	.833
Item 2: "In Mathematik ist es wichtig für mich, den Stoff so gründlich wie möglich zu verstehen." ("It is important for me to understand mathematics as thoroughly as I can.")					
Item 3: " In Mathematik möchte ich soviel wie möglich lernen." ("I want to learn as much as possible in mathematics.")					
Performance-approach goals	*α*=	.890	.892	.887	.928
Item 1: "In Mathematik möchte ich im Vergleich zu den anderen Schülerinnen und Schülern gut abschneiden." ("It is important for me to do well in math compared to other students.")	ω=	.898	.892	.888	.929
Item 2: "Mein Ziel in Mathematik ist es, eine bessere Leistung als die anderen Schülerinnen und Schüler zu zeigen." ("It is important for me to do better in math than other students.")					
Item 3: "In Mathematik ist es wichtig für mich, besser als andere Schülerinnen und Schüler abzuschneiden." ("My goal in math is to get a better grade than most of the other students.")					
Performance-avoidance goals	*α*=	.793	.819	.825	.891
Item 1: "In Mathematik ist es wichtig für mich, zu vermeiden, im Vergleich zu anderen Schülerinnen und Schülern schlecht abzuschneiden." ("It is important for me in math to avoid doing poorly compared to other students.")	ω=	.800	.826	.823	.891
Item 2: "In Mathematik ist es mein Ziel, keine schlechtere Leistung als andere Schülerinnen und Schüler zu zeigen." ("It is important for me in math to avoid performing worse than other students.")					
Item 3: "In Mathematik möchte ich es vermeiden, im Vergleich zu anderen Schülerinnen und Schülern eine schlechte Leistung zu zeigen." („My goal in math is to avoid a worse grade than most of the other students.")					
Interest	*α*=	.895	.891	.881	.873
Item 1: "In Mathe mache ich meine Hausaufgaben, weil mir dieses Fach Spaß macht." ("I do my math homework because it is fun.")	ω=	.896	.890	.879	.873
Item 2: "Im Mathe-Unterricht arbeite ich mit, weil ich großes Interesse an Mathematik habe." ("I work on math tasks because I enjoy it.")					
Item 3: "In Mathe strenge ich mich an, weil mich das Fach interessiert." ("I make effort in math because I am interested in this subject.")					

Table B2. *Missings on the Investigated Variables in Study 1*

	T1	T2	T3	T4
Mastery goals	140	84	134	214
PAP goals	140	83	132	213
PAV goals	140	83	133	213
Interest	140	83	132	213
ACH	154	151	237	286
Grades	-	381	451	180
Cognitive ability	133	-	-	-
Parent education	219	-	-	-
Language background	20	-	-	-
Missings on all variables	16	26	76	111

Note. PAP = performance-approach. PAV = performance-avoidance. ACH = math achievement test score. Grades = grades in math.

Table B3. *Statistics for Attrition Analysis in Study 1*

	Dropped out (group 1)			Joined later (group 2)			Participated at T1 and T4 (group 3)			*d* (1 v. 2)	*d* (1 v. 3)	*d* (2 v. 3)
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*			
M T1	89	4.58	.70	-	-	-	516	4.54	.57	-	.06	-
M T2	99	4.13	.76	1	2.67	-	561	4.27	.76	-	-.18	-
M T3	54	3.88	1.01	7	3.57	1.32	550	4.06	.81	-	-.20	-
M T4	-	-	-	11	3.83	0.87	520	4.08	.86	-	-	-.28
PAP T1	89	3.60	1.23	-	-	-	516	3.54	1.13	-	.06	-
PAP T2	99	3.45	1.03	1	3.00	-	562	3.29	1.13	-	.15	-
PAP T3	54	3.35	.94	7	2.81	1.20	552	3.17	1.11	-	.17	-
PAP T4	-	-	-	11	3.09	1.45	521	3.31	1.12	-	-	-.17
PAV T1	89	3.96	1.03	-	-	-	516	3.97	1.00	-	-.01	-
PAV T2	99	3.64	1.01	1	2.67	-	562	3.58	1.06	-	.05	-
PAV T3	54	3.52	.84	7	2.90	1.41	551	3.40	1.04	-	.13	-
PAV T4	-	-	-	11	3.30	1.35	521	3.43	1.09	-	-	-.10
Int. T1	89	4.12	1.01	-	-	-	516	4.04	0.98	-	.08	-
Int. T2	99	3.56	1.13	1	3.00	-	562	3.65	1.07	-	-.08	-
Int. T3	54	3.36	.99	7	3.05	1.54	552	3.41	1.08	-	-.05	-
Int. T4	-	-	-	11	3.27	1.25	521	3.45	1.03	-	-	-.15
ACH T1	90	-1.15	.84	-	-	-	501	-.81	.84	-	-.40	-
ACH T2	80	-0.81	1.01	-	-	-	514	-.51	.96	-	-.31	-
ACH T3	27	-0.73	1.22	3	1.08	.32	478	-.21	1.10	-	-.45	-
ACH T4	-	-	-	4	.40	.49	455	.06	.85	-	-	-
Grades T2	20	3.98	.95	3	5.00	1.00	341	4.73	.71	-	-.91	-
Grades T3	16	3.41	.55	5	4.60	1.08	273	4.36	.91	-	-1.30	-
Grades T4	-	-	-	10	4.33	1.28	555	4.05	.92	-	-	.25
Cog. a.	87	103.53	10.65	-	-	-	525	107.82	9.64	-	-.42	-
Sex	111	.41	.49	14	.50	.52	620	.43	.50	-.17	-.03	.14
P. edu.	73	4.15	1.14	-	-	-	453	4.34	1.14	-	-.17	-
Language b.	109	1.17	.37	-	-	-	616	1.12	.33	-	.12	-

Note. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. INT = Interest in math. ACH = math achievement test score. Grades = grades in math. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. Sex: 0 = boys, 1 = girls. *d* = Cohen's *d*. Cohen's *d* was not calculated if one group contained less than 10 students.

Table B4. *Model Fit of the Confirmatory Factor-analyses of the Achievement Goal Scales in Study 1 by Measurement Point*

Model	χ^2	df	SCF	*p*	CFI	RMSEA
T1: One-factor	370.312	27	1.518	< .001	.799	.143
T1: Two-factor	168.169	26	1.194	< .001	.917	.094
T1: Three-factor	67.009	24	1.353	< .001	.975	.054
T2: One-factor	489.569	27	1.401	< .001	.803	.161
T2: Two-factor	159.635	26	1.358	< .001	.943	.088
T2: Three-factor	80.279	24	1.353	< .001	.976	.060
T3: One-factor	358.008	27	1.650	< .001	.822	.141
T3: Two-factor	75.611	26	1.625	< .001	.973	.056
T3: Three-factor	48.364	24	1.601	= .002	.987	.041
T4: One-factor	331.950	27	1.678	< .001	.850	.146
T4: Two-factor	63.113	26	1.601	< .001	.982	.052
T4: Three-factor	35.173	24	1.642	= .066	.994	.030

Note. SCF = Scaling correction factor. One-factor = all nine achievement goals items are loading on one factor. Two-factor = the mastery goal items are loading on one factor while the PAP and PAV goal items are loading on a second factor. Three-factor = mastery, PAP, and PAV goals items are loading on 3 separate factors.

Table B5. *Results for the Tests of Measurement Invariance over Time for the Scales Assessing Mastery, PAP goals, PAV goals, and Math Interest in Study 1*

Model	χ^2	df	SCF	$\Delta \chi^2$	Δdf	Δ *p*	CFI	ΔCFI	RMSEA
Mastery goals									
Configural	32.100	30	1.099				.999		.010
Metric	40.548	36	1.130	8.206	6	= .223	.998	.001	.013
Scalar	54.514	42	1.128	14.063	6	= .029	.993	.005	.020
PAP goals									
Configural	32.646	30	1.260				.999		.011
Metric	56.111	36	1.250	24.134	6	< .001	.995	.004	.027
Scalar	118.959	42	1.205	78.159	6	< .001	.979	.016	.050
Partial scalar	83.006	41	1.206	33.722	5	< .001	.989	.006	.037
PAV goals									
Configural	29.490	30	1.126				1		0
Metric	52.413	36	1.107	24.559	6	< .001	.994	.006	.025
Scalar	63.922	42	1.102	11.580	6	= .072	.992	.002	.027
Interest									
Configural	60.652	30	1.118				.991		.037
Metric	65.163	36	1.106	4.299	6	= .636	.992	.001	.033
Scalar	80.494	42	1.090	15.766	6	= .015	.989	.003	.035

Note. SCF = Scaling correction factor. PAP = Performance-approach. PAV = Performance-avoidance. In the partial scalar model of performance-approach goals, we set free the intercept of the first item at T1.

Table B6. *Fit Indices of Latent Growth Curve Models in Study 1*

Model	χ^2	df	SCF	Δ χ^2	Δdf	Δ *p*	CFI	ΔCFI	RMSEA
Mastery									
Linear	286.523	5	1.041				.577		.276
Unspecified	11.616	3	.817	209.717	2	< .001	.987	.410	.062
Unspecified_m	24.734	4	1.123	379.375	1	< .001	.969	.392	.084
PAP									
Linear	81.175	5	1.066				.942		.143
Unspecified	43.522	3	.522	33.908	2	< .001	.969	.027	.135
PAV									
Linear	244.945	5	.985				.884		.254
Unspecified	26.247	3	.797	173.916	2	< .001	.989	.105	.102
Unspecified_m	28.483	4	.822	133.084	1	< .001	.988	.104	.091
Interest									
Linear	199.507	5	1.247				.766		.229
Unspecified	58.089	3	1.069	123.279	2	< .001	.934	.168	.157
ACH									
Linear	9.185	5	2.615				.986		.036
Unspecified	9.246	3	1.496	2.373	2	= .305	.979	.007	.057

Note. SCF = Scaling correction factor. PAP = Performance-approach goals. PAV = Performance-avoidance goals. ACH = math achievement test score. Linear = model only included the linear slope (to account for the different interval duration between measurement points, factor loadings of the linear slope factor were fixed to 0, 2, 5.5, and 8.5 for T1, T2, T3, and T4, respectively). Unspecified = model included a non-linear slope (Factor loadings of the slope factor were fixed to 0 and 8.5 for T1 and T4 respectively, whereas factor loadings for T2 and T3 were freely estimated). Unspecified_m = in this model, we fixed the residual variance of the T1 goals factor to zero as the T1 goals factor showed a slightly negative, non-significant, residual variance in the unspecified model of mastery goals and the model of PAV goals.

Table B7a. *Standardized Regression Coefficients of the Time Variant Control Variables from the Mastery Goals Correlated Change SEM in Study 1*

	Mastery T1	Mastery T2	Mastery T3	Mastery T4	Interest T1	Interest T2	Interest T3	Interest T4	ACH T1	ACH T2	ACH T3	ACH T4
On												
PAP T1	.14**				.21***				.00			
PAP T2		.18***				.30***				.04		
PAP T3			.14**				.31***				.03	
PAP T4				-.02				.28***				.11
PAV T1	.22***				.08				.03			
PAV T2		.24***				.08				.04		
PAV T3			.31***				.09				.04	
PAV T4				.42***				.11				-.01

Note. *$p < .05$. **$p < .01$. ***$p < .001$. PAP = Performance-approach goals. PAV = Performance-avoidance goals. ACH = math achievement test score.

Table B7b. *Standardized Regression Coefficients of the Time Variant Control Variables from the Performance-Approach Goals Correlated Change SEM in Study 1*

	PAP T1	PAP T2	PAP T3	PAP T4	Interest T1	Interest T2	Interest T3	Interest T4	ACH T1	ACH T2	ACH T3	ACH T4
On												
M T1	.08**				.39***				.00			
M T2		.09***				.47***				.00		
M T3			.02				.44***				.02	
M T4				.03				.41***				.08*
PAV T1	.67***				.09**				.02			
PAV T2		.69***				.13***				.08*		
PAV T3			.74***				.18***				.08	
PAV T4				.81***				.16***				.05

Note. *$p < .05$. **$p < .01$. ***$p < .001$. M = mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. ACH = math achievement test score.

Table B7c. *Standardized Regression Coefficients of the Time Variant Control Variables from the Performance-Avoidance Goals Correlated Change SEM in Study 1*

	PAV T1	PAV T2	PAV T3	PAV T4	Interest T1	Interest T2	Interest T3	Interest T4	ACH T1	ACH T2	ACH T3	ACH T4
On												
M T1	.15***				.38***				.00			
M T2		.11***				.45***				.01		
M T3			.16***				.42***				.03	
M T4				.14***				.39***				.07
PAPT1	.62***				.14***				.01			
PAPT2		.69***				.21***				.07*		
PAPT3			.70***				.25***				.07*	
PAPT4				.80***				.25***				.08*

Note. *p < .05. **p < .01. ***p < .001. M = mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. ACH = math achievement test score.

Table B8a. *Correlation Coefficients for the Control Variables form the Mastery Goals Correlated Change SEM in Study 1*

	PAP T1	PAP T2	PAP T3	PAP T4	PAV T1	PAV T2	PAV T3	PAV T4	Cog. a.	Sex	P. edu.
With											
PAP T2	.68*										
PAP T3	.52*	.65*									
PAP T4	.52*	.61*	.71*								
PAV T1	.72*	.51*	.42*	.37*							
PAV T2	.56*	.75*	.56*	.48*	.67*						
PAV T3	.48*	.57*	.79*	.61*	.55*	.71*					
PAV T4	.44*	.51*	.64*	.86*	.44*	.54*	.72*				
Cog. a.	.00	.01	.08*	.05	.00	.07	.06	.04			
Sex	-.14*	-.13*	-.11*	-.10*	-.15*	-.13*	-.12*	-.07	-.04		
P. edu.	-.08*	-.09*	-.04	-.04	-.04	-.08*	-.10*	-.07	.12*	.01	
Language	-.12*	-.10*	-.09*	-.10*	-.11*	-.10*	-.10*	-.07*	.13*	.07	.16*

Note. *$p < .05$. PAP = Performance-approach goals. PAV = Performance-avoidance goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B8b. *Correlation Coefficients for the Control Variables form the Performance-approach Goals Correlated Change SEM in Study 1*

	M T1	M T2	M T3	M T4	PAV T1	PAV T2	PAV T3	PAV T4	Cog. a.	Sex	P. edu.
With											
M T2	.65*										
M T3	.38*	.60*									
M T4	.35*	.50*	.63*								
PAV T1	.32*	.26*	.24*	.23*							
PAV T2	.26*	.40*	.33*	.26*	.67*						
PAV T3	.20*	.31*	.43*	.33*	.55*	.71*					
PAV T4	.19*	.25*	.36*	.46*	.44*	.54*	.72*				
Cog. a.	.04	.07	.09	.09*	.00	.07	.07	.04			
Sex	-.02	.00	.03	.08	-.15*	-.13*	-.12*	-.07	-.04		
P. edu.	-.04	-.04	.00	-.04	-.04	-.08*	-.10*	-.07	.12*	.01	
Language	-.04	-.03	-.02	.00	-.11	-.10*	-.10*	-.07*	.13*	.07	.18*

Note. $*p < .05$. M = mastery goals. PAV = Performance-avoidance goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B8c. *Correlation Coefficients for the Control Variables form the Performance-avoidance Goals Correlated Change SEM in Study 1*

	M T1	M T2	M T3	M T4	PAP T1	PAP T2	PAP T3	PAP T4	Cog. a.	Sex	P. edu.
With											
M T2	.65*										
M T3	.38*	.60*									
M T4	.35*	.50*	.63*								
PAP T1	.29*	.25*	.20*	.18*							
PAP T2	.21*	.34*	.24*	.21*	.68*						
PAP T3	.14*	.23*	.37*	.27*	.52*	.65*					
PAP T4	.16*	.21*	.30*	.36*	.52*	.61*	.71*				
Cog. a.	.04	.07	.09	.09*	.00	.00	.08*	.06			
Sex	-.02	.00	.03	.08	-.14*	-.13*	-.11*	-.10*	-.04		
P. edu.	-.04	-.04	.01	-.04	-.08*	-.09*	-.04	-.04	.12*	.01	
Language	-.04	-.03	-.02	.00	-.12*	-.10*	-.09*	-.10*	.13*	.07	.17*

Note. *p < .05. M = mastery goals. PAP = Performance-approach goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B9. *Fit Indices of the Investigated Structural Equation Models in Study 1*

Model	χ^2	df	SCF	CFI	RMSEA
Mastery correlated change SEM	470.692	158	1.062	.924	.052
PAP correlated change SEM	359.893	158	1.032	.967	.041
PAV correlated change SEM	355.175	158	1.054	.968	.041
Goals and Interest REM	198.642	48	1.040	.974	.078
Goals and ACH REM	185.944	48	1.083	.972	.075
Goals and Grades REM	191.439	47	1.002	.969	.078

Note. SCF = Scaling correction factor. PAP = Performance-approach goals. PAV = Performance-avoidance goals. REM = reciprocal effect model. In the correlated change SEM of mastery goals, the residual variance of the T1 mastery goals factor score was fixed to zero as the T1 factor score showed a slightly negative, non-significant, residual variance. In the correlated change SEM of PAP goals, the residual variance of the T1 PAP goals factor score was fixed to zero as the T1 factor score showed a slightly negative, non-significant, residual variance. In the correlated change SEM of PAV goals, the residual variance of the T4 interest goals factor score was fixed to zero as the T4 factor score showed a slightly negative, non-significant, residual variance.

Table B10a. *Regression Coefficients of the REM Including the Achievement Goals and Math Interest in Study 1*

	T2 on T1			T3 on T2			T4 on T3		
	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*
Interest on									
Mastery	.11	.04	= .009	.05	.04	= .234	.02	.04	= .675
PAP	.09	.04	= .038	-.06	.07	= .336	-.03	.05	= .585
PAV	.01	.04	= .798	.13	.08	= .094	.08	.05	= .089
Interest	.48	.05	< .001	.56	.04	< .001	.64	.05	< .001
Cog. a.	.13	.04	= .002	.04	.03	= .203	.06	.03	= .049
Sex	-.03	.04	= .511	-.09	.04	= .035	-.01	.04	= .759
P. edu.	.02	.03	= .663	.04	.04	= .328	.00	.03	= .922
Language	.01	.04	= .790	.01	.03	= .891	.03	.03	= .281
Mastery on									
Mastery	.57	.04	< .001	.52	.04	< .001	.61	.04	< .001
PAP	.00	.05	= .965	-.08	.05	= .124	-.06	.06	= .278
PAV	.09	.05	= .078	.17	.06	= .002	.12	.04	= .009
Interest	.06	.04	= .135	.07	.03	= .051	-.01	.04	= .758
Cog. a.	.04	.03	= .189	.03	.04	= .394	.04	.04	= .337
Sex	.06	.05	= .269	.01	.04	= .693	.07	.04	= .114
P. edu.	-.01	.05	= .786	.02	.04	= .590	-.04	.04	= .292
Language	.00	.03	= .947	-.02	.04	= .643	.05	.03	= .126
PAP on									
Mastery	-.07	.03	= .040	-.02	.05	= .731	-.02	.04	= .697
PAP	.57	.06	< .001	.48	.07	< .001	.58	.05	< .001
PAV	.07	.06	= .246	.20	.07	= .002	.11	.07	= .101
Interest	.10	.04	= .013	.05	.04	= .270	.09	.05	= .062
Cog. a.	.02	.03	= .531	.06	.03	= .045	-.01	.04	= .878
Sex	-.02	.04	= .672	-.01	.03	= .703	.01	.03	= .735
P. edu.	-.04	.04	= .221	.01	.03	= .719	-.01	.02	= .816
Language	-.06	.04	= .165	.02	.04	= .647	-.06	.03	= .043
PAV on									
Mastery	.00	.04	= .915	-.01	.04	= .784	.05	.04	= .237
PAP	.12	.05	= .011	.00	.05	= .956	.18	.06	= .003
PAV	.53	.05	< .001	.69	.06	< .001	.52	.06	< .001
Interest	.05	.03	= .116	.07	.03	= .022	.06	.04	= .099
Cog. a.	.10	.03	< .001	.02	.03	= .396	-.03	.04	= .481
Sex	-.01	.04	= .796	-.03	.03	= .264	.03	.04	= .370
P. edu.	-.04	.04	= .295	-.05	.03	= .057	.00	.03	= .905
Language	-.05	.03	= .037	-.05	.04	= .181	-.01	.02	= .681

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. Each control variable (cognitive ability, sex, parents' highest educational degree, and language background) was gathered at T1.

Table B10b. *Regression Coefficients of the REM Including the Achievement Goals and Math Achievement Test Score in Study 1*

	T2 on T1			T3 on T2			T4 on T3		
	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*
ACH on									
Mastery	.04	.03	= .209	.07	.05	= .161	.07	.04	= .102
PAP	.02	.05	= .729	.00	.05	= .969	.08	.06	= .173
PAV	-.04	.06	= .490	.03	.07	= .660	-.12	.07	= .112
ACH	.38	.05	< .001	.36	.06	< .001	.49	.06	< .001
Cog. a.	.23	.05	< .001	.26	.05	< .001	.22	.06	< .001
Sex	-.11	.06	= .049	-.07	.06	= .186	-.06	.04	= .156
P. edu.	.03	.04	= .498	.02	.04	= .578	.04	.04	= .333
Language	.01	.04	= .714	.03	.04	= .506	.00	.04	= .958
Mastery on									
Mastery	.59	.04	< .001	.55	.03	< .001	.60	.03	< .001
PAP	.01	.05	= .824	-.06	.05	= .207	-.07	.05	= .207
PAV	.08	.05	= .091	.17	.06	= .002	.12	.05	= .010
ACH	.11	.05	= .023	-.02	.05	= .668	.02	.04	= .607
Cog. a.	-.01	.05	= .839	.05	.05	= .288	.03	.05	= .521
Sex	.06	.06	= .261	.00	.04	= .979	.07	.04	= .086
P. edu.	-.01	.04	= .276	.03	.04	= .550	-.04	.04	= .284
Language	-.01	.03	= .847	-.01	.04	= .715	.05	.03	= .140
PAP on									
Mastery	-.02	.02	= .465	.01	.03	= .779	.02	.03	= .480
PAP	.58	.06	< .001	.49	.05	< .001	.59	.05	< .001
PAV	.07	.06	= .285	.20	.07	= .002	.11	.07	= .104
ACH	.06	.05	= .200	-.04	.03	= .252	.02	.04	= .608
Cog. a.	-.01	.04	= .846	.08	.03	= .012	-.01	.04	= .823
Sex	-.03	.05	= .504	-.03	.03	= .443	.00	.03	= .989
P. edu.	-.04	.03	= .247	.01	.03	= .660	.00	.03	= .934
Language	-.06	.04	= .160	.02	.04	= .566	-.05	.03	= .052
PAV on									
Mastery	.03	.03	= .440	.03	.03	= .373	.08	.04	= .030
PAP	.13	.05	= .006	.02	.05	= .777	.19	.06	< .001
PAV	.53	.05	< .001	.69	.06	< .001	.53	.06	< .001
ACH	.05	.05	= .270	-.05	.03	= .109	.03	.04	= .516
Cog. a.	.07	.03	= .031	.05	.03	= .077	-.03	.04	= .411
Sex	-.01	.05	= .757	-.05	.03	= .090	.03	.04	= .488
P. edu.	-.04	.04	= .330	-.05	.03	= .059	.00	.03	= .965
Language	-.06	.03	= .028	-.04	.03	= .206	-.01	.02	= .709

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. Each control variable (cognitive ability, sex, parents' highest educational degree, and language background) was gathered at T1.

Table B10c. *Regression Coefficients of the REM Including the Achievement Goals and Math Grades in Study 1*

	T2 on T1			T3 on T2			T4 on T3		
	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*
Grades on									
Mastery	-	-	-	.12	.04	= .005	.03	.03	= .378
PAP	-	-	-	-.02	.07	= .814	.05	.06	= .434
PAV	-	-	-	-.06	.08	= .453	-.02	.06	= .708
Grades	-	-	-	.68	.05	< .001	.62	.06	< .001
Cog. a.	-	-	-	.16	.06	= .011	.16	.04	< .001
Sex	-	-	-	.06	.05	= .225	-.01	.04	= .880
P. edu.	-	-	-	.08	.05	= .079	-.02	.04	= .610
Language	-	-	-	.03	.05	= .556	.02	.05	= .744
Mastery on									
Mastery	.59	.04	< .001	.55	.03	< .001	.59	.03	< .001
PAP	.01	.05	= .924	-.07	.05	= .162	-.07	.05	= .217
PAV	.09	.05	= .081	.18	.06	= .002	.12	.05	= .019
Grades	-	-	-	.02	.04	= .640	.15	.06	= .014
Cog. a.	.02	.04	= .556	.03	.04	= .385	-.01	.05	= .870
Sex	.05	.06	= .343	.01	.04	= .869	.07	.04	= .067
P. edu.	-.02	.05	= .740	.02	.04	= .570	-.06	.04	= .106
Language	.00	.03	= .909	-.02	.04	= .688	.04	.03	= .164
PAP on									
Mastery	-.02	.02	= .531	.01	.04	= .784	.02	.03	= .601
PAP	.58	.06	< .001	.49	.06	< .001	.59	.05	< .001
PAV	.07	.06	= .265	.20	.07	= .002	.11	.07	= .116
Grades	-	-	-	-.02	.04	= .551	.10	.04	= .013
Cog. a.	.01	.04	= .756	.07	.04	= .036	-.03	.04	= .418
Sex	-.04	.05	= .416	-.02	.03	= .484	.00	.03	= .912
P. edu.	-.04	.03	= .194	.01	.03	= .642	-.01	.02	= .600
Language	-.06	.04	= .191	.02	.04	= .598	-.06	.03	= .026
PAV on									
Mastery	.03	.03	= .392	.02	.03	= .446	.07	.03	= .038
PAP	.13	.05	= .007	.01	.06	= .865	.19	.06	< .001
PAV	.53	.05	< .001	.69	.06	< .001	.52	.07	< .001
Grades	-	-	-	.01	.05	= .772	.09	.04	= .033
Cog. a.	.10	.03	= .002	.03	.03	= .402	-.05	.04	= .185
Sex	-.02	.05	= .612	-.04	.03	= .139	.02	.03	= .530
P. edu.	-.04	.04	= .302	-.05	.03	= .056	-.01	.03	= .666
Language	-.05	.03	= .044	-.05	.03	= .188	-.01	.02	= .576

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. Each control variable (cognitive ability, sex, parents' highest educational degree, and language background) was gathered at T1.

Table B11. *Cronbach's Alpha α and McDonald's Omega* ω *by Measurement Point as well as Item Wording of the Scales in Study 2*

Math scales		T1	T2	T3	T4
Mastery goals	*α*=	.766	.777	.774	.728
Item 1: "In Mathematik strenge ich mich an, weil ich in diesem Fach etwas können möchte." ("In mathematics, I strive toward mastery.")	ω=	.761	.791	.780	.726
Item 2: "In Mathe tue ich etwas, weil ich über Mathematik mehr wissen möchte." ("I work for mathematics because I want to learn more about it.")					
Item 3: "Wenn ich mich mit Mathe beschäftige, dann konzentriere ich mich voll auf das Lösen der Aufgaben." ("When I study math, I concentrate fully on solving the tasks.")					
Performance-approach goals	*α*=	.814	.794	.837	.809
Item 1: "In Mathe versuche ich gerne, besser zu sein als die anderen Schüler/Schülerinnen." ("It is important for me to do well in math class compared to others.")	ω=	.811	.748	.819	.809
Item 2: "Ich lerne für Mathe, weil ich zu den Besten gehören möchte." ("I study for math because I want to be one of the best in the class.")					
Item 3: "Ich tue etwas für Mathe, weil ich zeigen möchte, dass ich intelligenter bin als andere." ("I make an effort in math because I want to demonstrate that I am more intelligent than others.")					
Performance-avoidance goals	*α*=	.789	.769	.742	.720
Item 1: "In Mathematik strenge ich mich an, weil ich nicht versagen möchte." ("I make an effort in math because I don't want to fail.")	ω=	.792	.777	.848	.733
Item 2: "Ich lerne für Mathe, weil ich mich nicht vor anderen blamieren möchte." ("I study for math because I don't want to embarrass myself in front of others.")					
Item 3: "Ich tue etwas für Mathe, weil ich nicht zu den schlechten Schülern gehören möchte." ("I make an effort in math because I don't want to be one of the students who do badly in class.")					
Interest	*α*=	.823	.807	.817	.792
Item 1: "Für Mathe interessiere ich mich" ("I am interested in math.")	ω=	.832	.812	.747	.747
Item 2: "Der Matheunterricht macht mir bisher Spaß." ("Math lessons are fun.")					
Item 3: "Oft bin ich nach dem Unterricht schon neugierig auf die nächste Mathe-Stunde." ("After a math class, I am often curious about what we are going to do in the next lesson.")					

Table B12. *Missings on the Investigated Variables in Study 2*

	T1	T2	T3	T4
Mastery goals	406	400	433	450
PAP goals	406	396	428	441
PAV goals	406	399	430	447
Interest	406	397	427	441
Grades	-	410	435	460
Cognitive ability	476	-	-	-
Parent education	944	-	-	-
Language background	338	-	-	-
Missings on all variables	227	391	409	433

Note. PAP = performance-approach. PAV = performance-avoidance. ACH = math achievement test score. Grades = grades in math.

Table B13. *Statistics for Attrition Analysis in Study 2*

	Dropped out (group 1)			Joined later (group 2)			Participated at T1 and T4 (group 3)			*d* (1 v. 2)	*d* (1 v. 3)	*d* (2 v. 3)
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*			
M T1	298	3.94	.90	-	-	-	716	3.94	0.87	-	-.01	-
M T2	307	3.68	.96	3	4.33	.33	710	3.81	0.89	-	-.14	-
M T3	276	3.53	.99	21	3.59	.83	690	3.60	0.95	-.06	-.07	-.01
M T4	-	-	-	150	3.13	.86	820	3.18	0.90	-	-	-.05
PAP T1	298	2.82	1.21	-	-	-	716	2.65	1.16	-	.14	-
PAP T2	308	2.60	1.12	3	2.44	.51	713	2.55	1.09	-	.05	-
PAP T3	278	2.61	1.16	21	2.81	1.23	693	2.50	1.15	-.17	.10	.26
PAP T4	-	-	-	153	2.34	1.05	826	2.26	0.99	-	-	.08
PAV T1	298	3.47	1.10	-	-	-	716	3.33	1.12	-	.12	-
PAV T2	308	3.28	1.07	3	3.33	.88	710	3.25	1.08	-	.02	-
PAV T3	278	3.09	1.07	21	2.79	1.06	691	3.01	1.04	.27	.07	-.21
PAV T4	-	-	-	151	2.71	.97	822	2.83	0.97	-	-	-.12
Int. T1	298	3.61	1.02	-	-	-	716	3.60	1.06	-	.00	-
Int. T2	308	3.21	1.05	3	3.11	.77	712	3.35	1.04	-	-.14	-
Int. T3	279	3.14	1.06	21	3.25	1.16	693	3.20	1.05	-.10	-.06	.05
Int. T4	-	-	-	153	2.66	1.03	826	2.71	0.99	-	-	-.05
Grades T2	300	4.61	.93	3	5.00	.00	707	4.95	0.66	-	-.43	-
Grades T3	270	4.03	1.01	21	4.29	1.01	694	4.35	0.91	-.25	-.33	-.07
Grades T4	-	-	-	141	4.18	1.29	819	4.40	1.13	-	-	-.18
Cog. a.	267	105.04	10.94	-	-	-	677	109.05	11.07	-	-.36	-
Sex	422	.39	.49	149	0.46	.50	834	.52	0.50	-.14	-.26	-.12
P. edu.	52	4.62	.97	-	-	-	424	4.44	1.10	-	.17	-
Language b.	319	.82	.39	-	-	-	763	.89	0.32	-	-.20	-

Note. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals. INT = Interest in math. Grades = grades in math. Cog. a. = cognitive ability score. P. edu. = parents' highest educational degree. Language b. = language background. Sex: 0 = boys, 1 = girls. *d* = Cohen's *d*. Cohen's *d* was not calculated if one group contained less than 10 students.

Table B14. *Model Fit of the Confirmatory Factor-Analyses of the Achievement Goal Scales in Study 2 by Measurement Point*

Model	χ^2	df	SCF	*p*	CFI	RMSEA
T1: One-factor	673.502	27	1.360	< .001	.777	.154
T1: Two-factor	399.534	26	1.270	< .001	.871	.119
T1: Three-factor	251.425	24	1.175	< .001	.921	.097
T2: One-factor	777.079	27	1.258	< .001	.745	.165
T2: Two-factor	482.308	26	1.248	< .001	.845	.131
T2: Three-factor	272.590	24	1.147	< .001	.915	.101
T3: One-factor	679.350	27	1.312	< .001	.753	.156
T3: Two-factor	307.828	26	1.294	< .001	.893	.104
T3: Three-factor	175.529	24	1.263	< .001	.943	.080
T4: One-factor	601.501	27	1.287	< .001	.734	.148
T4: Two-factor	385.581	26	1.328	< .001	.833	.119
T4: Three-factor	206.597	24	1.252	< .001	.915	.088

Note. SCF = Scaling correction factor. One-factor = all 9 achievement goal items are loading on one factor. Two-factor = the mastery goal items are loading on one factor while the PAP and PAV goal items are loading on a second factor. Three-factor = mastery, PAP, and PAV goal items are loading on 3 separate factors.

Table B15. *Results for the Tests of Measurement Invariance over Time for the Scales Assessing Mastery, PAP goals, PAV goals, and Math Interest in Study 2*

Model	χ^2	df	SCF	$\Delta\chi^2$	Δdf	Δ *p*	CFI	ΔCFI	RMSEA
Mastery goals									
Configural	26.135	30	1.188				1		0
Metric	34.617	36	1.193	34.617	6	=.209	1	0	0
Scalar	95.551	42	1.167	95.551	6	<.001	.983	.017	.030
Partial scalar	65.686	41	1.161	65.686	5	<.001	.992	.008	.021
PAP goals									
Configural	38.410	30	1.307				.998		.014
Metric	46.865	36	1.279	8.550	6	=.201	.997	.001	.015
Scalar	121.466	42	1.263	80.097	6	<.001	.981	.016	.037
Partial scalar	68.779	41	1.259	23.813	5	<.001	.993	.004	.022
PAV goals									
Configural	27.486	30	1.074				1		0
Metric	74.678	36	1.051	52.089	6	<.001	.988	.012	.028
Partial metric	50.385	35	1.060	24.476	5	<.001	.995	.005	.018
Scalar	113.047	41	1.068	60.319	6	<.001	.978	.017	.035
Partial scalar	84.805	39	1.0517	36.409	4	<.001	.986	.009	.029
Interest									
Configural	77.376	30	1.170				.988		.033
Metric	115.229	36	1.170	37.796	6	< .001	.980	.008	.039
Scalar	290.395	42	1.276	123.44	6	< .001	.938	.042	.065
Partial scalar	157.461	41	1.218	36.445	5	< .001	.071	.009	.045

Note. SCF = Scaling correction factor. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Mastery partial scalar = the intercept of item 1 was set free at T1. PAP partial scalar = the intercept of item 2 was set free at T4. PAV partial metric = the factor loading of item 2 was set free at T1. PAV partial scalar = the intercepts of item 1 and item 2 were set free at T4. Interest partial scalar = the intercept of item 3 at T2 was set free.

Table B16. *Fit Indices of the Latent Growth Curve Models in Study 2.*

Model	χ^2	df	SCF	$\Delta\chi^2$	Δdf	Δ *p*	CFI	ΔCFI	RMSEA
Mastery									
Linear	126.421	5	1.573				.926		.131
Unspecified	97.993	3	1.186	38.375	2	< .001	.942	.016	.150
PAP									
Linear	177.803	5	1.264				.941		.157
Unspecified	192.744	3	.970	22.159	2	< .001	.936	.005	.212
Unspecified_m	185.578	4	1.013	16.205	1	< .001	.938	.003	.180
PAV									
Linear	78.433	5	1.254				.968		.102
Unspecified	35.296	3	1.366	46.170	2	< .001	.986	.018	.087
Interest									
Linear	63.848	5	2.119				.962		.091
Unspecified	86.685	3	1.340	5.821	2	= .054	.947	.015	.141

Note. SCF = Scaling correction factor. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Linear = model only includes the linear slope. Quadratic = model includes a linear and a quadratic slope. Unspecified_m = in this model, we fixed the residual variance of the T4 PAP goals factor to zero as the T4 PAP goals factor showed a slightly negative residual variance in the unspecified model of PAP goals.

Table B17a. *Standardized Regression Coefficients of the Time Variant Control Variables from the Mastery Goals Correlated Change SEM in Study 2*

	Mastery T1	Mastery T2	Mastery T3	Mastery T4	Interest T1	Interest T2	Interest T3	Interest T4
On								
PAP T1	.21***				.24***			
PAP T2		.22***				.28***		
PAP T3			.25***				.28***	
PAP T4				.32***				.35***
PAV T1	.35***				.08**			
PAV T2		.33***				.07***		
PAV T3			.29***				.07**	
PAV T4				.25***				.01

Note. *$p < .05$. **$p < .01$. ***$p < .001$. PAP = Performance-approach goals. PAV = Performance-avoidance goals. P. edu. = parents' highest educational degree.

Table B17b. *Standardized Regression Coefficients of the Time Variant Control Variables from the Performance-Approach Goals Correlated Change SEM in Study 2*

	PAP T1	PAP T2	PAP T3	PAP T4	Interest T1	Interest T2	Interest T3	Interest T4
On								
M T1	.11***				.59***			
M T2		.16***				.66***		
M T3			.18***				.63***	
M T4				.25***				.64***
PAV T1	.39***				-.07*			
PAV T2		.39***				-.07**		
PAV T3			.40***				-.06**	
PAV T4				.37***				-.08**

Note. *p < .05. **p < .01. ***p < .001. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals.

Table B17c. *Standardized Regression Coefficients of the Time Variant Control Variables from the Performance-Avoidance Goals Correlated Change SEM in Study 2*

	PAV T1	PAV T2	PAV T3	PAV T4	Interest T1	Interest T2	Interest T3	Interest T4
On								
M T1	.25***				.53***			
M T2		.28***				.59***		
M T3			.23***				.58***	
M T4				.20***				.56***
PAP T1	.44***				.09**			
PAP T2		.40***				.11***		
PAP T3			.48***				.10***	
PAP T4				.42***				.12***

Note. *$p < .05$. **$p < .01$. ***$p < .001$. M = Mastery goals. PAP = Performance-approach goals. PAV = Performance-avoidance goals.

Table B18a. *Correlation Coefficients for the Control Variables form the Mastery Goals Correlated Change SEM in Study 2*

	PAP T1	PAP T2	PAP T3	PAP T4	PAV T1	PAV T2	PAV T3	PAV T4	Cog. a.	Sex	P. edu.
With											
PAP T2	.88*										
PAP T3	.73*	.86*									
PAP T4	.52*	.64*	.65*								
PAV T1	.59*	.53*	.45*	.27*							
PAV T2	.53*	.55*	.47*	.29*	.87*						
PAV T3	.47*	.51*	.58*	.36*	.73*	.73*					
PAV T4	.31*	.34*	.35*	.51*	.49*	.49*	.65*				
Cog. a.	-.06	-.04	-.04	.00	-.10*	-.11*	-.11*	-.08*			
Sex	-.25*	-.25*	-.21*	-.18*	-.08*	-.04	-.02	.01	.01		
P. edu.	-.12*	-.14*	-.14*	-.09*	-.08	-.08*	-.08	-.06	.00	-.06*	
Language	-.12*	-.10*	-.07*	-.04	-.06	-.04	-.03	-.01	.11*	-.04	-.04

Note. **p* < .05. PAP = Performance-approach goals. PAV = Performance-avoidance goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B18b. *Correlation Coefficients for the Control Variables form the Performance-Approach Goals Correlated Change SEM in Study 2*

	Mastery T1	Mastery T2	Mastery T3	Mastery T4	PAV T1	PAV T2	PAV T3	PAV T4	Cog. a.	Sex	P. edu.
With											
M T2	.80*										
M T3	.63*	.73*									
M T4	.47*	.56*	.61*								
PAV T1	.46*	.41*	.31*	.23*							
PAV T2	.43*	.48*	.33*	.24*	.87*						
PAV T3	.33*	.35*	.40*	.28*	.73*	.73*					
PAV T4	.22*	.22*	.23*	.38*	.49*	.49*	.65*				
Cog. a.	-.01	.00	.05	.06	-.10*	-.11*	-.10*	-.08			
Sex	-.11*	-.09*	-.06	-.11*	-.08*	-.04	-.02	.01	.01		
P. edu.	-.15*	-.14*	-.15*	-.08	-.13*	-.13*	-.15*	-.11*	.02	-.07	
Language	-.08*	-.05	.03	.03	-.06*	-.04	-.03	-.01	.11*	-.04	-.04

Note. **p* < .05. M = Mastery goals. PAV = Performance-avoidance goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B18c. *Correlation Coefficients for the Control Variables form the Performance-Avoidance Goals Correlated Change SEM in Study 2*

	Mastery T1	Mastery T2	Mastery T3	Mastery T4	PAP T1	PAP T2	PAP T3	PAP T4	Cog. a.	Sex	P. edu.
With											
M T2	.80*										
M T3	.63*	.73*									
M T4	.47*	.56*	.61*								
PAP T1	.41*	.38*	.28*	.26*							
PAP T2	.39*	.43*	.34*	.32*	.88*						
PAP T3	.33*	.38*	.43*	.34*	.73*	.86*					
PAP T4	.22*	.27*	.27*	.47*	.52*	.64*	.65*				
Cog. a.	-.02	.00	.05	.06	-.07*	-.05	-.04	.00			
Sex	-.11*	-.09*	-.06	-.11*	-.25*	-.25*	.21*	.18*	.01		
P. edu.	-.12	-.09	-.10	-.03	-.13*	-.15*	-.14*	-.09*	.00	-.05	
Language	-.07*	-.04	.03	.03	-.12*	-.10*	-.07*	-.04	.11*	-.04	-.04

Note. *$p < .05$. M = Mastery goals. PAP = Performance-approach goals. P. edu. = parents' highest educational degree. Language = language background (a German native speaker yes = 1; no = 0). Sex (0 = boys, 1 = girls).

Table B19. *Fit Indices of the Investigated Structural Equation Models in Study 2*

Model	χ^2	df	SCF	CFI	RMSEA
Mastery correlated change SEM	869.291	96	1.205	.910	.076
PAP correlated change SEM	696.925	98	1.224	.942	.066
PAV correlated change SEM	411.966	96	1.242	.967	.048
Goals and Interest REM	346.086	48	1.278	.981	.066
Goals and Grades REM	349.700	43	1.134	.976	.071

Note. SCF = Scaling correction factor. PAP = Performance-approach goals. PAV = Performance-avoidance goals. REM = reciprocal effect model.

Table B20a. *Regression Coefficients of the Cross-lagged Structural Equation Models Including the Achievement Goals and Math Interest in Study 2*

	T2 on T1			T3 on T2			T4 on T3		
	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*
Interest on									
Mastery	.06	.02	= .005	.06	.03	= .084	.02	.03	= .601
PAP	.05	.02	= .005	.03	.03	= .316	.06	.03	= .040
PAV	-.03	.02	= .154	-.04	.02	= .062	-.03	.03	= .377
Interest	.80	.02	< .001	.71	.03	< .001	.64	.03	< .001
Cog. a.	.03	.02	= .111	.03	.02	= .252	.03	.02	= .156
Sex	.02	.01	= .179	-.02	.02	= .441	-.05	.03	= .051
P. edu.	-.01	.02	= .763	-.03	.04	= .453	.06	.04	= .147
Language	.02	.01	= .101	.05	.02	= .026	-.02	.02	= .298
Mastery on									
Mastery	.71	.03	< .001	.65	.03	< .001	.49	.04	< .001
PAP	.04	.02	= .028	.05	.03	= .055	.07	.03	= .019
PAV	.05	.02	= .041	-.03	.03	= .308	.02	.03	= .468
Interest	.09	.03	< .001	.10	.04	= .003	.13	.03	< .001
Cog. a.	.01	.02	= .514	.03	.03	= .317	.02	.03	= .618
Sex	.02	.02	= .379	.01	.02	= .566	-.04	.03	= .121
P. edu.	.04	.03	= .225	-.06	.04	= .115	.05	.05	= .316
Language	.02	.02	= .263	.05	.02	= .047	.01	.03	= .634
PAP on									
Mastery	.02	.01	= .143	.01	.02	= .564	-.03	.03	= .346
PAP	.85	.01	< .001	.85	.02	< .001	.65	.03	< .001
PAV	.01	.02	= .589	-.01	.03	= .687	-.01	.03	= .843
Interest	.01	.02	= .582	.01	.02	= .494	.04	.04	= .291
Cog. a.	.01	.02	= .430	-.01	.02	= .606	.02	.02	= .243
Sex	-.03	.01	= .024	.00	.02	= .997	-.04	.02	= .070
P. edu.	-.02	.02	= .449	-.02	.03	= .417	.01	.03	= .838
Language	.00	.01	= .848	.02	.02	= .285	-.01	.03	= .859
PAV on									
Mastery	.02	.02	= .239	.00	.03	= .904	-.03	.03	= .433
PAP	.03	.02	= .131	.18	.03	< .001	-.02	.03	= .503
PAV	.84	.02	< .001	.65	.02	< .001	.67	.02	< .001
Interest	.01	.02	= .626	-.06	.03	= .039	-.01	.04	= .729
Cog. a.	-.03	.01	= .050	-.02	.02	= .367	-.01	.02	= .545
Sex	.04	.02	= .011	.04	.02	= .070	.02	.03	= .592
P. edu.	.04	.03	= .289	-.01	.05	= .759	-.02	.03	= .566
Language	.02	.02	= .157	.02	.02	= .364	-.01	.02	= .752

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. Each control variable (cognitive ability, sex, Parents' highest educational degree, and language background) was gathered at T1.

Table B20b. *Regression Coefficients of the Cross-lagged Structural Equation Models Including the Achievement Goals and Math Grades in Study 2*

	T2 on T1			T3 on T2			T4 on T3		
	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*	Std. Estimate	S.E.	*p*
Grades on									
Mastery	-	-	-	.14	.03	< .001	.10	.03	< .001
PAP	-	-	-	.03	.03	= .425	-.04	.04	= .317
PAV	-	-	-	-.05	.03	= .087	-.04	.04	= .213
Grades	-	-	-	.36	.04	< .001	.45	.04	< .001
Cog. a.	-	-	-	.30	.03	< .001	.12	.04	= .002
Sex	-	-	-	-.01	.03	= .865	.07	.03	= .012
P. edu.	-	-	-	.05	.05	= .287	.04	.05	= .367
Language	-	-	-	.04	.04	= .335	-.05	.04	= .180
Mastery on									
Mastery	.77	.02	< .001	.72	.02	< .001	.56	.03	< .001
PAP	.05	.02	= .008	.06	.03	= .014	.08	.03	= .007
PAV	.03	.02	= .126	-.04	.02	= .070	.01	.03	= .674
Grades	-	-	-	.00	.03	= .978	.07	.02	< .001
Cog. a.	.02	.02	= .355	.04	.03	= .189	.01	.03	= .866
Sex	.02	.02	= .474	.01	.02	= .591	-.05	.03	= .098
P. edu.	.04	.03	= .258	-.06	.04	= .111	.04	.05	= .440
Language	.02	.02	= .195	.05	.02	= .032	.01	.03	= .634
PAP on									
Mastery	.03	.02	= .100	.02	.02	= .142	-.01	.03	= .600
PAP	.85	.01	< .001	.86	.02	< .001	.65	.03	< .001
PAV	.01	.02	= .567	-.01	.03	= .570	.00	.03	= .904
Grades	-	-	-	-.01	.01	= .391	.05	.02	= .011
Cog. a.	.01	.01	= .453	-.01	.02	= .773	.01	.02	= .694
Sex	-.03	.01	= .020	.00	.02	= .977	-.04	.02	= .058
P. edu.	-.02	.02	= .503	-.02	.03	= .550	.00	.03	= .857
Language	.00	.01	= .880	.02	.02	= .223	-.01	.03	= .768
PAV on									
Mastery	.03	.02	= .074	-.03	.02	= .143	-.03	.03	= .249
PAP	.03	.02	= .099	.17	.03	< .001	-.02	.03	= .559
PAV	.84	.02	< .001	.65	.02	< .001	.66	.02	< .001
Grades	-	-	-	-.03	.02	= .131	-.04	.02	= .048
Cog. a.	-.03	.01	= .034	-.02	.02	= .509	.01	.03	= .678
Sex	.04	.02	= .011	.04	.02	= .073	.02	.03	= .586
P. edu.	.03	.03	= .335	-.01	.05	= .824	-.02	.03	= .597
Language	.02	.02	= .151	.02	.02	= .359	.01	.02	= .744

Note. PAP = Performance-approach goals. PAV = Performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. Each control variable (cognitive ability, sex, Parents' highest educational degree, and language background) was gathered at T1.

Systematic search for meta-analyses that investigated the relations of achievement goals and academic interest or academic achievement.

Search strategy 1: Database PsycINFO keyword search.
Search string: (goal(s) orientation **OR** achievement goal(s) **OR** performance goal(s) **OR** mastery goal(s) **OR** performance-avoidance **OR** performance-approach **AND** (meta-analysis) **in** (peer reviewed journals in English language) + 35 articles

Search strategy 2: Keyword search in relevant journals.
Search string a: (Title **OR** abstract search: goal(s) orientation **OR** achievement goal(s) **OR** performance goal(s) **OR** mastery goal(s) **OR** performance-avoidance **OR** performance-approach **AND** (meta-analysis) **in** (peer reviewed journals in English language) **in** (*Psychological Bulletin* **OR** *Educational Research Review* **OR** *Educational Psychology Review* **OR** *PLOS ONE* **OR** *Review of Educational Research* **or** *Personality and Social Psychology Review)* + 13 articles

Overlap between the search strategies - 4

= 44 articles

Inclusion and exclusion criteria of articles
(a) Not a meta-analysis - 5 articles
(b) No achievement goal measures in an academic setting or no relations with academic achievement and/or academic interest investigated - 22 articles
(c) Experimental manipulation of the achievement goals - 6
(d) No students from secondary or middle school were considered - 2

= 9 articles

Figure B1. Search process of the meta-analyses that are reported in Table 1.

Systematic search for meta-analyses that investigated the relations of the separate achievement goal constructs among each other and to academic achievement or academic interest.

Search strategy 1: Database PsycINFO keyword search. Search string: (*goal(s) orientation* **or** *achievement goal(s)* **OR** *performance goal(s)* **or** *mastery goal(s)* **OR** *performance-avoidance* **OR** *performance-approach*) **AND** (*longitudinal* **OR** *adolescent development* **OR** *childhood development* **OR** *secondary school* **OR** *middle school* **OR** *high school*) **AND** (*academic achievement* **or** *values* **or** *motivation* **or** *interest*) **in** (*peer reviewed journals in English language*). — + 588 articles

Search strategy 2: References from the Scherrer and Preckel (2018) meta-analysis — + 20 articles

Overlap between the search strategies — - 14 articles

= 594 articles

Inclusion and exclusion criteria of articles (several exclusion criteria per article possible)

- **(a)** No achievement goal measure related to school — in 99 articles
- **(b)** No dependent variable (i.e., measures of achievement and interest) — in 46 articles
- **(c)** No relation of achievement goals and the dependent variables are reported — in 30 articles
- **(d)** Experimental manipulation or not a longitudinal study — in 330 articles
- **(e)** The sample comprises no middle school or high school students — in 73 articles

↓ ↓ ↓

Articles that do not meet inclusion or exclusion criteria — - 578 articles

Articles that meet all inclusion and exclusion criteria — = 16 articles

Figure B2. Search process of the longitudinal articles that are reported in Table 2.

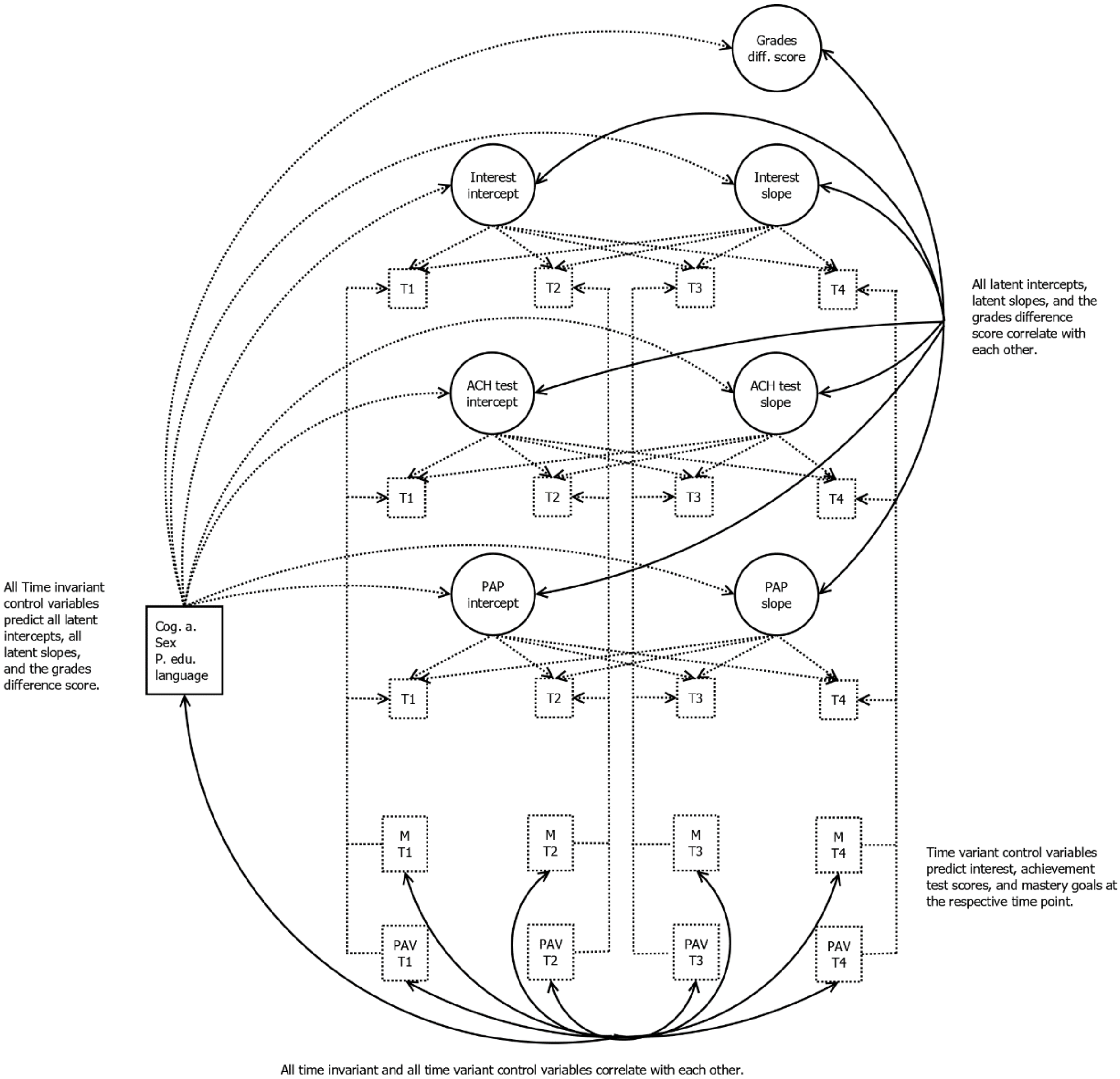


Figure B3. Correlated change of PAP goals, interest, achievement test scores, and grades controlled for students' cognitive ability, sex, parents' highest educational degree, and language background as time invariant predictors and for mastery and PAV goals as time variant predictors in Study 1. M = mastery goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. ACH = math achievement test score. Grades diff. score = latent change score for grades in math. Correlations are depicted as continuous lines. Regression paths are depicted as dotted lines.

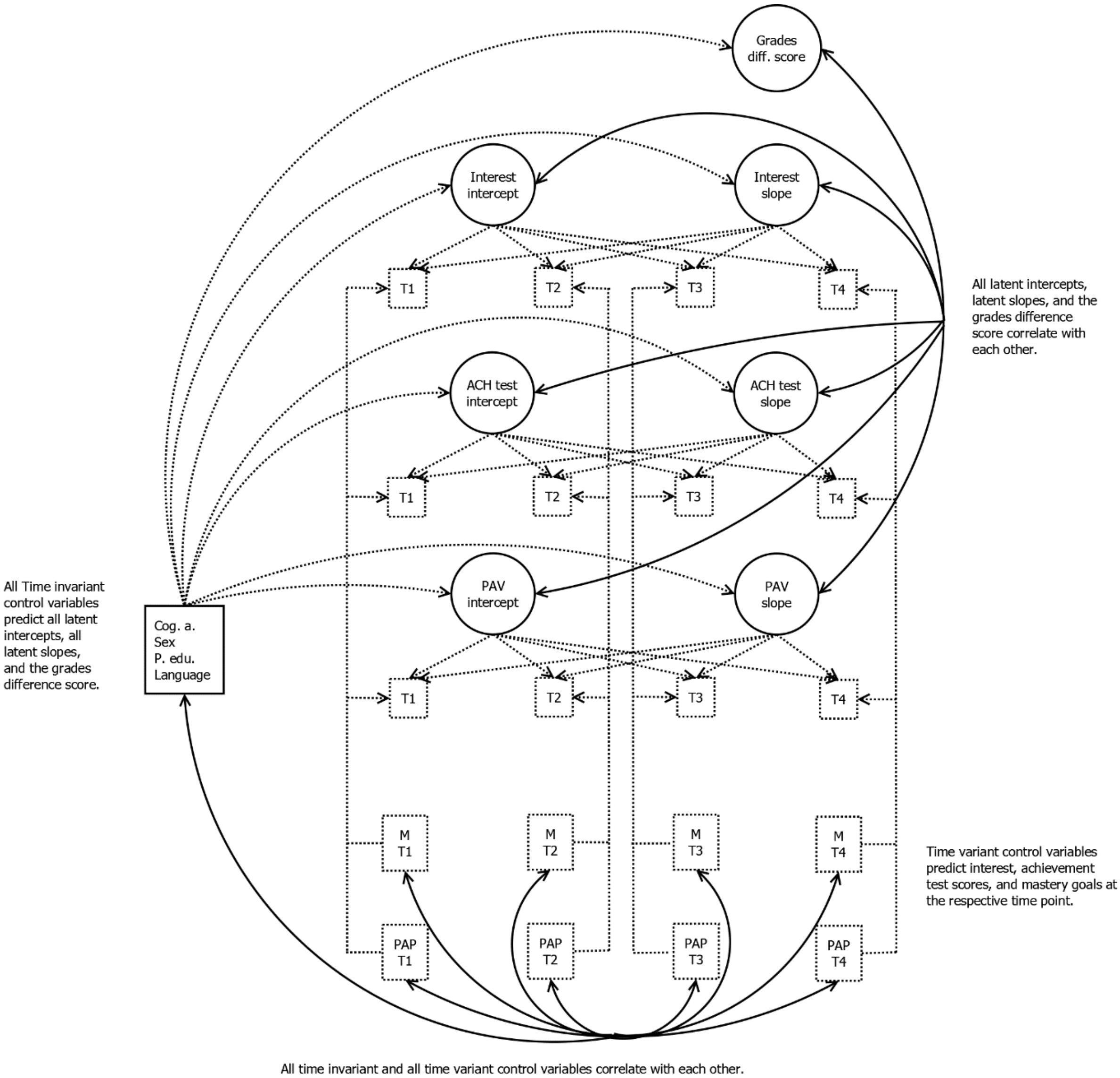


Figure B4. Correlated change of PAV goals, interest, achievement test scores, and grades controlling for students' cognitive ability, sex, parents' highest educational degree, and language background as time invariant predictors and for mastery and PAP goals as time variant predictors in Study 1. M = mastery goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree. Language = language background. ACH = math achievement test score. Grades diff. score = latent change score for grades in math. Correlations are depicted as continuous lines. Regression paths are depicted as dotted lines.

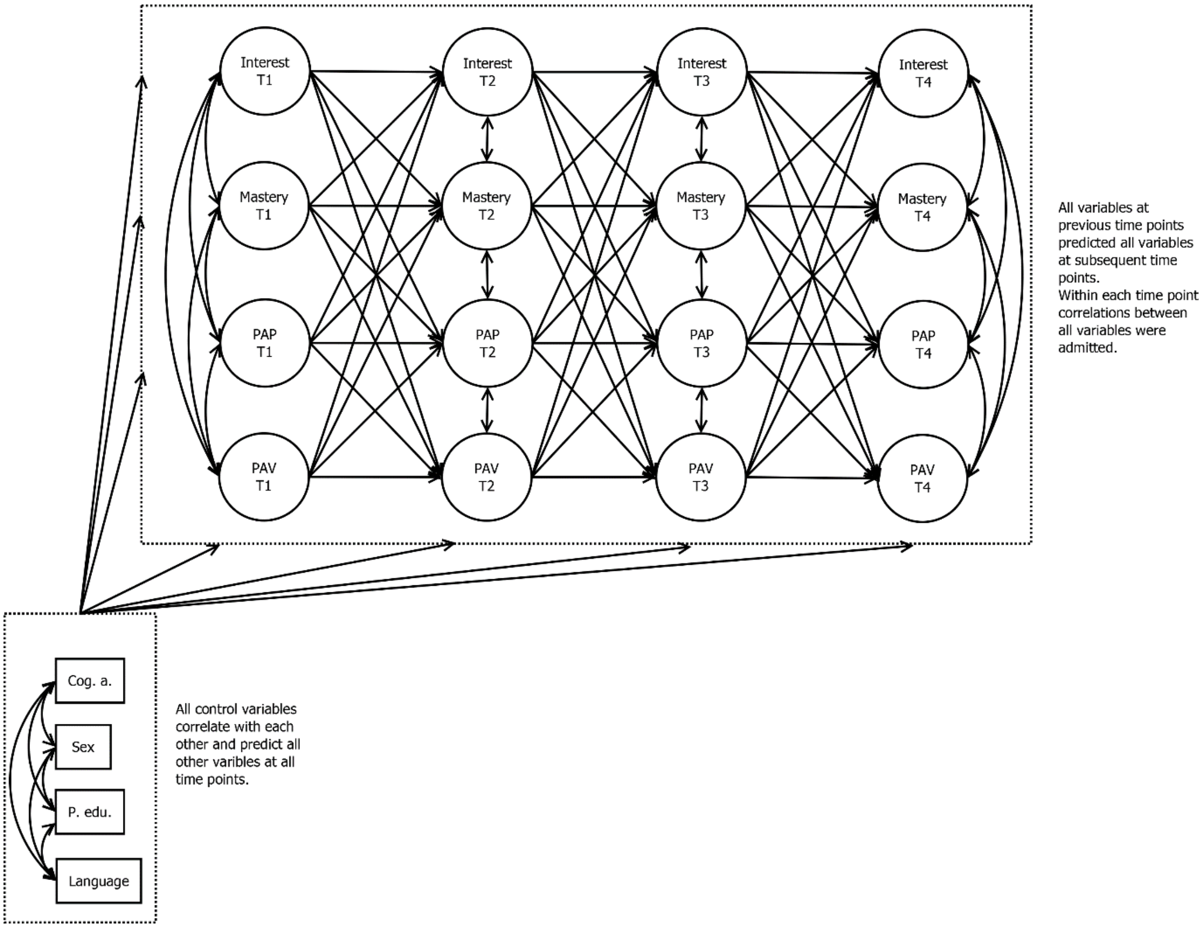


Figure B5. Reciprocal effect model of mastery, PAP, and PAV goals and interest controlling for cognitive ability, sex, and parents' highest educational degree. PAP = performance-approach goals. PAV = performance-avoidance goals. Cog. a. = cognitive ability. P. edu. = parents' highest educational degree.

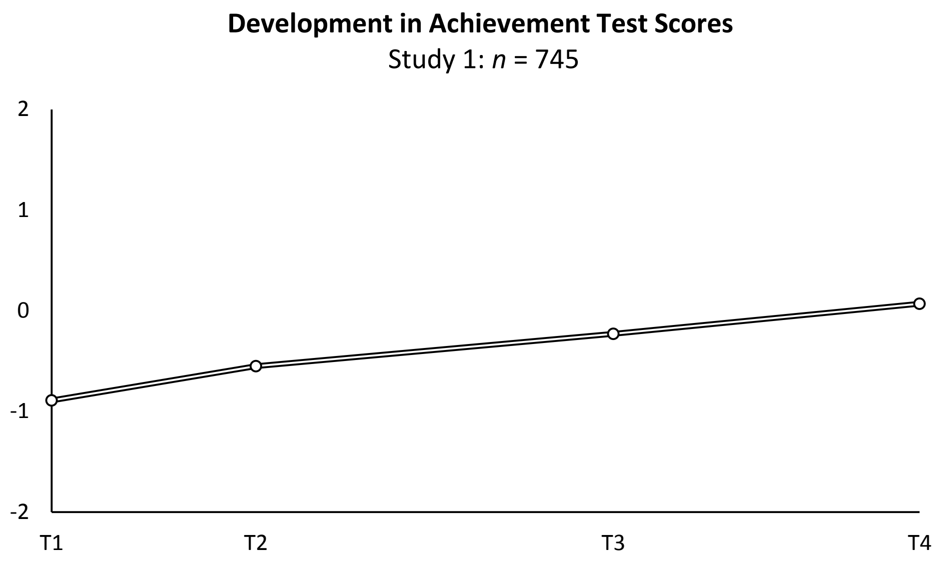
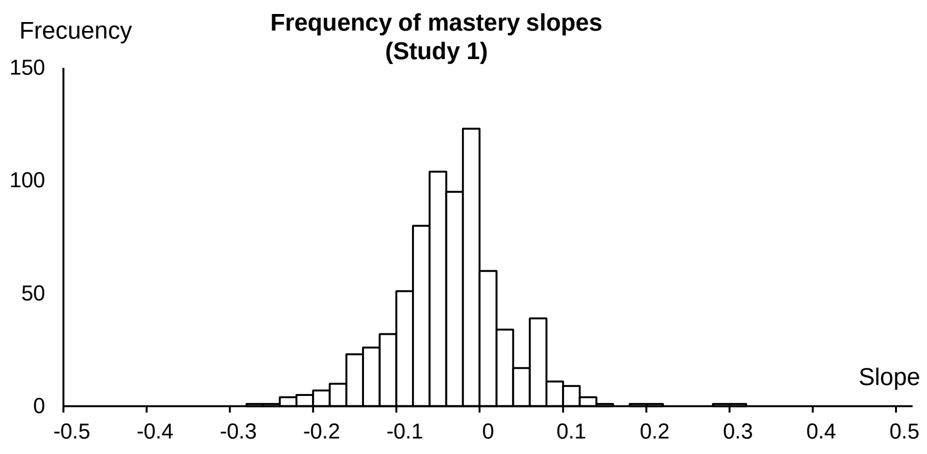
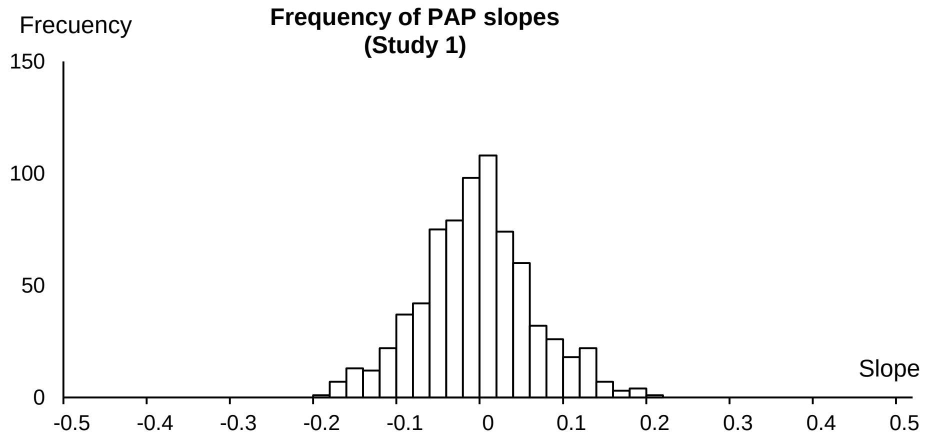
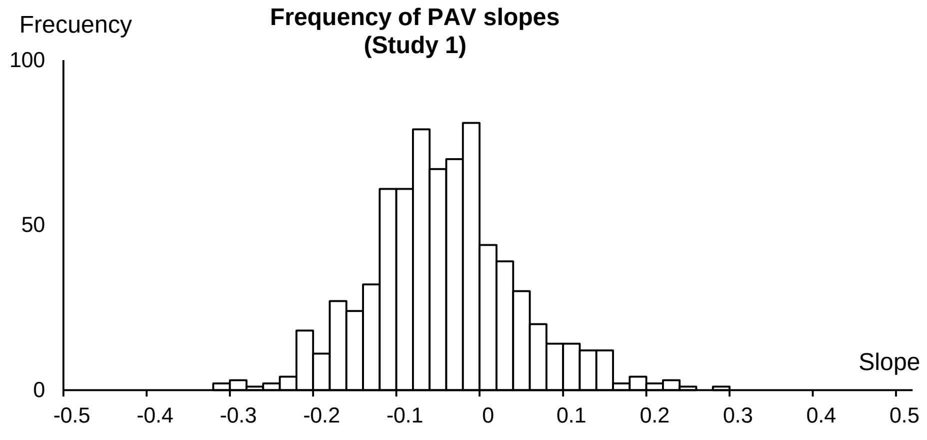
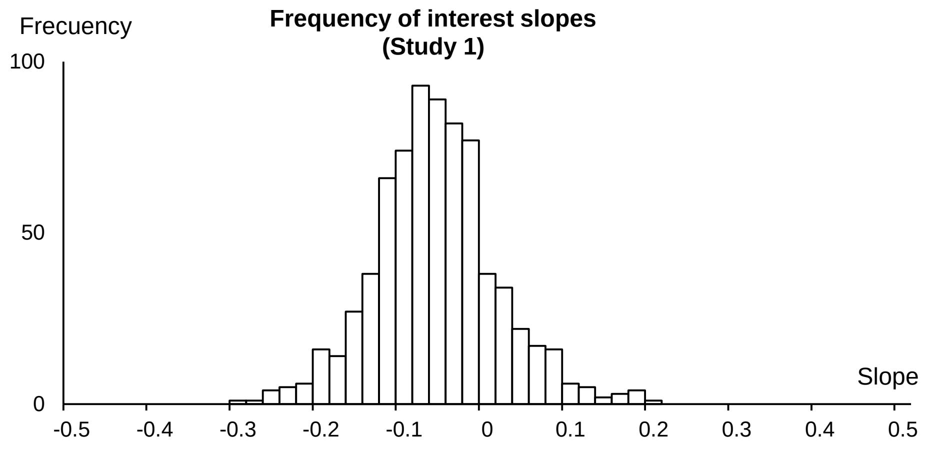


Figure B6. Developmental trajectory of achievement test scores based on person parameters estimated by weighted maximum likelihood estimation in Study 1.


Frecuency
Frequency of mastery slopes
(Study 1)
150
100
50
0
Slope
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5


Frecuency
Frequency of PAP slopes
(Study 1)
150
100
50
0
Slope
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5


Frecuency
Frequency of PAV slopes
(Study 1)
100
50
0
Slope
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5



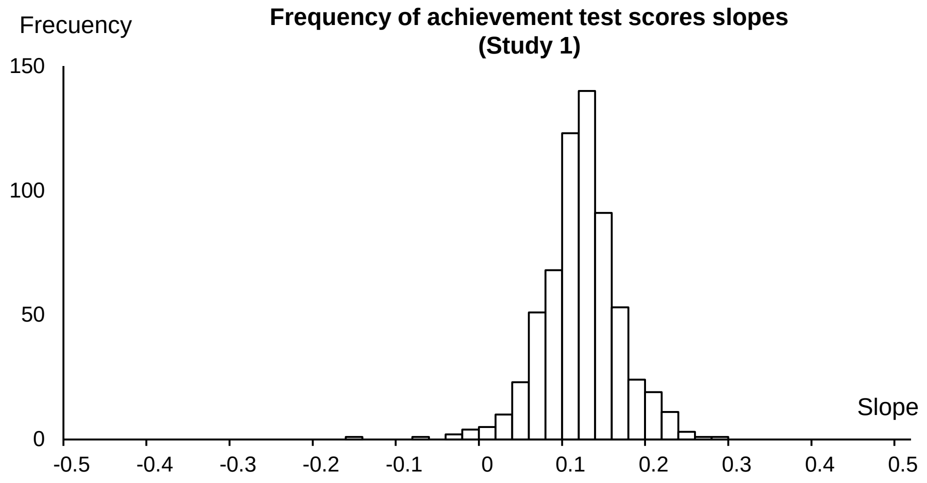
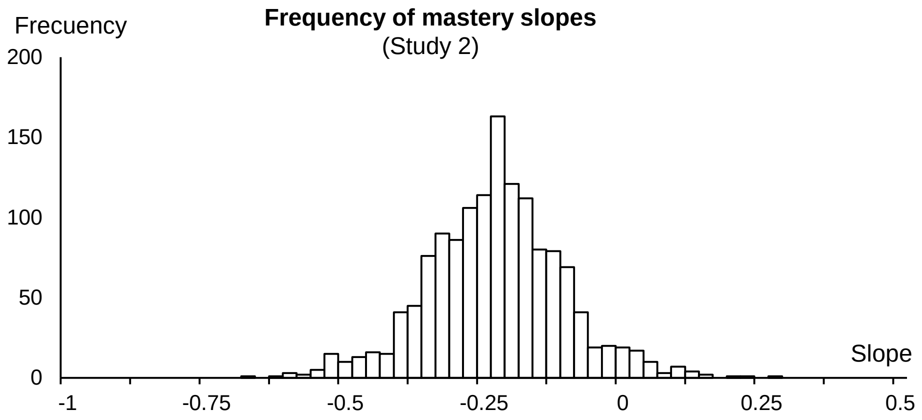
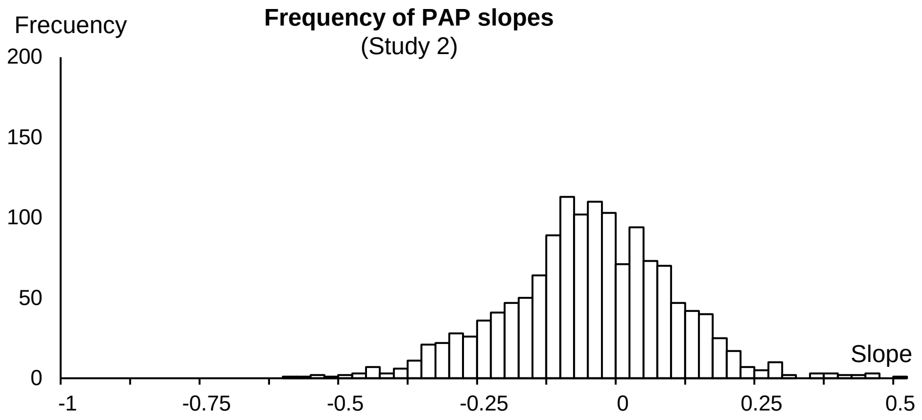
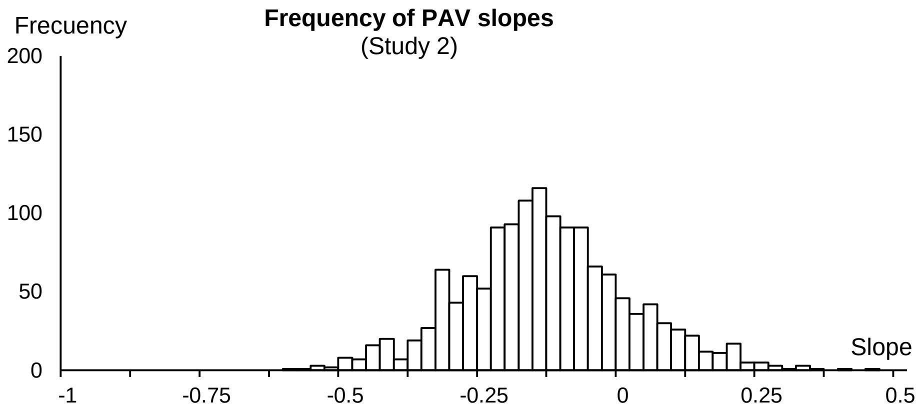


Figure B7. Frequencies of achievement goal, interest, and achievement test score slopes in Study 1. All slopes were taken from unconditional latent growth curve models. PAP = performance-approach goals. PAV = performance-avoidance goals.


Frecuency
Frequency of mastery slopes
(Study 2)
200
150
100
50
0
Slope
-1
-0.75
-0.5
-0.25
0
0.25
0.5


Frecuency
Frequency of PAP slopes
(Study 2)
200
150
100
50
0
Slope
-1
-0.75
-0.5
-0.25
0
0.25
0.5


Frecuency
Frequency of PAV slopes
(Study 2)
200
150
100
50
0
Slope
-1
-0.75
-0.5
-0.25
0
0.25
0.5

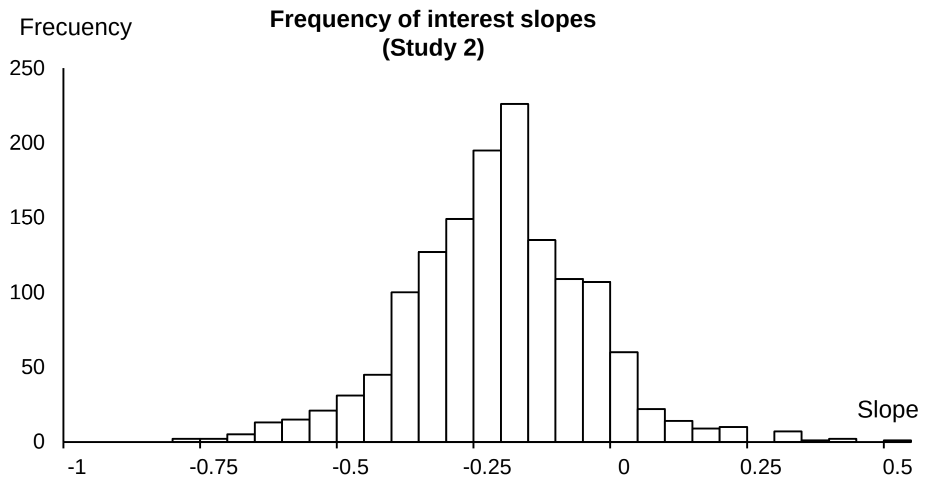


Figure B8. Frequencies of achievement goal and interest slopes in Study 2. Mastery and PAV slopes were taken from unconditional latent growth curve models. PAP goals and interest slopes were taken from linear latent growth curve models. PAP = performance-approach goals. PAV = performance-avoidance goals.

Appendix C

Propensity Score Matching

We estimated a propensity score (PS) for each student by boosted regression (McCaffrey, Ridgeway, & Morral, 2004) using the Twang Package by Ridgeway, McCaffrey, Morral, Burgette, and Griffin (2012) in the statistical program R. A PS is defined as “the conditional probability of assignment to a particular treatment given a vector of observed covariates” (Rosenbaum & Rubin, 1983, p. 41). In our study, the treatment comprised attending a gifted class.

Boosted regression is a data-adaptive and nonparametric approach that recursively partitions the data for each covariate. The boosted regression algorithm first partitions the dataset into two regions on the basis of one cut-off value of a single covariate, then divides each of the partitions into two further sections on the basis of another cut-off value of this covariate or on the basis of one further covariate and so on until the boosted regression includes the allowed number of partitions (McCaffrey, Ridgeway, & Morral, 2004). Each partition allows for additional interactions between variables and the algorithm initially selects partitions that provide the most information about the treatment assignment. An advantage of the boosted regression approach is that it uses the correct functional form for each covariate (i.e., categorical, ordinal, or continuous forms). Furthermore, it automatically takes into account interactions, three-way interactions, and missing values for all covariates included in the PS estimation (Stone & Tang, 2013).

As covariates, we used students’ IQ, age, gender, parents’ highest educational degree (either mother or father), school attended by the student, math achievement, reading speed, reading comprehension, general ASC, math ASC, verbal (German) ASC, and need for cognition (NFC) assessed at the first wave of measurement. The covariates were significantly related to the treatment variable:

Covariate	*n*	Relation with class type	Description of measure
IQ	887	*r* = .597**	Cognitive Ability Test for Grades 4 to 12 (KFT 4-12+R; Heller & Perleth, 2000) described in main text.
Students' age	907	*r* = -.354**	Age in months.
Math achievement	850	*r* = .468**	Standardized achievement test Version 1 (Authors, 2013a, 2013b) described in main text.
Reading speed	851	*r* = .168**	Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12 (LGVT 6-12) [Reading speed and comprehension test for grades 6 to 12]. Göttingen: Hogrefe.
Text comprehension	809	*r* = .348**	Souvignier, E., Trenk-Hinterberger, I., Adam-Schwebe, S. & Gold, A. (2008). Frankfurter Leseverständnistest (FLVT 5-6) [Frankfurter Reading comprehension test]. Göttingen: Hogrefe.
General ASC	864	*r* = .082*	Four items of the Self-Description Questionnaire (SDQ; Marsh, 1990b)
Math ASC	864	*r* = .179**	Four items of the Self-Description Questionnaire (SDQ; Marsh, 1990b) described in main text.
Verbal (German) ASC	864	*r* = .115**	Three items of the Self-Description Questionnaire (SDQ; Marsh, 1990b), German short version (SDQ II; Kunter et al., 2002). Items: "I learn German quickly," "I have always done well in German," and "German is one of my best subjects."
NFC	864	*r* = .216**	19-item NFC scale for adolescents (Preckel, F. (2014). Assessing need for cognition in early adolescence: Validation of a German adaption of the Cacioppo/Petty scale. *European Journal of Psychological Assessment*, *30*, 65-72.)
Gender	922	Chi^2=8.364**, *df* = 1	0 = boy, 1 = girl
Parents' education	778	*r* = .246**	Parents' highest educational degree on an ordinal scale: *no educational degree, certificate of secondary education, secondary school level I certificate, vocational/university entrance diploma, college/university degree, PhD*.

The allowed maximum number of splits for the boosted regression algorithm was set at 8000 and the optimal PS solution was found after 1250 partitions. The covariates had the following relative weights for the PS estimation: IQ = 56.575%, age = 19.588%, gender = 0.000%, parents' highest educational degree = 2.019%, particular school attended by the student = 6.975%, math achievement = 7.564%, reading speed = 5.099%, text comprehension = 0.590%, general ASC = 0.000%, math ASC = 0.084%, verbal (German) ASC = 0.844%, and NFC = 1.423%.

Using the PS, matching of students in gifted and regular classes was implemented with the MatchIt Package (Ho, Imai, King, & Stuart, 2011) in the statistical program R. Overall, we applied five different matching strategies: use of the nearest neighbor one-to-one procedure, use of caliper values, weighting of students in regular classes, use of the discard option for students in regular classes, and use of the discard option for students in gifted classes. Different combinations of the five matching strategies resulted in five different matching solutions or models (see Table C1). For Model 1, we used the nearest neighbor one-to-one procedure and matched the students without replacement, weighting, or a caliper rule. For Models 2 to 4 we applied nearest neighbor one-to-one procedure without replacement, used no weighting, and specified a caliper of 1.00 (Model 2), 1.25 (Model 3), and 1.50 (Model 4). The caliper value denotes the maximal value of the PS standard deviation within which two units are allowed to be matched together. For example, a caliper of 1.25 means that two matched units may differ by a maximum of 1.25 PS standard deviations. Units of the treatment group and the control group that exceed the caliper criteria are discarded through this procedure. Note that exclusion of treatment units (i.e., students in gifted classes) can lead to a systematic distortion of this group. For Model 5, we used a full matching algorithm with replacement, weighting of students in regular classes, and discard option of control units (i.e., students in regular classes). Thus, all gifted class students were maintained, whereas control group units that did not lie in the PS range of the treatment group were discarded. Furthermore, in order to align the two groups, each of the maintained control group units received a specific weight that was taken into account in further analyses.

Results for Model 1 to 5

Table C1 reports standardized mean differences in covariates between students in regular and gifted classes before and after the matching for all models (effect sizes d_1). Our decision criterion considered a covariate balanced if the standardized bias was $\leq$ 0.20. Although there are no strict cut-offs for choosing a specific criterion (Rubin, 2001), the 0.20 value provides an approximate criterion, in part based on results in Cochran and Rubin (1973), and the classification of effect sizes by Cohen (1988).

Table C1 further reports standardized mean differences in covariates between all students in gifted classes and those students in gifted classes selected by the matching procedure (effect sizes d_2). These values indicate whether the matched students from gifted classes were representative of all students in the gifted classes or whether a systematic distortion accrued due to the exclusion of some students.

Before the matching, there were large effect sizes for mean differences between students in regular classes and gifted classes regarding cognitive ability, math achievement, and reading comprehension (d_1 > .80), medium sized effect sizes for age and reading speed (d_1 = .50 to .80), and small to non-relevant effect sizes for the other covariates (d_1 < .50). Although in Model 1 the matching reduced these differences, a large effect size remained for cognitive ability (d_1 = .81), medium effect sizes remained for math achievement and reading comprehension (d_1 = .50 to .80), and small effect sizes were found for reading speed, verbal ASC, and NFC (d_1 = .20 to .50). These mean differences can be explained by the regular classes' lack of students with high PS (i.e., students with very high cognitive ability and high scores on the math and reading achievement tests).

In Models 2 to 4, we used a caliper rule for the matching, with the consequence that some gifted class students with high PS were discarded from the matching procedure. These models therefore provide a better fit than Model 1.

For Model 2 (caliper = 1), there were no relevant mean differences between students from both class types on any covariate (all $d_1 \leq .20$). However, the matching reduced the included number of students from the gifted classes from 283 to 131. These 131 students showed small sized differences when compared to all students in gifted classes with regard to cognitive ability (d_2 = .42), age (d_2 = .39), math achievement (d_2 = .33), and reading comprehension (d_2 = .26), and NFC (d_2 = .23).

Using a less conservative caliper rule (caliper = 1.25) in Model 3, including 148 students from gifted classes, also eliminated all relevant mean differences between class types on all covariates (all $d_1 \leq .2$). The 148 students showed small sized mean differences when compared to all students in gifted classes with regard to cognitive ability (d_2 = .37), age (d_2 = .35), math achievement (d_2 = .27), and reading comprehension (d_2 = .21). Of note, these d_2 values were smaller than those from Model 2.

In Model 4 (caliper = 1.5), 167 students from the gifted classes remained in the sample after matching. However, with this larger subset there were small sized mean differences between class types for cognitive ability (d_1 = .33), math performance (d_1 = .25), and reading comprehension (d_1 = .22). The 167 matched students from gifted classes showed small sized mean differences when compared to all students from the gifted classes regarding age (d_2 = .35), cognitive ability (d_2 = .27), and math achievement (d_2 = .21).

In Model 5, 195 students from the regular classes were matched to all 283 students in the gifted classes. Taking into account the calculated weights for the students in regular classes, we observed a medium sized mean difference between class-types in German ASC (d_1 = .51) and small sized mean differences in math achievement (d_1 = .35), reading comprehension (d_1 = .32), general ASC (d_1 = .37), and NFC (d_1 = .34). When inspecting the weights assigned to students in regular classes, the two students from regular classes with the highest weights accounted for 95 students from gifted classes.

Model Comparison and Rationale for Model Selection

In Model 1, no students from the gifted classes were discarded; however, there were meaningful differences on the covariates between class types. In Model 2, these mean differences were completely eliminated; however, 152 students from the gifted classes were discarded by the matching. Further, there were small sized mean differences between the selected students from the gifted classes and all students from these classes with regard to age, cognitive ability, math achievement, reading comprehension, and NFC. In Model 3, by discarding 135 students from the gifted classes, relevant mean differences between class types were also eliminated. Further, mean differences between the selected students from gifted classes and all students from these classes were smaller than those found in Model 2. In Model 4, there were small sized mean differences between class types and between selected students from gifted classes and all students from the gifted classes. In Model 5, no students from gifted classes were discarded. However, small to medium mean differences remained between class types and two students from regular classes received enormous relative weights.

Based on this model comparison, we chose Model 3 as the best matching solution. It offered the most balanced solution taking into account (a) the balancing of covariates and (b) the representativeness of the selected students from the gifted classes for all students in the gifted classes. There were no relevant mean differences between the matched students from regular and gifted classes on the covariates (all $d_1 \leq .2$) and only a few small mean differences ($d_1 = .2$ to .5) between the students from gifted classes selected by the matching and all students from gifted classes. Therefore, all further analyses were computed with the subsamples from Model 3.

However, to test the robustness of the findings from the different matching solutions, we additionally ran all analyses with the subsamples from Model 1, 2, and 4. Model 5 was excluded from further consideration because of the strongly overproportioned weights of just two of the students from the regular classes. Findings of robustness checks are reported in Appendix S2. Figure A1 at the end of this Appendix S1 shows the group means in math academic self-concept and math achievement across the four waves of measurement by class type for the whole (unmatched) data set and the matched data sets (Model 1 to 4).

Tables

Table C1

Description of the Unmatched Sample (Overall and by Class Type) and the Matched Samples (Model 1-5) on the Covariates Used for Matching with Effect Sizes for Mean Differences between Gifted Classes and Regular Classes (Cohens' d_1) for all Models and between Matched Students and all Students from Gifted Classes (Cohens' d_2) for Models 2 to 4

Matching strategy					**Model 1**		**Model 2**				**Model 3**				**Model 4**				**Model 5**	
Matching 1:1					yes		yes				yes				yes				no	
Caliper					no		1				1.25				1.5				no	
Weighting students in RC					no		no				no				no				yes	
Discard students from RC					yes		yes				yes				yes				yes	
Discard students from GC					no		yes				yes				yes				no	
	All	**RC**	**GC**		**RC**		**GC**	**RC**			**GC**	**RC**			**GC**	**RC**			**RC**	
Covariates	*N*=922	*n*=639	*n*=283	d_1	*n*=283	d_1 [a]	*n*=131	*n*=131	d_1	d_2	*n*=148	*n*=148	d_1	d_2	*n*=167	*n*=167	d_1	d_2	*n*=195	d_1 [a]
Cognitive ability *M (SD)*	111.98 (12.31)	107.21 (9.90)	122.61 (10.35)	1.49	114.20 (7.97)	.81	118.30 (8.69)	118.29 (7.49)	.00	.42	118.77 (8.68)	117.98 (6.75)	.12	.37	119.79 (8.99)	117.35 (7.30)	.33	.27	122.94 (6.12)	.03
Age in month *M (SD)*	126.34 (6.62)	127.90 (5.16)	122.83 (8.04)	.63	126.61 (5.69)	.47	125.93 (6.31)	126.39 (5.92)	.07	.39	125.64 (5.96)	125.94 (5.84)	.05	.35	125.65 (6.98)	126.16 (5.73)	.09	.35	122.89 (5.18)	.01
Math ACH *M (SD)*	-.57 (.96)	-.87 (.83)	.10 (.90)	1.08	-.48 (.80)	.65	-.19 (.86)	-.21 (.75)	.02	.33	-.16 (.85)	-.23 (.71)	.12	.27	-.09 (.85)	-.28 (.73)	.25	.21	-.21 (.61)	.35
Reading speed *M (SD)*	7.81 (5.43)	7.22 (5.91)	9.21 (3.71)	.54	8.25 (5.41)	.26	8.73 (3.59)	9.13 7.07	.11	.13	8.78 (3.78)	8.23 (3.28)	.17	.12	8.93 (3.86)	8.74 (6.54)	.03	.08	9.58 (2.64)	.09
Text com-prehension *M (SD)*	23.48 (6.01)	22.10 (5.95)	26.62 (4.85)	.93	24.03 (5.27)	.53	25.34 (4.72)	25.06 (5.18)	.06	.26	25.60 (4.63)	24.55 (5.37)	.20	.21	25.79 (4.90)	24.58 (5.39)	.22	.17	28.19 (4.08)	.32
General ASC *M (SD)*	4.14 (.65)	4.11 (.66)	4.23 (.62)	.19	4.17 (.63)	.09	4.18 (.64)	4.22 (.57)	.06	.08	4.19 (.65)	4.24 (.58)	.09	.06	4.18 (.64)	4.20 (.58)	.04	.08	4.45 (.68)	.37
Math ASC *M (SD)*	4.09 (.86)	3.98 (.89)	4.32 (.75)	.45	4.18 (.80)	.19	4.19 (.80)	4.17 (.81)	.02	.18	4.24 (.82)	4.24 (.75)	.01	.11	4.24 (.79)	4.20 (.80)	.05	.11	4.52 (.68)	.26

German ASC *M (SD)*	4.09 (.73)	4.03 (.74)	4.22 (.70)	.26	4.07 (.73)	.21	4.18 (.69)	4.13 (.70)	.07	.05	4.19 (.71)	4.16 (.63)	.05	.04	4.17 (.69)	4.09 (.70)	.11	.07	4.57 (.55)	.51
Need for cognition *M (SD)*	3.52 (.68)	3.42 (.66)	3.74 (.67)	.48	3.53 (.68)	.31	3.58 (.71)	3.54 (.65)	.06	.23	3.64 (.68)	3.59 (.61)	.10	.14	3.62 (.70)	3.55 (.63)	.12	.17	3.97 (.75)	.34
Gender																				
male *n*	541	355	186		168		77	75			89	85			96	96			53	
female *n*	381	284	97		115		54	56			59	63			71	71			142	
Parents' highest educational degree (degree: *n* parents)	1: 2 2: 27 3: 155 4: 119 5: 343 6: 132	1: 2 2: 26 3: 123 4: 98 5: 207 6: 70	1: 0 2: 1 3: 32 4: 21 5: 136 6: 62		1: 1 2: 7 3: 37 4: 39 5: 116 6: 45		1: 0 2: 1 3: 20 4: 13 5: 52 6: 31	1: 1 2: 2 3: 15 4: 16 5: 55 6: 25			1: 0 2: 1 3: 22 4: 15 5: 59 6: 36	1: 1 2: 2 3: 19 4: 16 5: 56 6: 32			1: 0 2: 1 3: 21 4: 18 5: 68 6: 42	1: 1 2: 5 3: 18 4: 20 5: 64 6: 32			1: 0 2: 0 3: 10 4: 13 5: 120 6: 47	

Note. Educational degree: 1 = no educational degree, 2 = certificate of secondary education, 3 = secondary school level I certificate, 4 = vocational/university entrance diploma, 5 = college/university degree, 6 = PhD. GC = gifted classes. RC = regular classes. ACH = achievement. ASC = academic self-concept. [a] = comparison with all 283 students from GC. *d* = standardized mean difference Cohen's *d* (d_1 : difference of *M* GC and *M* RC divided by *SD* GC; d_2 : difference of *M* $GC_{\text{matched sample}}$ and *M* GC divided by *SD* GC).

Table C2
Common-Item-Design for Equating, as well as Item Parameters and Reliability of Person Parameters of the Vertically Linked Math Test Versions (Test 1 to 4)

	Test 1	Test 2	Test 3	Test 4	*N*	IP min	IP max	WLE-Rel
Test 1	26				857	-2.34	1.22	.77
Test 2	17	31			997	-2.19	1.66	.84
Test 3	9	11	37		899	-1.56	3.74	.88
Test 4	6	10	17	39	804	-1.51	2.66	.86

Note. Sum of items per test version in the diagonal. *N* = Number of participants for parameter estimation. IP min = Minimum value of item parameters. IP max = Maximum value of item parameters. WLE-Rel = Reliability of WLE person parameters.

Table C3
BFLPE analyses for the Whole Sample: Regression Coefficients from Multilevel Analyses for each Wave of Measurement (T1 – T4) relating Math Academic Self-Concept (ASC) to Math Ability, Prior Math ASC, Class-Average Math Ability, and Class type (Gifted vs. Regular)

	Math ASC T1			**Math ASC T2**			**Math ASC T3**			**Math ASC T4**		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.311	.046	<.001	.274	.040	<.001	.275	.038	<.001	.281	.037	<.001
Prior ASC_{T-1}	--	--	--	.524	.028	<.001	.539	.038	<.001	.611	.035	<.001
Class level												
Math ability	-.277	.149	.063	-.198	.108	.069	-.142	.080	.075	-.238	.108	.027
Class type[a]	.360	.130	.006	.166	.124	.179	.056	.087	.523	.168	.124	.176

Notes. [a] Class type coded as 0 = regular classes vs. 1 = gifted classes.

Table C4

Results for the Tests of Measurement Invariance over Class Type (RC = Regular Classes, GC = Gifted Classes) for the Academic Self-Concept Measure by Wave of Measurement (Matched Sample; Model 3)

Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Start grade 5							
GC	10.556	2	1.437		.951		.175
RC	0.102	2	1.100		1		.000
Configural	12.062	4	1.266		.967		.120
Metric	15.656	7	1.255	3.526 (3), .317	.964	.003	.094
Scalar	18.373	10	1.188	2.117 (3), .549	.965	.001	.077
End grade 5							
GC	0.601	2	1.346		1		.000
RC	3.554	2	1.392		.988		.076
Configural	4.205	4	1.369		.999		.019
Metric	12.253	7	1.485	7.587 (3), .055	.973	.026	.073
Partial metric (loading item acasm1_2 set free)	4.315	6	1.495	0.398 (2), .820	1	.001	.000
Scalar (based on the partial metric model)	8.615	9	1.307	5.166 (3), .160	1	.000	.000
End grade 6							
GC	0.295	2	2.263		1		.000
RC	0.522	2	1.931		1		.000
Configural	0.826	4	2.097		1		.000
Metric	1.946	7	1.640	1.415 (3), .702	1	.000	.000
Scalar	4.091	10	1.441	2.770 (3), .429	1	.000	.000
Middle grade 7							
GC	0.186	2	0.930		1		.000
RC	0.326	2	1.489		1		.000
Configural	0.545	4	1.209		1		.000
Metric	2.440	7	1.127	2.057 (3), .561	1	.000	.000
Scalar	7.534	10	1.071	5.658 (3), .130	1	.000	.000

Note. SCF = Scaling correction factor.

Table C5

Results for the Tests of Measurement Invariance over Time for the Academic Self-Concept Measure by Class Type (Matched Sample; Model 3)

Model	Chi2	df	SCF	ΔChi2 (df), *p*	CFI	ΔCFI	RMSEA
Regular classes							
Configural	93.171	74	1.114		.977		.042
Metric	109.419	83	1.094	17.149 (9), .046	.968	.009	.047
Scalar	115.107	92	1.084	5.119 (9), .824	.972	.004	.041
Gifted classes							
Configural	110.940	74	1.214		.966		.058
Metric	127.881	83	1.222	16.764 (9), .053	.959	.007	.060
Scalar	150.015	92	1.195	24.310 (9), .004	.947	.012	.065

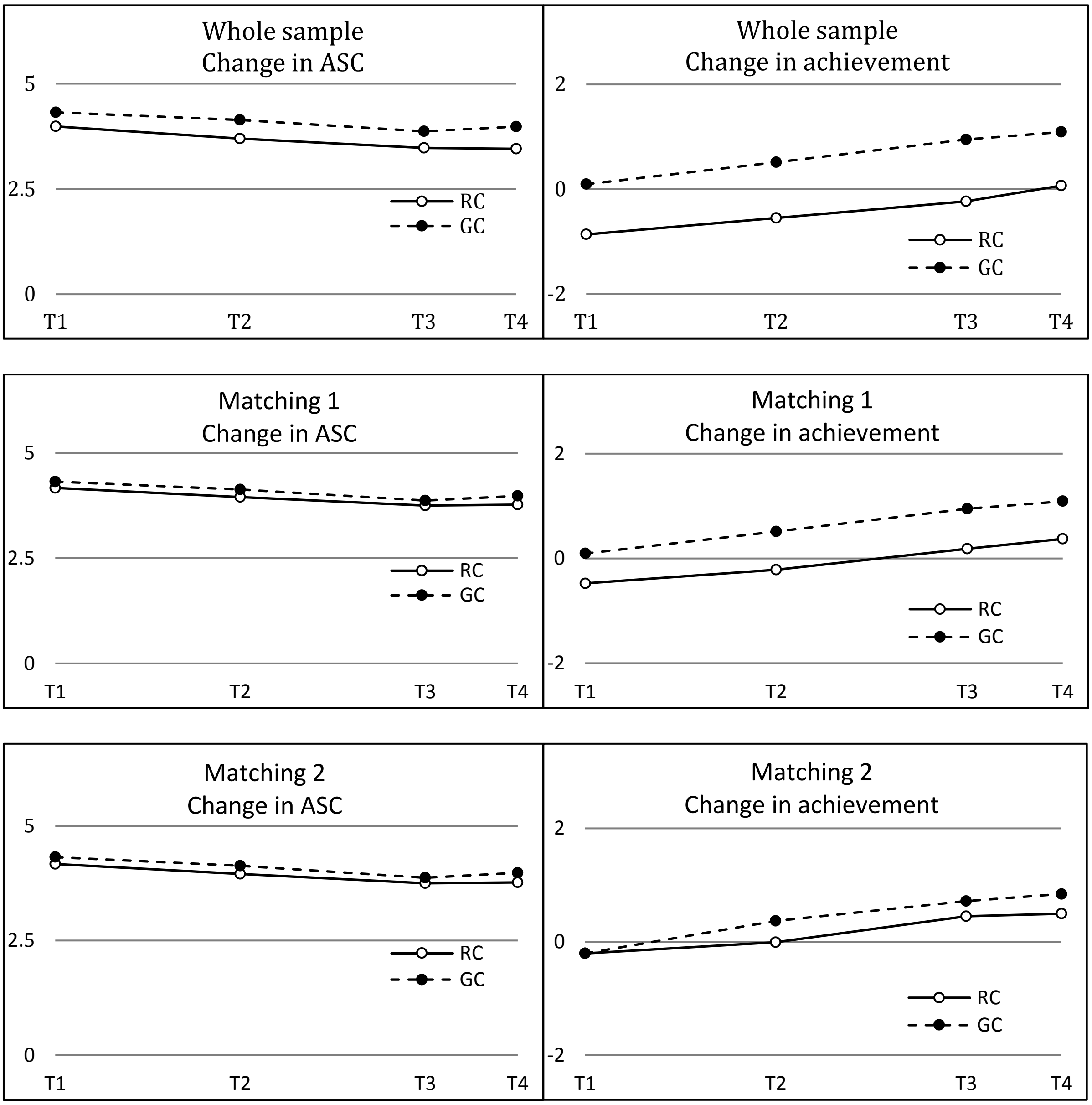
Note. SCF = Scaling correction factor.

Table C6
Results of the Univariate Latent Growth Curve Analyses with Linear Slopes (Model A) or Linear and Quadratic Slopes (Model B) for Academic Self-Concept or Achievement by Class Type (Matched Sample; Model 3)

	Gifted Classes		**Regular Classes**	
Academic self-concept				
Model A: linear growth				
Model fit				
χ^2 (*df*)	9.711 (5)		17.971** (5)	
Scaling correction factor	1.098		.971	
CFI	.980		.933	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.410***	.098	-.165*	.077
Slope linear	-.120***	.020	-.062***	.015
Variances				
Intercept	1.288***	.155	.713***	.081
Slope linear	.038***	.008	.023***	.006
Covariances				
Intercept x Slope linear	.162***	.030	.063***	.014
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	2.905 (1)		6.312 (1)	
Scaling correction factor	1.108		.794	
$\Delta\chi^2$ (*df*)	6.795 (4) , *p* = .147		12.251 (4), *p* = .016	
CFI	.992		.972	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.360***	.100	-.129	.075
Slope linear	-.710	.054	.011	.047
Slope quadratic	.010	.010	.013	.009
Variances				
Intercept	1.298***	.201	.802***	.123
Slope linear	.166	.089	.171**	.062
Slope quadratic	.002	.003	.006**	.002
Covariances				
Intercept x Slope linear	.216*	.097	.127	.074
Intercept x Slope quadratic	.007	.015	.009	.011
Slope linear x Slope quadratic	.018	.014	.028**	.010
Academic achievement				
Model A: linear growth				
Model fit				
χ^2 (*df*)	7.528 (5)		6.356 (5)	
Scaling correction factor	1.429		1.331	
CFI	.982		.964	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.991***	.092	.533***	.069
Slope linear	.225***	.017	.159***	.015

Variances				
Intercept	.981***	.142	.461***	.101
Slope linear	.011	.006	.015**	.005
Covariances				
Intercept x Slope	.082***	.023	.049***	.020
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	0.189 (1)		1.476 (1)	
Scaling correction factor	1.134		1.920	
$\Delta\chi^2$ (*df*)	7.015 (4), *p* = .135		4.752 (4), *p* = .314	
CFI	1		.987	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.897***	.097	.540***	.073
Slope linear	.088	.069	.214**	.062
Slope quadratic	-.025*	.013	.011	.011
Variances				
Intercept	.839	.216	.711***	.199
Slope linear	.067	.126	.186	.155
Slope quadratic	.001	.004	.003	.437
Covariances				
Intercept x Slope linear	.004	.149	.276*	.135
Intercept x Slope quadratic	-.015	.025	.036	.020
Slope linear x Slope quadratic	.008	.021	.025	.025

Figures



Whole sample
Change in ASC
5
2.5
0
T1
T2
T3
T4
RC
GC
Whole sample
Change in achievement
2
0
-2
T1
T2
T3
T4
RC
GC
Matching 1
Change in ASC
5
2.5
0
T1
T2
T3
T4
RC
GC
Matching 1
Change in achievement
2
0
-2
T1
T2
T3
T4
RC
GC
Matching 2
Change in ASC
5
2.5
0
T1
T2
T3
T4
RC
GC
Matching 2
Change in achievement
2
0
-2
T1
T2
T3
T4
RC
GC

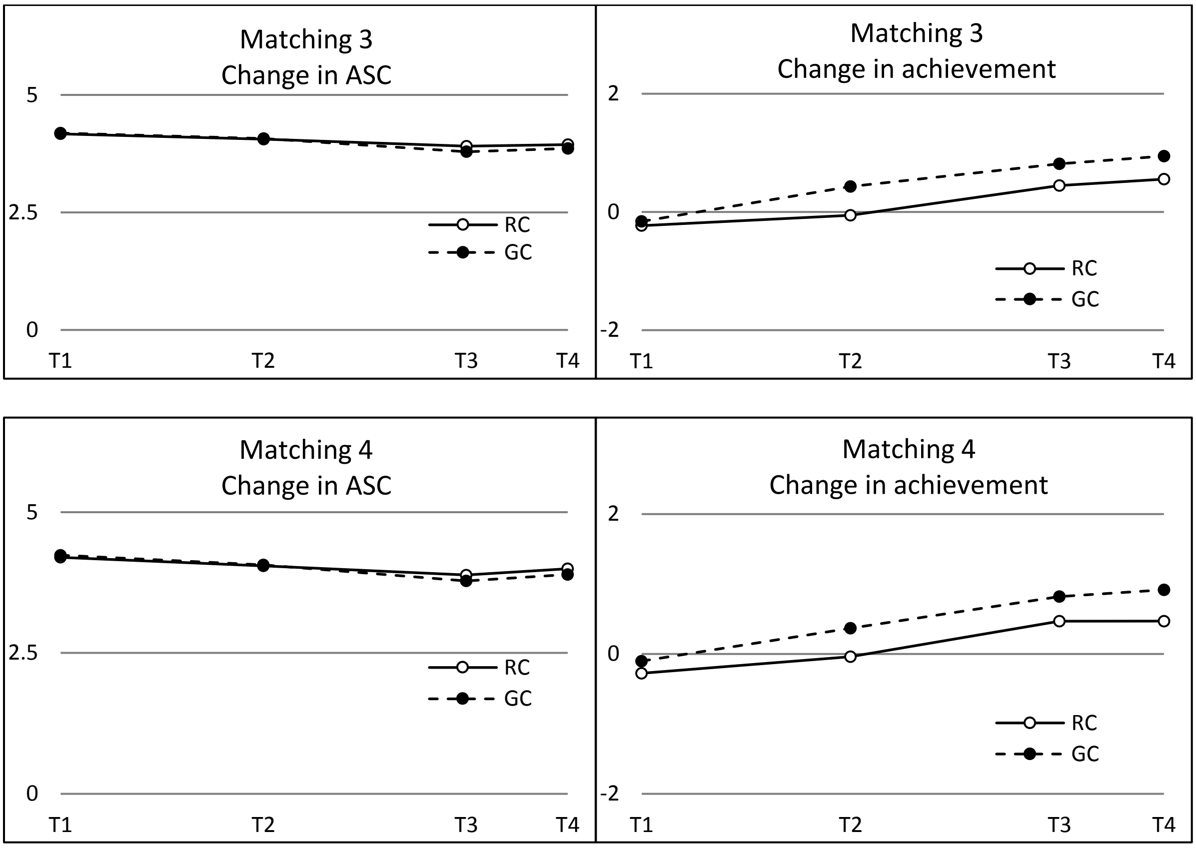


Figure A1. Group means (manifest values) in math academic self-concept (ASC) and math achievement across waves of measurement (T1 to T4) by class type (RC = regular classes, GC = gifted classes) for the whole data set and the matched data sets (Matching 1 to 4).

References

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhya, Ser. A., 35,* 417-446. doi: 10.1017/cbo9780511810725.005

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum. doi: 10.4324/9780203771587

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). *The MatchIt package* (Version 2.4–20) [Computer software]. Retrieved from http://gking.harvard.edu/matchit

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9,* 403-425. doi: 10.1037/1082-989X.9.4.403

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L., & Griffin B. A. (2012). *Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the Twang Package*. Retrieved from http://cran.r-project.org/src/contrib/Descriptions/twang.html.

Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55. doi: 10.1093/biomet/70.1.41

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2,* 169-188. doi: 10.1023/A:1020363010465

Stone, C. A., & Tang, Y. (2013). Comparing propensity score methods in balancing covariates and recovering impact in small sample educational program evaluations. *Practical Assessment, Research & Evaluation, 18(13),* 1-12. Available online: http://pareonline.net/getvn.asp?v=18&n=13

Appendix D

Results of the Robustness Checks for Propensity Score Matching Solutions or Models (Match 1, Match 2, Match 3, Match 4)

Table D1

Regression Coefficients from Multilevel Analyses for each Wave of Measurement (T1 – T4) relating Math Academic Self-Concept (ASC) to Math Ability, Prior Math Academic Self-concept, Class-Average Math Ability, and Class Type (Gifted vs. Regular)

MATCH 1	**Math ASC T1**			**Math ASC T2**			**Math ASC T3**			**Math ASC T4**		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.245	.052	<.001	.292	.055	<.001	.324	.055	<.001	.201	.039	<.001
Prior ASC_{T-1}	--	--	--	.490	.046	<.001	.537	.051	<.001	.656	.046	<.001
Class level												
Math ability	-.150	.176	.396	-.243	.153	.053	-.133	.094	.156	-.179	.097	.065
Class type[a]	.146	.127	.251	.095	.145	.513	-.028	.096	.770	.077	.122	.529

MATCH 2	**Math ASC T1**			**Math ASC T2**			**Math ASC T3**			**Math ASC T4**		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.236	.071	<.001	.264	.082	.001	.266	.091	.003	.256	.043	<.001
Prior ASC_{T-1}	--	--	--	.441	.067	<.001	.475	.077	<.001	.623	.064	<.001
Class level												
Math ability	-.261	.183	.153	-.083	.186	.655	-.092	.144	.520	-.077	.139	.577
Class type[a]	.255	.165	.121	-.073	.194	.706	-.064	.165	.701	-.115	.195	.554

MATCH 3	Math ASC T1			Math ASC T2			Math ASC T3			Math ASC T4		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.212	.079	.007	.284	.079	<.001	.233	.075	.002	.238	.049	<.001
Prior ASC_{T-1}	--	--	--	.444	.059	<.001	.535	.068	<.001	.606	.057	<.001
Class level												
Math ability	-.229	.214	.284	-.150	.170	.376	-.087	.125	.486	-.070	.120	.556
Class type[a]	.177	.201	.379	-.016	.166	.922	-.031	.132	.815	-.192	.176	.274

MATCH 4	Math ASC T1			Math ASC T2			Math ASC T3			Math ASC T4		
	B	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*	*B*	*SE*	*p*
Individual level												
Math ability	.180	.066	.006	.229	.064	<.001	.290	.078	<.001	.219	.037	<.001
Prior ASC_{T-1}	--	--	--	.491	.054	<.001	.512	.074	<.001	.617	.051	<.001
Class level												
Math ability	-.267	.209	.201	-.053	.146	.716	-.083	.127	.510	-.056	.105	.597
Class type[a]	.230	.186	.261	-.106	.148	.474	-.096	.138	.484	-.168	.138	.225

Notes. [a] Class type coded as 0 = regular classes vs. 1 = gifted classes.

Table D2

Results for the Tests of Measurement Invariance over Class-Type (RC = Regular Classes, GC = Gifted Classes) for the Academic Self-Concept Measure by Wave of Measurement

MATCH 1 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Start grade 5							
GC	8.050	2	1.494		.982		.107
RC	0.615	2	1.086		1		.000
Configural	9.841	4	1.290		.990		.074
Metric	15.838	7	1.311	6.026 (3), .110	.985	.005	.069
Scalar	17.099	10	1.237	0.369 (3), .947	.988	.003	.052
End grade 5							
GC	0.551	2	1.858		1		.000
RC	2.588	2	1.227		.998		.034
Configural	2.723	4	1.543		1		.000
Metric	8.298	7	1.525	5.631 (3), .131	.997	.003	.027
Scalar	10.296	10	1.370	1.443 (3), .696	.999	.002	.011
End grade 6							
GC	1.182	2	2.070		1		.000
RC	0.188	2	2.695		1		.000
Configural	1.240	4	2.383		1		.000
Metric	3.009	7	1.774	2.475 (3), .480	1	.000	.000
Scalar	4.078	10	1.543	0.950 (3), .813	1	.000	.000
Middle grade 7							
GC	0.240	2	1.033		1		.000
RC	3.760	2	1.169		.993		.064
Configural	4.218	4	1.101		1		.016
Metric	9.032	7	1.091	4.833 (3), .184	.996	.004	.036
Scalar	15.961	10	1.056	7.188 (3), .066	.990	.006	.051
MATCH 2 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Start grade 5							
GC	9.205	2	1.253		.953		.169
RC	0.230	2	1.150		1		0
Configural	9.820	4	1.201		.977		.107

Metric	20.203	7	1.207	10.365 (3), .016	.947	.030	.122
Partial metric (loading item acasm3_1 set free)	13.814	6	1.221	4.022 (2), .134	.969	.008	.102
Scalar (based on the partial metric model)	15.191	9	1.156	0.681 (3), .878	.975	.006	.074
End grade 5							
GC	1.960	2	1.163		1		.000
RC	7.766	2	1.011		.955		.154
Configural	9.322	4	1.087		.973		.104
Metric	20.876	7	1.373	10.563 (3), .014	.929	.044	.127
Partial metric (loading item acasm1_2 set free)	8.642	6	1.320	0.712 (2), .700	.986	.013	.060
Scalar (based on the partial metric model)	17.785	9	1.213	10.181 (3), .017	.955	.031	.089
Partial scalar (based on the partial metric model plus intercept item acasm1_2 set free)	10.611	8	1.256	1.804 (2), .406	.987	.001	.052
End grade 6							
GC	1.818	2	1.831		1		.000
RC	0.898	2	1.128		1		.000
Configural	2.935	4	1.480		1		.000
Metric	3.607	7	1.285	0.284 (3), .963	1	.000	.000
Scalar	5.444	10	1.193	1.901 (3), .593	1	.000	.000
Middle grade 7							
GC	0.000	2	0.907		1		.000
RC	0.428	2	1.266		1		.000
Configural	0.499	4	1.087		1		.000
Metric	4.050	7	1.105	3.481 (3), .323	1	.000	.000
Scalar	8.290	10	1.067	4.469 (3), .215	1	.000	.000

MATCH 3 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Start grade 5							
GC	10.556	2	1.437		.951		.175
RC	0.102	2	1.100		1		.000
Configural	12.062	4	1.266		.967		.120
Metric	15.656	7	1.255	3.526 (3), .317	.964	.003	.094
Scalar	18.373	10	1.188	2.117 (3), .549	.965	.001	.077
End grade 5							
GC	0.601	2	1.346		1		.000

RC	3.554	2	1.392		.988		.076
Configural	4.205	4	1.369		.999		.019
Metric	12.253	7	1.485	7.587 (3), .055	.973	.026	.073
Partial metric (loading item acasm1_2 set free)	4.315	6	1.495	0.398 (2), .820	1	.001	0
Scalar (based on the partial metric model)	8.615	9	1.307	5.166 (3), .160	1	.000	0
End grade 6							
GC	0.295	2	2.263		1		.000
RC	0.522	2	1.931		1		.000
Configural	0.826	4	2.097		1		.000
Metric	1.946	7	1.640	1.415 (3), .702	1	.000	.000
Scalar	4.091	10	1.441	2.770 (3), .429	1	.000	.000
Middle grade 7							
GC	0.186	2	0.930		1		.000
RC	0.326	2	1.489		1		.000
Configural	0.545	4	1.209		1		.000
Metric	2.440	7	1.127	2.057 (3), .561	1	.000	.000
Scalar	7.534	10	1.071	5.658 (3), .130	1	.000	.000

MATCH 4 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Start grade 5							
GC	10.960	2	1.355		.953		.168
RC	1.040	2	1.080		1		.000
Configural	13.119	4	1.217		.969		.121
Metric	24.091	7	1.202	10.990 (3), .012	.942	.027	.125
Partial metric (loading item acasm1_1 set free)	18.760	6	1.245	5.679 (2), .058	.956	.013	.117
Scalar (based on the partial metric model)	20.422	9	1.177	0.660 (3), .882	.961	.005	.090
End grade 5							
GC	0.636	2	1.338		1		.000
RC	7.864	2	1.024		.960		.139
Configural	7.540	4	1.181		.985		.075
Metric	20.017	7	1.334	11.573 (3), .009	.945	.040	.109
Partial metric (loading item acasm1_2 set free)	8.051	6	1.298	1.009 (2), .604	.991	.006	.047
Scalar (based on the partial metric model)	13.393	9	1.182	5.666 (3), .129	.981	.010	.056

End grade 6							
GC	0.428	2	2.132		1		.000
RC	0.133	2	1.924		1		.000
Configural	0.577	4	2.028		1		.000
Metric	1.409	7	1.588	1.065 (3), .786	1	.000	.000
Scalar	2.431	10	1.406	1.205 (3), .752	1	.000	.000
Middle grade 7							
GC	0.607	2	0.974		1		.000
RC	1.097	2	1.252		1		.000
Configural	1.767	4	1.113		1		.000
Metric	3.846	7	1.096	2.097 (3), .553	1	.000	.000
Scalar	6.959	10	1.051	3.272 (3), .352	1	.000	.000

Note. SCF = Scaling correction factor.

Table D3
Results for the Tests of Measurement Invariance over Time for the Academic Self-Concept Measure by Class Type

MATCH 1 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Regular classes							
Configural	92.229	74	1.187		.990		.030
Metric	116.636	83	1.166	26.714 (9), .002	.982	.008	.038
Scalar	129.546	92	1.148	12.954 (9), .165	.980	.002	.038
Gifted classes							
Configural	84.156	74	1.334		.995		.022
Metric	98.995	83	1.326	15.079 (9), .089	.992	.003	.026
Scalar	115.034	92	1.293	17.672 (9), .039	.989	.003	.030
MATCH 2 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Regular classes							
Configural	96.642	74	1.044		.974		.048
Metric	117.460	83	1.042	21.015 (9), .013	.961	.013	.056
Scalar	127.119	92	1.040	9.623 (9), .382	.960	.001	.054
Gifted classes							
Configural	122.277	74	1.137		.955		.071
Metric	143.752	83	1.144	21.158 (9), .012	.943	.012	.075
Scalar	170.449	92	1.127	28.492 (9), .001	.927	.016	.081
MATCH 3 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Regular classes							
Configural	93.171	74	1.114		.977		.042
Metric	109.419	83	1.094	17.149 (9), .046	.968	.009	.047
Scalar	115.107	92	1.084	5.119 (9), .824	.972	.004	.041
Gifted classes							
Configural	110.940	74	1.214		.966		.058
Metric	127.881	83	1.222	16.764 (9), .053	.959	.007	.060
Scalar	150.015	92	1.195	24.310 (9), .004	.947	.012	.065
MATCH 4 Model	**Chi2**	**df**	**SCF**	**ΔChi2 (df), *p***	**CFI**	**ΔCFI**	**RMSEA**
Regular classes							
Configural	89.468	74	1.126		.985		.036
Metric	111.283	83	1.112	22.997 (9), .006	.973	.012	.046
Scalar	122.829	92	1.099	11.480 (9), .244	.971	.002	.045
Gifted classes							
Configural	104.426	74	1.242		.975		.050
Metric	120.234	83	1.236	15.937 (9), .068	.969	.006	.052
Scalar	135.301	92	1.214	15.474 (9), .079	.964	.005	.053

Note. SCF = Scaling correction factor.

Table D4

Results of the Univariate Latent Growth Curve Analyses with Linear Slopes (Model A) or Linear and Quadratic Slopes (Model B) for Academic Self-Concept or Achievement by Class Type

Match 1	**Gifted Classes**		**Regular Classes**	
Academic self-concept				
Model A: linear growth				
Model fit				
χ^2 (*df*)	52.819*** (5)		20.455*** (5)	
SCF	1.2047		1.0847	
CFI	.897		.966	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.220**	.074	-.321***	.067
Slope linear	-.125***	.014	-.095***	.012
Variances				
Intercept	1.393***	.124	1.121***	.092
Slope linear	.036***	.006	.027***	.005
Covariances				
Intercept x Slope linear	.169***	.024	.110***	.016
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	13.957*** (1)		4.610* (1)	
SCF	1.1245		.9069	
$\Delta\chi^2$ (*df*)	39.140 (4), $p <$ .001		15.947 (4), $p = .003$	
CFI	.972		.992	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.086	.074	-.295***	.067
Slope linear	-.015	.040	-.019	.034
Slope quadratic	.019**	.007	.015*	.006
Variances				
Intercept	1.514***	.159	1.138***	.122
Slope linear	.316***	.058	.111	.061
Slope quadratic	.006***	.002	.005**	.002
Covariances				
Intercept x Slope linear	.348***	.078	.116	.076
Intercept x Slope quadratic	.028**	.011	.000	.012
Slope linear x Slope quadratic	.044***	.009	.019*	.009
Academic achievement				
Model A: linear growth				
Model fit				
χ^2 (*df*)	8.440 (5)		5.560 (5)	
SCF	1.255		1.525	

CFI	.988		.995	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	1.105***	.066	.349***	.053
Slope linear	.205***	.012	.168***	.012
Variances				
Intercept	.899***	.113	.519***	.079
Slope linear	.009	.005	.018***	.005
Covariances				
Intercept x Slope	.062**	.018	.045**	.015
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	0.166 (1)		0.070 (1)	
SCF	1.004		0.992	
$\Delta\chi^2$ (*df*)	7.910 (4), p = .095		5.071 (4), p = .280	
CFI	1		1	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	1.081***	.069	.367***	.055
Slope linear	.170***	.047	.219***	.043
Slope quadratic	-.006	.009	.010	.008
Variances				
Intercept	.878***	.182	.751***	.156
Slope linear	.159	.098	.198	.125
Slope quadratic	.003	.003	.002	.003
Covariances				
Intercept x Slope linear	.085	.115	.237*	.099
Intercept x Slope quadratic	.002	.019	.029*	.014
Slope linear x Slope quadratic	.024	.016	.023	.020

Match 2	**Gifted Classes**		**Regular Classes**	
Academic self-concept				
Model A: linear growth				
Model fit				
χ^2 (*df*)	10.773 (5)		12.513* (5)	
SCF	1.033		.950	
CFI	.972		.965	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.358***	.109	-.083	.075
Slope linear	-.117***	.023	-.044**	.014
Variances				
Intercept	1.449***	.193	.735***	.092
Slope linear	.046***	.012	.023***	.006
Covariances				
Intercept x Slope linear	.204***	.043	.054***	.014
Model B: quadratic growth				

Model fit				
χ^2 (*df*)	2.469 (1)		2.936 (1)	
SCF	1.107		.900	
$\Delta\chi^2$ (*df*)	8.276 (1), *p* = .082		9.605 (1), *p* = .048	
CFI	.993		.991	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.302**	.111	-.081	.074
Slope linear	-.072	.061	.009	.045
Slope quadratic	.008	.011	.010	.009
Variances				
Intercept	1.567***	.267	.764***	.129
Slope linear	.291**	.104	.117	.070
Slope quadratic	.007*	.003	.005*	.002
Covariances				
Intercept x Slope linear	.379**	.133	.057	.076
Intercept x Slope quadratic	.030	.018	-.001	.011
Slope linear x Slope quadratic	.041*	.016	.021	.012
Academic achievement				
Model A: linear growth				
Model fit				
χ^2 (*df*)	6.395 (5)		3.986 (5)	
SCF	1.328		1.168	
CFI	.990		1	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.883***	.103	.487***	.080
Slope linear	.211***	.020	.145***	.015
Variances				
Intercept	1.063***	.194	.551***	.119
Slope linear	.018*	.008	.009	.005
Covariances				
Intercept x Slope	.107***	.031	.042*	.021
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	0.243 (1)		0.053 (1)	
SCF	1.187		0.927	
$\Delta\chi^2$ (*df*)	6.017(4), *p* = .198		3.751 (4), *p* = .441	
CFI	1		1	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.715***	.122	.644***	.111
Slope linear	.085	.105	.126	.079
Slope quadratic	-.029	.019	-.006	.015
Variances				
Intercept	.589**	.187	.960**	.348

Slope linear	.129	.122	.111	.246
Slope quadratic	.006	.004	-.002	.007
Covariances				
Intercept x Slope linear	.020	.152	.321	.204
Intercept x Slope quadratic	-.009	.025	.042	.028
Slope linear x Slope quadratic	.026	.019	.006	.040

Match 3	**Gifted Classes**		**Regular Classes**	
Academic self-concept				
Model A: linear growth				
Model fit				
χ^2 (*df*)	9.711 (5)		17.971** (5)	
SCF	1.098		0.971	
CFI	.980		.933	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.410***	.098	-.165*	.077
Slope linear	-.120***	.020	-.062***	.015
Variances				
Intercept	1.288***	.155	.713***	.081
Slope linear	.038***	.008	.023***	.006
Covariances				
Intercept x Slope linear	.162***	.030	.063***	.014
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	2.905 (1)		6.312 (1)	
SCF	1.108		.794	
$\Delta\chi^2$ (*df*)	6.795 (4), *p* = .147		12.251 (4), *p* = .016	
CFI	.992		.972	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.360***	.100	-.129	.075
Slope linear	-.71	.054	.011	.047
Slope quadratic	.010	.010	.013	.009
Variances				
Intercept	1.298***	.201	.802***	.123
Slope linear	.166	.089	.171**	.062
Slope quadratic	.002	.003	.006**	.002
Covariances				
Intercept x Slope linear	.216*	.097	.127	.074
Intercept x Slope quadratic	.007	.015	.009	.011
Slope linear x Slope quadratic	.018	.014	.028**	.010
Academic achievement				
Model A: linear growth				
Model fit				
χ^2 (*df*)	7.528 (5)		6.356 (5)	

SCF	1.429		1.331	
CFI	.982		.964	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.991***	.092	.533***	.069
Slope linear	.225***	.017	.159***	.015
Variances				
Intercept	.981***	.142	.461***	.101
Slope linear	.011	.006	.015**	.005
Covariances				
Intercept x Slope	.082***	.023	.049***	.020
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	0.189 (1)		1.476 (1)	
SCF	1.134		1.920	
$\Delta\chi^2$ (*df*)	7.015 (4), p = .135		4.752 (4), p = .314	
CFI	1		.987	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.897***	.097	.540***	.073
Slope linear	.088	.069	.214**	.062
Slope quadratic	-.025*	.013	.011	.011
Variances				
Intercept	.839***	.216	.711***	.199
Slope linear	.067	.126	.186	.155
Slope quadratic	.001	.004	.003	.437
Covariances				
Intercept x Slope linear	.004	.149	.276*	.135
Intercept x Slope quadratic	-.015	.025	.036	.020
Slope linear x Slope quadratic	.008	.021	.025	.025

Match 4	**Gifted Classes**		**Regular Classes**	
Academic self-concept				
Model A: linear growth				
Model fit				
χ^2 (*df*)	76.496*** (5)		75.778*** (5)	
SCF	1.091		1.088	
CFI	.745		.680	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.505***	.142	-.158	.101
Slope linear	-.165***	.033	-.093***	.020
Variances				
Intercept	2.819***	.398	1.294***	.139
Slope linear	.101***	.020	.015	.010
Covariances				
Intercept x Slope linear	.572***	.080	.219***	.031

Model B: quadratic growth				
Model fit				
χ^2 (*df*)	5.046* (1)		1.650 (1)	
SCF	1.337		.911	
$\Delta\chi^2$ (*df*)	74.512 (4), $p <$ .001		71.497 (4), $p < .001$	
CFI	.986		.997	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	-.422***	.120	-.132	.091
Slope linear	-.028	.078	.179*	.070
Slope quadratic	.028	.016	.061***	.016
Variances				
Intercept	2.248***	.330	1.271***	.171
Slope linear	.411*	.161	.434***	.128
Slope quadratic	.021***	.006	.022***	.006
Covariances				
Intercept x Slope linear	.123	.179	-.034	.147
Intercept x Slope quadratic	-.082**	.029	-.058*	.028
Slope linear x Slope quadratic	.083**	.030	.095***	.026

Academic achievement

Model A: linear growth				
Model fit				
χ^2 (*df*)	5.583 (5)		6.413 (5)	
SCF	1.317		1.312	
CFI	.996		.974	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.968***	.091	.487***	.065
Slope linear	.214***	.017	.157***	.014
Variances				
Intercept	1.037***	.169	.454***	.099
Slope linear	.013	.007	.011*	.005
Covariances				
Intercept x Slope	.097***	.027	.038*	.018
Model B: quadratic growth				
Model fit				
χ^2 (*df*)	0.266 (1)		1.948 (1)	
SCF	0.962		1.722	
$\Delta\chi^2$ (*df*)	5.048 (4), $p = .282$		4.182 (4), $p = .382$	
CFI	1		.982	
Parameters	Unstand. Estimate	SE	Unstand. Estimate	SE
Means				
Intercept	.909***	.098	.475***	.070
Slope linear	.136*	.066	.148**	.056
Slope quadratic	-.014	.012	-.002	.010
Variances				

Intercept	1.017***	.271	.706***	.199
Slope linear	.181	.140	.154	.149
Slope quadratic	.004	.005	.002	.004
Covariances				
Intercept x Slope linear	.141	.168	.257*	.126
Intercept x Slope quadratic	.006	.027	.035	.018
Slope linear x Slope quadratic	.027	.024	.019	.024

Table D5
Results of the Conditional Dual-Process Latent Growth Curve Model for Achievement (ACH) and Academic Self-Concept (ASC)

Match 1	Unstandardized Estimates		SE	
	Regular classes [a]	Gifted classes [a]	Regular classes [a]	Gifted classes [a]
Means				
I-ASC	-.288, p < .001	-.106, p = .338	.071	.111
S-ASC	-.020, p = .570	-.013, p = .828	.035	.061
Q-ASC	.015, p = .022	.019, p = .070	.006	.011
I-ACH	.338, p = .003	1.097, p < .001	.112	.120
S-ACH	.167, p < .001	.203, p < .001	.015	.012
Residual Variances				
I-ASC	1.319, p < .001		.113	
S-ASC	.209, p < .001		.048	
Q-ASC	.006, p < .001		.001	
I-ACH	.686, p < .001		.086	
S-ACH	.013, p = .006		.005	
Covariances				
I-ASC with S-ASC	.227, p < .001		.058	
I-ASC with Q-ASC	.014, p = .106		.008	
S-ASC with Q-ASC	.031, p < .001		.007	
I-ASC with I-ACH	.521, p < .001		.065	
I-ACH with S-ASC	.008, p = .767		.026	
I-ACH with Q-ASC	-.008, p = .042		.004	
I-ASC with S-ACH	.038, p < .001		.010	
S-ASC with S-ACH	.003, p = .597		.005	
Q-ASC with S-ACH	.000, p = .796		.001	
I-ACH with S-ACH	.050, p < .001		.014	
Class type on ASC [a]				
I-ASC	.182, p = .165		.131	
S-ASC	.006, p = .927		.069	
Q-ASC	.004, p = .718		.012	
I-ACH	.760, p < .001		.164	
S-ACH	.036, p = .063		.019	

Match2	Unstandardized Estimates		SE	
	Regular classes [a]	Gifted classes [a]	Regular classes [a]	Gifted classes [a]
Means				
I-ASC	-.086, p = .386	-.288, p = .041	.099	.141
S-ASC	.005, p = .927	-.068, p = .370	.053	.076
Q-ASC	.010, p = .316	.008, p = .580	.010	.014
I-ACH	.494, p < .001	.877, p < .001	.124	.134
S-ACH	.147, p < .001	.212, p < .001	.015	.022
Residual Variances				
I-ASC	1.197, p < .001		.164	
S-ASC	.225, p < .001		.055	
Q-ASC	.006, p < .001		.002	
I-ACH	.786, p < .001		.128	

S-ACH	.015, *p* = .012	.006
Covariances		
I-ASC with S-ASC	.241, *p* = .002	.077
I-ASC with Q-ASC	.018, *p* = .097	.011
S-ASC with Q-ASC	.034, *p* < .001	.009
I-ASC with I-ACH	.513, *p* < .001	.093
I-ACH with S-ASC	.042, *p* = .303	.041
I-ACH with Q-ASC	-.001, *p* = .826	.006
I-ASC with S-ACH	.040, *p* = .002	.013
S-ASC with S-ACH	.009, *p* = .219	.008
Q-ASC with S-ACH	.001, *p* = .492	.001
I-ACH with S-ACH	.074, *p* = .001	.022
Class type on ASC [a]		
I-ASC	-.202, *p* = .242	.172
S-ASC	-.073, *p* = .422	.094
Q-ASC	-.003, *p* = .869	.017
I-ACH	.383, *p* = .035	.182
S-ACH	.065, *p* = .015	.027

Match 3	Unstandardized Estimates		SE	
	Regular classes [a]	Gifted classes [a]	Regular classes [a]	Gifted classes [a]
Means				
I-ASC	-.144, *p* = .081	-.343, *p* = .006	.082	.126
S-ASC	.003, *p* = .942	-.066, *p* = .302	.038	.064
Q-ASC	.013, *p* = .058	.009, *p* = .403	.007	.011
I-ACH	.530, *p* < .001	.972, *p* < .001	.110	.147
S-ACH	.159, *p* < .001	.224, *p* < .001	.017	.018
Residual Variances				
I-ASC	1.066, *p* < .001		.126	
S-ASC	.181, *p* < .001		.046	
Q-ASC	.004, *p* = .011		.001	
I-ACH	.695, *p* < .001		.105	
S-ACH	.015, *p* = .005		.005	
Covariances				
I-ASC with S-ASC	.184, *p* = .002		.059	
I-ASC with Q-ASC	.010, *p* = .294		.009	
S-ASC with Q-ASC	.024, *p* = .001		.007	
I-ASC with I-ACH	.432, *p* < .001		.070	
I-ACH with S-ASC	.012, *p* = .735		.034	
I-ACH with Q-ASC	-.005, *p* = .439		.006	
I-ASC with S-ACH	.037, *p* = .003		.012	
S-ASC with S-ACH	.006, *p* = .419		.008	
Q-ASC with S-ACH	.001, *p* = .713		.001	
I-ACH with S-ACH	.067, *p* < .001		.018	
Class type on ASC [a]				
I-ASC	-.199, *p* = .184		.150	
S-ASC	-.069, *p* = .355		.074	
Q-ASC	-.004, *p* = .758		.013	

I-ACH	.441, *p* = .016		.184	
S-ACH	.065, *p* = .010		.025	

Match4	Unstandardized Estimates		SE	
	Regular classes [a]	Gifted classes [a]	Regular classes [a]	Gifted classes [a]
Means				
I-ASC	-.138, *p* = .205	-.387, *p* = .036	.109	.185
S-ASC	.175, *p* = .028	-.022, *p* = .853	.079	.117
Q-ASC	.061, *p* = .001	.028, *p* = .312	.019	.027
I-ACH	.502, *p* < .001	.944, *p* < .001	.117	.145
S-ACH	.160, *p* < .001	.210, *p* < .001	.018	.017
Residual Variances				
I-ASC	1.820, *p* < .001		.216	
S-ASC	.466, *p* < .001		.090	
Q-ASC	.023, *p* < .001		.004	
I-ACH	.713, *p* < .001		.115	
S-ACH	.013, *p* = .034		.006	
Covariances				
I-ASC with S-ASC	.099, *p* = .340		.104	
I-ASC with Q-ASC	-.061, *p* = .301		.018	
S-ASC with Q-ASC	.096, *p* < .001		.017	
I-ASC with I-ACH	.642, *p* < .001		.103	
I-ACH with S-ASC	-.040, *p* = .422		.050	
I-ACH with Q-ASC	-.038, *p* < .001		.010	
I-ASC with S-ACH	.067, *p* < .001		.015	
S-ASC with S-ACH	.003, *p* = .806		.010	
Q-ASC with S-ACH	-.002, *p* = .270		.002	
I-ACH with S-ACH	.066, *p* = .001		.021	
Class type on ASC [a]				
I-ASC	-.249, *p* = .247		.216	
S-ASC	-.196, *p* = .155		.138	
Q-ASC	-.033, *p* = .309		.033	
I-ACH	.442, *p* = .018		.187	
S-ACH	.050, *p* = .044		.025	

Note. [a] = Because of the use of the categorical predictor class-type (0, regular classes; 1, gifted classes) means of latent factors refer to the group coded as 0. Therefore, the model was tested twice, once with regular classes coded as 0 and once with gifted classes coded as 0. Estimates for (co-)variances are unaffected by this coding. The size of the effects of the categorical predictor class type is also unaffected by this coding but not the prefix; the prefixes reported here refer to the coding of regular classes as 0.

References

All references that were part of the meta-analysis reported in Article 1 are reported in Appendix A and marked with an asterisk.

Adams, R. J., Smart, P., & Huff, A. S. (2016). Shades of grey: guidelines for working with the grey Literature in systematic reviews for management and organizational studies. *International Journal of Management Reviews*, *19*, 432–454. doi:10.1111/ijmr.12102

Adelson, J. L., & Carpenter, B. D. (2011). Grouping for achievement gains: for whom does achievement grouping increase kindergarten reading growth? *Gifted Child Quarterly, 55,* 265-278. doi: 10.1177/0016986211417306

Adelson, J. L., McCoach, D. B., & Gavin, M. K. (2012). Examining the effects of gifted programming in mathematics and reading using the ECLS-K. *Gifted Child Quarterly, 56,* 25-39. doi: 10.1177/0016986211431487

Alicke, M. D., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection.* New York, NY: The Guilford Press.

Allemand, M., & Martin, M. (2016). On correlated change in personality. *European Psychologist*, 21, 237–253. doi:10.1027/1016-9040/a000256

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261–271. doi:10.1037/0022-0663.84.3.261

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research, 64,* 287–309. doi:10.3102/00346543064002287

Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemporary Educational Psychology, 22,* 269–298. doi:10.1006/ceps.1996.0926

Anderman, E. M., Austin, C. C., & Johnson, D. M. (2002). The development of goal orientation. In A., Wigfield, & J. S., Eccles. (Eds.), *Development of achievement motivation* (pp. 197–220). San Diego, CA: Academic Press.

Anderman, E. M., Maehr, M. L., & Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research & Development in Education, 32,* 131–147. Retrieved from https://psycnet.apa.org/record/1999-05484-001

Anderman, L. H., & Anderman, E. M. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology, 24,* 21-37. doi:10.1006/ceps.1998.0978

Anger, C., Plünnecke, A., & Schüler, R. M. *INSM-Bildungsmonitor 2018* [INSM education monitor]. Cologne: INSM.

Au, W. (2011). Teaching under the new Taylorism: high-stakes testing and the standardization of the 21stcentury curriculum. *Journal of Curriculum Studies, 43,* 25-45. doi:10.1080/00220272.2010.521261

Bachman, J. G., & O'Malley, P. M. (1986). Self-concepts, self-esteem, and educational experiences: The frog pond revisited (again). *Journal of Personality and Social Psychology, 50,* 35–46. doi:10.1037/0022-3514.50.1.35

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28,* 117–148. doi:10.1207/s15326985ep2802_3

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs onacademic functioning. *Child Development, 67,* 1206–1222. doi:10.1111/j.1467-8624.1996.tb01791.x

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development, 72,* 187-206. doi:10.1111/1467-8624.00273

Baranik, L. E., Stanley, L. J., Bynum, B. H., & Lance, C. E. (2010). Examining the construct validity of mastery-avoidance achievement goals: A meta-analysis. *Human Performance, 23,* 265–282. doi:10.1080/08959285.2010.488463

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117,* 497. doi*:*10.1037/0033-2909.117.3.497

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370. doi: 10.1037/1089-2680.5.4.323

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1-44. doi:10.1111/1529-1006.01431

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396-403.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology, 104,* 682-699, doi: 10.1037/a0027608

Belfi, B., Goos, M., De Fraine, B., & Van Damme, J. (2012). The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. *Educational Research Review, 7,* 62-74. doi:10.1016/j.edurev.2011.09.002

Bernacki, M., Nokes-Malach, T., Richey, J. E., & Belenky, D. M. (2014). Science diaries: a brief writing intervention to improve motivation to learn science. *Educational Psychology, 36*, 26–46. doi:10.1080/01443410.2014.895293

Beycioglu, K., & Kondakci, Y. (2014). Principal leadership and organizational change in schools: a cross-cultural perspective. *Journal of Organizational Change Management, 27*. doi:10.1108/jocm-06-2014-0111

Bipp, T., Steinmayr, R., & Spinath, B. (2008). Personality and achievement motivation: Relationship among big five domain and facet scales, achievement goals, and intelligence. *Personality and Individual Differences, 44,* 1454–1464. doi:10.1016/j.paid.2008.01.001

Birkeland, M. S., Melkevik, O., Holsen, I., & Wold, B. (2012). Trajectories of global self-esteem development during adolescence. *Journal of Adolescence, 35,* 43–54. doi:10.1016/j.adolescence.2011.06.006

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93,* 23–34. doi:10.1037/0022-0663.93.1.23

Bong, M. (2005). Within-grade changes in Korean girls' motivation and perceptions of the learning environment across domains and achievement levels. *Journal of Educational Psychology, 97,* 656–672. doi:10.1037/0022-0663.97.4.656

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

Booth, M. Z., & Gerard, J. M. (2012). Adolescents' stage-environment fit in middle and high school. *Youth & Society*, *46*, 735–755. doi:10.1177/0044118x12451276

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. West Sussex: Wiley

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 125–230. doi:10.3102/00346543073002125

Bouffard, T., Boileau, L., & Vezeau, C. (2001). Students' transition from elementary to high school and changes of the relationship between motivation and academic performance. *European Journal of Psychology of Education, 16,* 589–604. doi:10.1007/bf03173199

Bowles, T. (1999). Focusing on time orientation to explain adolescent self concept and academic achievement. *Journal of Applied Health Behaviour, 1(2),* 1-8.

Bronstein, P., Ginsburg, G. S., & Herrera, I. S. (2005). Parental predictors of motivational orientation in early adolescence: A longitudinal study. *Journal of Youth and Adolescence, 34,* 559–575. doi:10.1007/s10964-005-8946-0

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

Buchanan, C. M. (2003). Mothers' generalized beliefs about adolescents: links to expectations for a specific child. *Journal of Early Adolescence, 23,* 29–50. doi:10.1177/0272431602239129

Buff, A. (2001). Warum lernen Schülerinnen und Schüler? Eine explorative Studie zur Lernmotivation auf der Basis qualitativer Daten. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 33,* 157-164. doi:10.1026//0049-8637.33.3.157

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139,* 655–701. doi:10.1037/a0029531

Burrus, J., & Brenneman, M. (2016). Psychological skills: essential components of development and achievement in K-12. In Lipnevich, A. A., Preckel, F., & Roberts, R. D. (Eds.) *Psychological Skills and School Systems in the 21st Century* (pp. 3-27). Basel, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-28606-8_1

Butler, R. (2005). Competence assessment, competence, and motivation between early and middle childhood. In Elliot, A. J., & Dweck, C. S. (Eds.), *Handbook of Competence and Motivation* (pp. 202-221). New York: Guildford Press.

Byrne, B. M. (2002). Validating the measurement and structure of self-concept: Snapshots of past, present, and future research. *American Psychologist, 57,* 896-909. doi:10.1037/0003-066X.57.11.897

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116-131. doi: 10.1037/0022-3514.42.1.116

Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *British Journal of Educational Psychology, 81,* 78-96. doi:10.1348/2044-8279.002004

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, MA: Cambridge University Press.

Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. *Journal of thoracic disease, 7,* 537–540 doi:10.3978/j.issn.2072-1439.2015.10.63

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27,* 270-295. doi:10.1006/ceps.2001.1094

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722. doi:10.1037/0012-1649.27.5.703

Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., … Riester, D. (2010). Trait goal orientation, self-regulation, and performance: A meta-analysis. *Journal of Business and Psychology, 26,* 467–483. doi:10.1007/s10869-010-9201-6

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. structural equation modeling: *A Multidisciplinary Journal, 14,* 464–504. doi:10.1080/10705510701301834

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464-504. doi: 10.1080/10705510701301834

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models. *Sociological Methods & Research, 29,* 468–508. doi:10.1177/0049124101029004003

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19,* 211-229. doi:10.1037/a0032968

Chiu, M. (2012). The internal/external frame of reference model, big-fish-little-pond effect, and combined model for mathematics and science. *Journal of Educational Psychology, 104,* 87-107. doi: 10.1037/a0025734

Chmielewski, A. K., Dumont, H., & Trautwein, U. (2013). Tracking effects depend on tracking type. *American Educational Research Journal, 50,* 925–957. doi:10.3102/0002831213489843

Chmielewski, K., Dumont, H. & Trautwein, U. (2013) Tracking effects depend on tracking type: An international comparison of academic self-concept. *American Educational Research Journal, 50,* 925-957. doi: 10.3102/0002831213489843

Chubb, N. H., Fertman, C. I., & Ross, J. L. (1997). Adolescent self-esteem and locus of control: A longitudinal study of gender and age differences. *Adolescence, 32* (125), 113-130.

Chung, J. M., Robins, R. W., Trzesniewski, K. H., Noftle, E. E., Roberts, B. W., & Widaman, K. F. (2014). Continuity and change in self-esteem during emerging adulthood. *Journal of Personality and Social Psychology, 106*, 469. doi: 10.1037/a0035135

Cohen, G. L., Garcia, J., & Goyer, P. (2017). Turning point: Targeted, tailored, and timely psychological intervention. In A. J. Elliot, C. S. Dweck, & D. S. Yaeger (Eds.), Handbook of competence and motivation (2nd ed.) (pp. 657-686). New York: The Guilford Press.

Cohen, P., Cohen, J., Aiken, L. S., & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34,* 315-346. doi:10.1207/s15327906mbr3403_2

Colangelo, N., Assouline, S. G., & Gross, M. U. M. (2004). *A nation deceived: How schools hold back America's brightest students.* The Templeton national report on acceleration. Iowa City: University of Iowa.

Craven, R. G., Marsh, H. W., & Debus, R. L. (1991). Effects of internally focused feedback and attributional feedback on enhancement of academic self-concept. *Journal of Educational Psychology, 83,* 17–27. doi:10.1037/0022-0663.83.1.17

Craven, R. G., Marsh, H. W., & Print, M. (2000). Selective, streamed and mixed-ability programs for gifted students: Impact on self-concept, motivation and achievement. *Australian Journal of Education, 44,* 51-57. doi: 10.1177/000494410004400106

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience* (pp. 15–35). New York: Cambridge University Press.

Cury, F., Elliot, A., Sarrazin, P., Da Fonseca, D., & Rufo, M. (2002). The trichotomous achievement goal model and intrinsic motivation: a sequential mediational analysis. *Journal of Experimental Social Psychology, 38*, 473–481. doi:10.1016/s0022-1031(02)00017-3

Dai, D. Y. (2004). How universal is the big-fish-little-pond effect? *American Psychologist, 59,* 267-268. doi: 10.1037/0003-066X.59.4.267

Dai, D. Y., & Rinn, A. N. (2008). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review, 20,* 283-317. doi: 10.1007/s10648-008-9071-x

Dai, D. Y., Swanson, J. A., & Cheng, H. (2011). State of research on giftedness and gifted education: A survey of empirical studies published during 1998-2010. *Gifted Child Quarterly, 55,* 126-138. doi: 10.1177/0016986210397831

Davies, J., & Brember, I. (1999). Self-esteem and national tests in years 2 and 6: a 4 year longitudinal study. *Educational Psychology, 19,* 337–345. doi:10.1080/0144341990190307

Deary, I., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21. doi:10.1016/j.intell.2006.02.001

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (1985a). Conceptualizations of intrinsic motivation and self-determination. *Intrinsic Motivation and Self-Determination in Human Behavior*, 11–40. doi:10.1007/978-1-4899-2271-7_2

Deci, E. L., & Ryan, R. M. (1985b). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19,* 109–134. doi:10.1016/0092-6566(85)90023-6

Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation. *The Corsini Encyclopedia of Psychology*, 1-2. doi:10.1002/9780470479216.corpsy0467

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627. doi:10.1037/0033-2909.125.6.627

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124,* 197–229. doi:10.1037/0033-2909.124.2.197

Denissen, J. J., 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,* 430–447. doi:10.1111/j.1467-8624.2007.01007.x

Dickhäuser O., Janke, S., Praetorius, A. K., & Dresel, M. (2017). The Effects of Teachers' Reference Norm Orientations on Students' Implicit Theories and Academic Self-Concepts. *Zeitschrift für Pädagogische Psychologie*. doi:10.1024/1010-0652/a000208

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463. doi:10.1111/j.0006-341x.2000.00455.x

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040–1048. doi:10.1037/0003-066x.41.10.1040

Dweck, C. S. (1989). Motivation. In A. Lesgold & R. Glaser (Eds.), *Foundations for a psychology of education* (pp. 87–136). New York: Erlbaum

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development.* Philadelphia, PA: Psychology Press.

Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 139–181). New York: Academic Press.

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology*. (3rd ed., pp. 404–434). Hoboken, NJ: Wiley. doi:10.1002/9780470479193.adlpsy001013

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., Vol. 1, pp. 404–434). Hoboken, NJ: John Wiley & Sons.

Eccles, J. S., & Templeton, J. (2002). Chapter 4: Extracurricular and other after-school activities for youth. *Review of Research in Education, 26*, 113–180. doi:10.3102/0091732x026001113

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53,* 109–132. doi:10.1146/annurev.psych.53.100901.135153

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

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*, 90. doi:10.1037/10254-034

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629–634. doi:10.1136/bmj.315.7109.629

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*, 169–189. doi:10.1207/s15326985ep3403_3

Elliot, A. J. (2005). Achievement Goals. In A. J., Elliot, & C. S., Dweck (Eds.). *Handbook of Competence and Motivation* (pp. 52–72). New York: Guildford Press.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*, 218.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*, 461–475. doi:10.1037/0022-3514.70.3.461

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80,* 501–519. doi:10.1037/0022-3514.80.3.501

Elliot, A. J., & Moller, A. C. (2003). Performance-approach goals: good or bad forms of regulation? *International Journal of Educational Research, 39*, 339–356. doi:10.1016/j.ijer.2004.06.003

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology, 82,* 804–818. doi:10.1037/0022-3514.82.5.804

Elliot, A. J., Dweck, C. S., & Yeager, D. S (2017). Competence and motivation: Theory and application. In Elliot, A. J., Dweck, C. S., & Yeager, D. S (Eds.). Competence and Motivation during Adolescence. *Handbook of Competence and Motivation: Theory and Application* (pp. 431-448). New York: Guildford Press.

Elliot, A. J., Hulleman, C. S. (2017). Achievement goals. In A. J., Elliot, C. S., Dweck, & D. S Yeager, (Eds.). *Handbook of competence and motivation: Theory and application* (2nd Ed., pp. 43–60). New York: Guildford Press.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3× 2 achievement goal model. *Journal of educational psychology, 103*, 632–648. doi: 10.1037/a0023952

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change. *Journal of Research on Technology in Education, 42,* 255–284. doi:10.1080/15391523.2010.10782551

Federal Statistical Office (2015). *Preliminary results of general and vocational school students: 2014-15 academic year.* [Schnellmeldungsergebnisse zu Schülerinnen und Schülern der allgemeinbildenden und beruflichen Schulen-Schuljahr 2014/15]. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/ BildungForschungKultur/ Schulen/SchnellmeldungSchueler5211003138004.html

Fenzel, L. M. (2000). Prospective study of changes in global self-worth and strain during the transition to middle school. *The Journal of Early Adolescence, 20*, 93–116. doi:10.1177/0272431600020001005

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*, 120–128. doi:10.1037/a0024445

Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences, 19*, 499–505. doi:10.1016/j.lindif.2009.05.004

Ferla, J., Valcke, M., & Schuyten, G. (2010). Judgments of self-perceived academic competence and their differential impact on students' achievement motivation, learning approach, and academic performance. *European Journal of Psychology of Education, 25*, 519-536. doi:10.1007/s10212-010-0030-9

Fisher, Z., & Tipton, E. (2015). Robumeta: Robust variance meta-regression. R package version 1.6. Retrieved from https://cran.r-project.org/web/packages/robumeta/robumeta.pdf

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence, 20*, 507–537. doi:10.1111/j.1532-7795.2010.00645.x

Gehlbach, H. (2006). How changes in students' goal orientations relate to outcomes in social studies. *The Journal of Educational Research*, *99*, 358–370. doi:10.3200/joer.99.6.358-370

Gest, S. D., Molloy, L. E., & Ram, N. (2015). Intraindividual variability in self-representations in adolescence. In M. Diehl, K. Hooker, & M. J. Sliwinski (Eds.), *Handbook of intraindividual variability across the life-span* (pp. 123-141). New York, London: Routledge.

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Keller, U., Fischbach, A., & 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*, 188-205. doi: 10.1016/j.cedpsych.2014.04.002

Gogol, K., Brunner, M., Martin, R., Preckel, F., & Goetz, T. (2017). Affect and motivation within and between school subjects: Development and validation of an integrative structural model of academic self-concept, interest, and anxiety. *Contemporary Educational Psychology, 49*, 46–65. doi:10.1016/j.cedpsych.2016.11.003

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *Journal of Educational Research, 86*, 313-326. doi: 10.1080/00220671.1990.10885977

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82,* 525–538. doi:10.1037/0022-0663.82.3.525

Gottfried, A. E. (2019). Academic intrinsic motivation: Theory, assessment, and longitudinal research. In In A. J., Elliot (Ed.) *Advances in Motivation Science*. (Vol. 6). New York: Elsevier.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93,* 3–13. doi:10.1037/0022-0663.93.1.3

Götz, L., Lingel, K., Schneider, W., & Schneider, W. (2013a). *DEMAT 5+: Deutscher Mathematiktest für fünfte Klassen* [German math achievement test for 5th grades]. Göttingen: Hogrefe.

Götz, L., Lingel, K., Schneider, W., & Schneider, W. (2013b). *DEMAT 6+: Deutscher Mathematiktest für sechste Klassen* [German math achievement test for 6th grades]. Göttingen: Hogrefe.

Granleese, J., & Joseph, S. (1994). Reliability of the Harter self-perception profile for children and predictors of global self-worth. *The Journal of Genetic Psychology, 155,* 487–492. doi:10.1080/00221325.1994.9914796

Grassinger, R., Dickhäuser, O., & Dresel, M. (2019). Grundvorstellungen zur Motivation von Lernenden. In D., Urhahne, M., Dresel, & F., Fischer, (Eds.). (2019). *Psychologie für den Lehrberuf* (pp. 207-227). Berlin: Springer.

Greene, M. L., Way, N., & Pahl, K. (2006). Trajectories of perceived adult and peer discrimination among Black, Latino, and Asian American adolescents: Patterns and psychological correlates. *Developmental Psychology, 42,* 218–236. doi:10.1037/0012-1649.42.2.218

Grek, S. (2009). Governing by numbers: the PISA "effect" in Europe. *Journal of Education Policy*, 24, 23–37. doi:10.1080/02680930802412669

Grigorenko, E. L. (2016). Foreword. In Lipnevich, A. A., Preckel, F., & Roberts, R. D. (Eds.) *Psychological Skills and School Systems in the 21st Century* (pp. v). Basel, Switzerland: Springer International Publishing.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. Journal *of Educational Psychology, 95*, 124–136. doi:10.1037/0022-0663.95.1.124

Gutman, L. M. (2006). How student and parent goal orientations and classroom goal structures influence the math achievement of African Americans during the high school transition. *Contemporary Educational Psychology, 31*, 44–63. doi:10.1016/j.cedpsych.2005.01.004

Gutman, L. M., & Eccles, J. S. (2007). Stage-environment fit during adolescence: Trajectories of family relations and adolescent outcomes. *Developmental Psychology, 43*, 522–537. doi:10.1037/0012-1649.43.2.522

Hackett, G., & Betz, N. E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education, 20,* 261. doi:10.2307/749515

Hall, G. S. (1904). *Adolescence (Vols. 1 & 2).* New York: Appleton.

Hansford, B. C., & Hattie, J. A. (1982). The Relationship between Self and Achievement/Performance Measures. *Review of Educational Research, 52,* 123. doi:10.2307/1170275

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology, 73,* 1284–1295. doi:10.1037/0022-3514.73.6.1284

Harris, M. A., Gruenenfelder-Steiger, A. E., Ferrer, E., Donnellan, M. B., Allemand, M., Fend, H., … Trzesniewski, K. H. (2015). Do parents foster self-esteem? Testing the prospective impact of parent closeness on adolescent self-esteem. *Child Development, 86,* 995–1013. doi:10.1111/cdev.12356

Harter, S. (1985). Competence as a dimension of self-evaluation. Toward a comprehensive model of self-worth. In R. Leahy (Ed.), *The development of the self* (pp. 55–121). New York: Academic Press.

Harter, S. (1990a). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In R. J. Sternberg & J. Kolligian (Eds.), *Competence considered* (pp. 67-97). New Haven, CT: Yale University Press.

Harter, S. (1990b). Developmental differences in the nature of self-representations: Implications for the understanding, assessment, and treatment of maladaptive behavior. *Cognitive Therapy and Research, 14,* 113–142. doi:10.1007/bf01176205

Harter, S. (1999). *Distinguished contributions in psychology. The construction of the self: A developmental perspective.* New York, NY, US: Guilford Press.

Hattie, J. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37,* 449-481. doi: 10.1016/S0883-0355(03)00015-6

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

Heckhausen, H. (2006). *Motivation und Handeln.* Berlin, Heidelberg, New York: Springer.

Heckhausen, J., & Heckhausen, H. (2018). Motivation and action: Introduction and overview. In J. Heckhausen & H. Heckhausen (Eds.). *Motivation and action* (pp. 1-14). Springer, Cham.

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6,* 203–217. doi:10.1037/1082-989x.6.3.203

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426–445. doi:10.1037/1082-989x.9.4.426

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. doi:10.1002/jrsm.5

Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R) [Cognitive Abilities Test (CogAT; Thorndike, L. & Hagen, E., 1954-1986) - German adapted version]*. Göttingen: Beltz Test.

Hendrick, S. S., Hendrick, C., & Adler, N. L. (1988). Romantic relationships: Love, satisfaction, and staying together. *Journal of Personality and Social Psychology, 54*, 980–988. doi:10.1037/0022-3514.54.6.980

Heo, G., & Kim, T. (2015). Autoregressive cross-lagged modeling of the reciprocal longitudinal relationship between self-esteem and career maturity. *Journal of Career Development, 43,* 273–288. doi:10.1177/0894845315598002

Herrmann, J. R., Schmidt, I., Kessels, U., & Preckel, F. (2016). Big fish in big ponds: Contrast and assimilation effects on math and verbal self-concepts of students in within-school gifted tracks. *British Journal of Educational Psychology, 86,* 222-240. doi: 10.1111/bjep.12100

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). *The MatchIt package* (Version 2.4–20) [Computer software]. Retrieved from http://gking.harvard.edu/matchit

Hong, Y., Chiu, C., Dweck, C. S., Lin, D. M.-S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*, 588–599. doi:10.1037/0022-3514.77.3.588

Hornstra, L., Majoor, M., & Peetsma, T. (2017). Achievement goal profiles and developments in effort and achievement in upper elementary school. *British Journal of Educational Psychology, 87*, 606–629. doi:10.1111/bjep.12167

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55. doi: 10.1080/10705519909540118

Huang, C. (2011). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review*, 23, 359–388. doi:10.1007/s10648-011-9155-x

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

Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 48–73. doi:10.1037/a0026223

Huguet, P., Dumas, F., Marsh, H. W., Regner, I., Wheeler, L., Suls, J., Seaton, M., & Nezlek, J. (2009). Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97,* 156-170. doi: 10.1037/a0015558

Hulleman, C. S., Barron, K. E., Kosovich, J. J., & Lazowski, R. A. (2016). Student motivation: Current theories, constructs, and interventions within an expectancy-value framework. In Lipnevich, A. A., Preckel, F., & Roberts, R. D. (Eds.) *Psychological skills and school systems in the 21st century* (pp. 241-278). Basel, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-28606-8_10

Hulleman, C. S., Durik, A.M., Schweigert, S., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*, 398-416. doi:10.1037/0022-0663.100.2.398

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin, 136,* 422–449. doi:10.1037/a0018947

Hunt. J. M. V. (1965). Intrinsic motivation and its role in psychological development. In D. Levine (Ed). *Nebraska Symposium on Motivation* Vol. 13. (pp. 189-282). Lincoln: University of Nebraska Press.

IBM Corp. (2015). *IBM SPSS Statistics for Windows, Version 23.0.* IBM Corp: New York.

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

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86,* 80–92. doi:10.1037/0021-9010.86.1.80

Kaplan, A., & Midgley, C. (1997). The effect of achievement goals: Does level of perceived academic competence make a difference? *Contemporary Educational Psychology, 22,* 415–435. doi:10.1006/ceps.1997.0943

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Thousand Oaks, CA: Sage

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (pp. 192–240). Lincoln, NE: University of Nebraska Press

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York, NY: Springer

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: methods and practices.* New York: Springer.

Korn, R. M., & Elliot, A. J. (2016). The 2 × 2 Standpoints Model of Achievement Goals. *Frontiers in Psychology, 7*, 1–12. doi:10.3389/fpsyg.2016.00742

Krapp, A., Schiefele, U., & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, 10*(2), 120-148.

Kulik, C.-L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19(3),* 415-428.

Kulik, J. A., & Kulik, C.-L. C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54,* 409-426. doi: 10.3102/00346543054003409

Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36,* 73-77. doi: 10.1177/001698629203600204

Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., … Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente. Materialien aus der Bildungsforschung (Band 72) [PISA 2000: Documentation of assessment tools. Instruments from educational research (Vol. 72)]*. Berlin: Max-Planck-Institut für Bildungsforschung.

Lazarides, R., & Raufelder, D. (2017). Longitudinal effects of student-perceived classroom support on motivation – A latent change model. *Frontiers in Psychology, 8,* 1–11. doi:10.3389/fpsyg.2017.00417

Lazarides, R., Viljaranta, J., Ranta, M., & Salmela-Aro, K. (2017). Adolescents' preparedness and motivation across the transition to post-comprehensive education. *Contemporary Educational Psychology, 49,* 151–159. doi:10.1016/j.cedpsych.2017.01.008

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation Interventions in Education. *Review of Educational Research*, 86, 602–640. doi:10.3102/0034654315617832

Lee, V. E., & Smith, J. B. (1997). High School Size: Which Works Best and for Whom? *Educational Evaluation and Policy Analysis, 19*, 205. doi:10.2307/1164463

Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play*."* *Intrinsic and Extrinsic Motivation,* 257–307. doi:10.1016/b978-012619070-0/50032-5

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology*, *97,* 184–196.

Li, J. (2005). Mind or Virtue. *Current Directions in Psychological Science*, *14,* 190–194. doi:10.1111/j.0963-7214.2005.00362.x

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development, 31,* 357–365. doi:10.1177/0165025407077757

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal, 13*, 59–72. doi:10.1207/s15328007sem1301_3

Lochbaum, M., & Gottardy, J. (2015). A meta-analytic review of the approach-avoidance achievement goals and performance relationships in the sport psychology literature. *Journal of Sport and Health Science, 4*, 164–173. doi:10.1016/j.jshs.2013.12.004

Lou, Y., Abrami, P. C., & Spence, J. C. (2000). Effects of within-class grouping on student achievement: An exploratory model. *The Journal of Educational Research, 94*, 101-112. doi: 10.1080/00220670009598748

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist, 55*, 137-150. doi: 10.1037/0003-066X.55.1.137

Madjar, N., & Chohat, R. (2016). Will I succeed in middle school? A longitudinal analysis of self-efficacy in school transitions in relation to goal structures and engagement. *Educational Psychology, 37*, 680–694. doi:10.1080/01443410.2016.1179265

Maehr, M. L., & Midgley, C. (1996). *Lives in context. Transforming school cultures.* Boulder, CO, US: Westview Press.

Makel, M. C., Lee, S. Y., Olszewski-Kubilius, P., & Putallaz, M. (2012). Changing the pond, not the fish: Following high ability students across different educational environments. *Journal of Educational Psychology, 104,* 778-792. doi: 10.1037/a0027558

Marsh, H. W. (1986). Verbal and Math Self-Concepts: An Internal/External Frame of Reference Model. *American Educational Research Journal, 23,* 129. doi:10.2307/1163048

Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77–172. doi:10.1007/bf01322177

Marsh, H. W. (1990). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: An interim test manual and a research monograph*. San Antonio, TX: The Psychological Corporation.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133–163. doi: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. *The British Journal of Educational Psychology*, *81*, 59-77. doi: 10.1348/000709910X503501.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be relatively large fish in a small pond even if you don’t learn to swim as well? *Journal of Personality and Social Psychology, 47,* 213-231. doi: 10.1037/0022-3514.47.1.213

Marsh, H. W., & Rowe, K. J. (1996). The negative effects of school-average ability on academic self-concept: An application of multilevel modeling. *Australian Journal of Education, 40(1),* 65-87.

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

Marsh, H. W., & Yeung, A. S. (1998). Top-down, bottom-up, and horizontal models: The direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology, 75,* 509–527. doi:10.1037/0022-3514.75.2.509

Marsh, H. W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: A multicohort–multioccasion study. *Child development, 69,* 1030-1053. doi:10.2307/1132361

Marsh, H. W., Relich, J. D., & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based upon the SDQ. *Journal of Personality and Social Psychology*, *45*, 173–187. doi:10.1037/0022-3514.45.1.173

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

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development*, *76*, 397–416. doi:10.1111/j.1467-8624.2005.00853.x

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

Martin, A. J. (2008). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology, 33*, 239–269. doi:10.1016/j.cedpsych.2006.11.003

Martin, A. J., & Elliot, A. J. (2015). The role of personal best (PB) and dichotomous achievement goals in students' academic motivation and engagement: a longitudinal investigation. *Educational Psychology, 36,* 1285–1302. doi:10.1080/01443410.2015.1093606

McArdle, J. J., & Nesselroade, J. R. (2003). *Growth curve analysis in contemporary psychological research*. Handbook of Psychology, Part Three, 18, 447-480. doi: 10.1002/0471264385.wei0218a

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9,* 403-425. doi: 10.1037/1082-989X.9.4.403

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The achievement motive. New York: Appleton-Century-Crofts.

Meece, J. L., & Miller, S. D. (2001). A longitudinal analysis of elementary school students' achievement goals in literacy activities. *Contemporary Educational Psychology, 26,* 454–480. doi:10.1006/ceps.2000.1071

Meece, J. L., Blumenfeld, P. C., Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80,* 514-523. *doi:http://dx.doi.org/10.1037/0022-0663.80.4.514*

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58,* 525-543. doi: 10.1007/BF02294825

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In L. M. Collins, & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 203-240). Washington, DC: American Psychological Association.

Middleton, M. J., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology, 89,* 710–718. doi:10.1037/0022-0663.89.4.710

Midgley, C., & Feldlaufer, H. (1987). Students' and teachers' decision-making fit before and after the transition to junior high school. *The Journal of Early Adolescence, 7,* 225–241. doi: 10.1177/0272431687072009

Midgley, C., Anderman, E., & Hicks, L. (1995). Differences between elementary and middle school teachers and students: A goal theory approach. *The Journal of Early Adolescence, 15,* 90–113. doi:10.1177/0272431695015001006

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., … Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS).* Ann Arbor: University of Michigan. Retrieved from http://www.umich.edu/~pals/PALS%202000_V12Word97.pdf

Midgley, C., Middleton, M.J., Gheen, M.H., & Kumar, R. (2002). Stage-environment fitrevisited: A goal theory approach to examining school transitions. In C. Midgley (Ed.), *Goals, goals structures, and patterns of adaptive learning.* (pp. 109–142). Mahwah, NJ: Erlbaum.

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

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. doi:10.1037/1082-989x.7.1.105

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38,* 30–38. doi:10.1037/0022-0167.38.1.30

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,* 1035–1070. doi:10.1037/a0028324

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25*, 3–53. doi:10.1006/ceps.1999.1019

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472-489. doi: 10.1037/0033-295X.110.3.472

Musu-Gillette, L. E., Wigfield, A., Harring, J. R., & Eccles, J. S. (2015). Trajectories of change in students' self-concepts of ability and values in math and college major choice. *Educational Research and Evaluation*, 21, 343–370. doi:10.1080/13803611.2015.105716

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus Version 7.4 [statistical software]*. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus user's guide. Eighth edition.* Los Angeles, CA: Muthén & Muthén

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,* 1033-1053. doi: 10.1037/a0027697

Nagy, G., Watt, H. M., 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,* 482-506. doi: 10.1111/j.1532-7795.2010.00644.x

Neel, C. G.-O., & Fuligni, A. (2012). A longitudinal study of school belonging and academic motivation across high school. *Child Development, 84*, 678–692. doi:10.1111/j.1467-8624.2012.01862.x

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. New York, NY: Routledge/Taylor & Francis Group.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. New York, NY, US: Routledge/Taylor & Francis Group.

Neyer, F. J., & Asendorpf, J. B. (2001). Personality–relationship transaction in young adulthood. *Journal of Personality and Social Psychology, 81*, 1190–1204. doi:10.1037/0022-3514.81.6.1190

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346. doi:10.1037/0033-295x.91.3.328

Nicholls, J. G. (1989). *The competitive ethos and democratic education.* Cambridge: Harvard University Press.

Nicholls, J. G., Patashnick, M., & Nolen, S. B. (1985). Adolescents' theories of education. *Journal of Educational Psychology*, *77*, 683.

Nicholls, J. G., Patashnick, M., Cheung, P. C., Thorkildsen, T. A., & Lauer, J. M. (1989). Can achievement motivation theory succeed with only one conception of success? In F. Halisch & J. Van den Beroken (Eds.), *Competence considered* (pp. 190–204). The Netherlands: Swets & Zeitlinger.

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

Niepel, C., Brunner, M., & Preckel, F. (2014b). Achievement goals, academic self-concept, and school grades in mathematics: Longitudinal reciprocal relations in above average ability secondary school students. *Contemporary Educational Psychology*, *39*, 301–313. doi:10.1016/j.cedpsych.2014.07.002

Nunn, G. D., & Jantz, P. B. (2009). Factors within response to intervention implementation training associated with teacher efficacy beliefs. *Education*, *129*, 599-608.

O'Connor, B. P., & Vallerand, R. J. (1998). Psychological adjustment variables as predictors of mortality among nursing home residents. *Psychology and Aging*, *13*, 368–374. doi:10.1037/0882-7974.13.3.368

O'Keefe, P. A., Ben-Eliyahu, A., & Linnenbrink-Garcia, L. (2012). Shaping achievement goal orientations in a mastery-structured environment and concomitant changes in related contingencies of self-worth. *Motivation and Emotion, 37*, 50–64. doi:10.1007/s11031-012-9293-6

Oakes, J. (2008). Keeping track: Structuring equality and inequality in an era of accountability. *Teachers College Record, 110*(3), 700-712.

O'Mara, A. J., Green, J., & Marsh, H. W. (2006). Administering self-concept interventions in schools: No training necessary? A meta-analysis. *International Education Journal, 7*, 524-533.

Orth, U., Erol, R. Y., & Luciano, E. C. (2018). Development of self-esteem from age 4 to 94 years: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144,* 1045–1080. doi:10.1037/bul0000161

Ortner, T., van de Vijver, F., & van de Vijver, F. J. R. (2015). Assessment beyond self-reports. In Ortner, T., vam de Vijver, F. J. R. (Eds.) *Behavior-based assessment in psychology,* (pp. 3-11). Hogrefe.

Oud, J. H. L., & Delsing, M. J. M. H. (2010). Continuous time modeling of panel data by means of SEM. In K. van Montfort, A. Satorra, & J. H. L. Oud (Eds.), *Longitudinal research with latent variables* (pp. 201–244). Berlin, Heidelberg: Springer.

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

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology, 24,* 124–139. doi:10.1006/ceps.1998.0991

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology, 42,* 190–198. doi:10.1037/0022-0167.42.2.190

Parker, A. K. (2010). A kongitudinal investigation of young adolescents' self-concepts in the middle grades. *RMLE Online, 33,* 1–13. doi:10.1080/19404476.2010.11462073

Patrick, H., Ryan, A. M., & Pintrich, P. R. (1999). The differential impact of extrinsic and mastery goal orientations on males' and females' self-regulated learning. *Learning and Individual differences, 11*, 153–171. doi:10.1016/s1041-6080(00)80003-5

Paulick, I., Watermann, R., & Nückles, M. (2013). Achievement goals and school achievement: The transition to different school tracks in secondary school. *Contemporary Educational Psychology, 38,* 75–86. doi:10.1016/j.cedpsych.2012.10.003

Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology, 92,* 128–150. doi:10.1037/0021-9010.92.1.128

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A. C., Götz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M.

Prenzel (Ed.), *Studies on the educational quality of schools* (pp. 17–37). Münster, Germany: Waxmann.

Pekrun, R., Vom Hofe, R., Blum, W., Götz, T., Wartha, S., Frenzel, A., & Jullien, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA) [A project for the analysis of achievement development in mathematics]. *Untersuchungen zur Bildungsqualität von Schule*, 21–53. Retrieved from http://kops.uni-konstanz.de/bitstream/handle/123456789/13900/Pekrun_PALMA.pdf?sequence=2

Peterson, J. (2013). Gender differences in the identification of gifted youth and in gifted program participation: A meta-analysis. *Contemporary Educational Psychology, 38*, 342-348. doi: 10.1016/j.cedpsych.2013.07.002

Piaget. J. (1928a), *Judgement and reasoning of the child*. New York: Harcourt, Brace & World.

Piaget. J. (1928b), *The child's conception of the world*. New York: Harcourt, Brace & World.

Pintrich, P. R. (2000a). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92–104. doi:10.1006/ceps.1999.1017

Pintrich, P. R. (2000b). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544–555. doi:10.1037/0022-0663.92.3.544

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*, 33–40. doi:10.1037/0022-0663.82.1.33

Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12*, 531–544. doi:10.1177/014920638601200408

Preckel, F., & Brüll, M. (2010). The benefit of being a big fish in a big pond: contrast and assimilation effects on academic self-concept. *Learning and Individual Differences, 20,* 522-531. doi:10.1016/j.lindif.2009.12.007.

Preckel, F., Lipnevich, A. A., Boehme, K., Brandner, L., Georgi, K., Könen, T., … Roberts, R. D. (2013). Morningness-eveningness and educational outcomes: the lark has an advantage over the owl at high school. *British Journal of Educational Psychology*, 83, 114–134. doi:10.1111/j.2044-8279.2011.02059.x

Preckel, F., Lipnevich, A. A., Schneider, S., & Roberts, R. D. (2011). Chronotype, cognitive abilities, and academic achievement: A meta-analytic investigation. *Learning and Individual Differences, 21*, 483–492. doi:10.1016/j.lindif.2011.07.003

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., Scherrer, V., & Schneider, W. (2017). High ability grouping: Benefits for gifted students' achievement development without costs in academic self-concept. *Child Development. Advance online publication.* doi:10.1111/cdev.12996

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., & Schneider, W. (2017). A Test of the Reciprocal-Effects Model of Academic Achievement and Academic Self-Concept in Regular Classes and Special Classes for the Gifted. *Gifted Child Quarterly, 61*, 103–116. doi:10.1177/0016986216687824

Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. (2008). Female 'big fish' swimming against the tide: The 'big-fish-little-pond effect' and gender-ratio in special gifted classes. *Contemporary Educational Psychology, 33,* 78-96. doi: 10.1016/j.cedpsych.2006.08.001

Raudenbusch S.W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rawsthorne, L. J., & Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review, 3,* 326–344. doi:10.1207/s15327957pspr0304_3

Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology, 15*, 119–138. doi:10.1017/s0954579403000075

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, *13, 83–103*. doi:10.1007/bf00992956

Renninger, K. A. (1992). Individual interest and development: Implications for theory and practice. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 361-376). Hillsdale, NJ: Lawrence Erlbaum Associates.

Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C.Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 375–407). New York: Academic.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin, 138,* 353–387. doi: 10.1037/a0026838

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science, 29*, 1358-1369. doi:10.31234/osf.io/kymhp

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin, 130,* 261–288. doi:10.1037/0033-2909.130.2.261

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17,* 423–434. doi:10.1037/0882-7974.17.3.423

Robinson, A., Shore, B. M., & Enersen, D. L. (2007). *Best practices in gifted education. An evidenced-based guide.* Waco, TX: Prufrock Press.

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, et al., (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews, 11,* 429–438. doi:10.1016/j.smrv.2007.07.005

Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of educational psychology, 88,* 408–422. doi:10.1037//0022-0663.88.3.408

Rogers, K. B. (2007). Lessons learned about educating the gifted and talented: A synthesis of the research on educational practice. *Gifted Child Quarterly, 51,* 382-396. doi: 10.1177/0016986207306324

Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70(1),* 41-55.

Rosenberg, M. (1989). *Society and the adolescent self-im*. (Revised edition.). Middletown, CT: Wesleyen University Press.

Rosenberg, M., Schooler, C., & Schoenbach, C. (1989). Self-esteem and adolescent problems: modeling reciprocal effects. *American Sociological Review, 54*, 1004. doi:10.2307/2095720

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review, 60*, 141. doi:10.2307/2096350

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118–137. doi:10.1016/j.intell.2015.09.002

Roznowski, M., Reith, M. & Hong, S. (2000). A further look at youth intellectual giftedness and its correlates: Values, interests, performance, and behaviour. *Intelligence, 28,* 87-113. doi: 10.1016/S0160-2896(99)00032-X

Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement, over- and underestimation, and students' beliefs for Māori and Pākehā students. *Contemporary Educational Psychology, 47*, 72–83. doi:10.1016/j.cedpsych.2016.01.001

Rudolph, U., & Körner, A. (2009). *Motivationspsychologie kompakt*. Beltz PVU.

Ruzek, E. A., & Schenke, K. (2019). The tenuous link between classroom perceptions and motivation: A within-person longitudinal study. *Journal of Educational Psychology, 111*, 903–917. doi:10.1037/edu0000323

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761. doi:10.1037/0022-3514.57.5.749

Ryan, R. M., & Deci, E. L. (2000a). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78. doi:10.1037/0003-066x.55.1.68

Ryan, R. M., & Deci, E. L. (2000b). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67. doi:10.1006/ceps.1999.1020

Scherer, R. (2013). Further evidence on the structural relationship between academic self-concept and self-efficacy: On the effects of domain specificity. *Learning and Individual Differences, 28,* 9–19. doi:10.1016/j.lindif.2013.09.008

Scherrer, V., & Preckel, F. (2019). Development of motivational variables and self-esteem during the school career: A meta-analysis of longitudinal studies. *Review of Educational Research, 89,* 211–258. doi: 10.3102/0034654318819127

Scherrer, V., Roberts, R., & Preckel, F. (2016). Parents’ assessment of circadian preference in elementary school-aged children: Validity and relations to educational outcomes. *Chronobiology International, 33*, 1188–1207. doi:10.1080/07420528.2016.1210618

Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Verlag für Psychologie, Hogrefe.

Schmidt, I., Brunner, M., Keller, L., Scherrer, V., Wollschläger, R., Baudson, T. G., & Preckel, F. (2017). Profile formation of academic self-concept in elementary school students in grades 1 to 4. *PLoS ONE* 12. doi:10.1371/journal.pone.0177854

Schnabel, K. U., Alfeld, C., Eccles, J. S., Köller, O., & Baumert, J. (2002). Parental influence on students' educational choices in the United States and Germany: Different ramifications—Same effect? *Journal of Vocational Behavior*, *60,* 178-198. doi:10.1006/jvbe.2001.1863

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*, 565–600. doi:10.1037/bul0000098

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. Journal of Educational Psychology, 81(3), 306–312. doi:10.1037/0022-0663.81.3.306

Schneider, W., Preckel, F., & Stumpf, E. (Eds.) (2014). *Hochbegabtenförderung in der Sekundarstufe* (*Fostering the of gifted and talented students in secendary school).* Karg Hefte – Beiträge zur Begabtenförderung und Begabungsforschung, Heft 7. Frankfurt, a. M.: Karg Stiftung.

Schneider, W., Stumpf, E., Preckel, F., & Ziegler, A. (2012). *Abschlussbericht 2012: Projekt zur Evaluation der Begabtenklassen in Bayern und Baden-Württemberg Laufzeit 2008 – 2012 (PULSS I)* [A project for the evaluation of special classes for gifted students in Bavaria and Baden-Wuerttemberg]. Unpublished report. University of Würzburg, University of Trier, and University of Nürnberg.

Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, 59, 84–95. doi:10.1016/j.intell.2016.08.00

Schunk, D. (1991). Self-Efficacy and Academic Motivation. *Educational Psychologist*, 26, 207–231. doi:10.1207/s15326985ep2603&4_2

Schwinger, M., & Wild, E. (2012). Prevalence, stability, and functionality of achievement goal profiles in mathematics from third to seventh grade. *Contemporary Educational Psychology, 37,* 1–13. doi:10.1016/j.cedpsych.2011.08.001

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101,* 403-419. doi: 10.1037/a0013838

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,* 390-433. doi:10.3102/0002831209350493

Seaton, M., Marsh, H. W., Dumas, F., Huguet, P., Monteil, J.-M., Régner, I., Blanton, H., Buunk, A. P., Gibbons, F. X., Kuyper, H., Suls, J., & Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47,* 73-103. doi: 10.1348/014466607X202309.

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

Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., & Yeung, A. S. (2013). The reciprocal relations between self-concept, motivation and achievement: juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educational Psychology, 34,* 49–72. doi:10.1080/01443410.2013.825232

Sebanc, A. M., Guimond, A. B., & Lutgen, J. (2014). Transactional relationships between latinos' friendship quality and academic achievement during the transition to middle school. *The Journal of Early Adolescence, 36,* 108–138. doi:10.1177/0272431614556347

Senko, C., & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-analytic evidence from multiple educational outcomes. *Journal of Educational Psychology, 109,* 574–598. doi:10.1037/edu0000160

Shapka, J. D., & Keating, D. P. (2005). Structure and Change in Self-Concept During Adolescence. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 37,* 83–96. doi:10.1037/h0087247

Shavelson, R. J., & Bolus, R. (1982). Self concept: The interplay of theory and methods. *Journal of Educational Psychology, 74,* 3–17. doi:10.1037/0022-0663.74.1.3

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-Concept: Validation of Construct Interpretations. *Review of Educational Research, 46,* 407–441. doi:10.2307/1170010

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology, 87,* 386–398. doi:10.1037/0022-0663.87.3.386

Shim, S. S., Ryan, A. M., & Anderson, C. J. (2008). Achievement goals and achievement during early adolescence: Examining time-varying predictor and outcome variables in growth-curve analysis. *Journal of Educational Psychology, 100,* 655–671. doi:10.1037/0022-0663.100.3.655

Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York: AldineTransaction.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417–453. doi:10.3102/00346543075003417

Skaalvik, E. M., & Rankin, R. J. (1995). A Test of the internal/external frame of reference model at different levels of math and verbal self-perception. *American Educational Research Journal, 32,* 161. doi:10.2307/1163217

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60(3),* 471-499.

Smeding, A., Darnon, C., Souchal, C., Toczek-Capelle, M.-C., & Butera, F. (2013). Reducing the socio-economic status achievement gap at university by promoting mastery-oriented assessment. *PLoS ONE, 8,* e71678. doi:10.1371/journal.pone.0071678

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin, 139,* 213–240. doi:10.1037/a0028931

Spinath, B. (2008). Das Feld der Motivationstheorien bleibt unübersichtlich. *Psychologische Rundschau, 59,* 167–168. doi:10.1026/0012-1924.59.3.167

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,* 87–102. doi: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*, 1555–1569. doi:10.1111/j.1467-8624.2008.01205.x

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

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research, 86,* 849-899. doi: 10.3102/0034654316675417

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology, 106,* 325–338. doi:10.1037/a0035133

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

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. *Oxford Bibliographies*. Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml

Stern, E. (2001). Intelligence, prior knowledge, and learning. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 11, pp. 7670-7674). Oxford: Elsevier Science.

Steyer, R., Eid, M., and Schwenkmezger, P. (1997). Modeling true intraindividual change: true change as a latent variable. *Methods Psychological Research Online 2,* 21–33. Retrieved from https://www.dgps.de/fachgruppen/methoden/mpr-online/issue2/art2/steyer.pdf

Stipek, D. J., & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the educational environment? *Journal of Educational Psychology, 80,* 352–356. doi:10.1037/0022-0663.80.3.352

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35,* 401–426. doi:10.1016/j.intell.2006.09.004

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. *Journal of Developmental and Life-Course Criminology, 2*, 85–112. doi:10.1007/s40865-016-0026-5

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. *Review of Educational Research, 83,* 357–385. doi:10.3102/0034654313483907

Thorndike, R., & Hagen, E. (1971). *Cognitive abilities test, verbal, nonverbal, and quantitative batteries: form 1, levels AH.* New York: Houghton Mifflin Co.

Tieso, C. (2005). The effects of grouping practices and curricular adjustments on achievement. *Journal for the Education of the Gifted, 29*, 60-89. doi: 10.1177/016235320502900104

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods, 4,* 169–187. doi:10.1002/jrsm.1070

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In Lipnevich, A. A., Preckel, F., & Roberts, R. D. (Eds.) *Psychological Skills and School Systems in the 21st Century* (pp. 187-214). Basel, Switzerland: Springer International Publishing.

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students' perceived standing of their class predicts academic self-concept. *Journal of Educational Psychology, 101,* 853-866. doi: 10.1037/a0016306

Trzesniewski, K. H., Donnellan, M. B., & Robins, R. W. (2003). Stability of self-esteem across the life span. *Journal of Personality and Social Psychology, 84,* 205–220. doi:10.1037/0022-3514.84.1.205

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68,* 202. doi:10.2307/1170754

Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2011). Stability and change in achievement goal orientations: A person-centered approach. *Contemporary Educational Psychology, 36,* 82–100. doi:10.1016/j.cedpsych.2010.08.002

Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *American Psychologist, 57,* 89–99. doi:10.1037/0003-066x.57.2.89

Urdan, T., & Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology, 28,* 524–551. doi:10.1016/s0361-476x(02)00060-7

Urhahne, D. (2008). Sieben Arten der Lernmotivation. *Psychologische Rundschau, 59,* 150–166. doi:10.1026/0012-1924.59.3.150

Van de gaer, Heidi Pustjens, Jan Van Damme, & Agnes De Munter. (2009). School engagement and language achievement: A longitudinal study of gender differences across secondary school. *Merrill-Palmer Quarterly, 55,* 373–405. doi:10.1353/mpq.0.0034

Van Yperen, N. W., Blaga, M., & Postmes, T. (2014). A Meta-analysis of self-reported achievement goals and nonself-report performance across three achievement domains (work, sports, and education). *PLoS ONE, 9,* doi:10.1371/journal.pone.0093594

Van Yperen, N. W., Blaga, M., & Postmes, T. (2015). A meta-analysis of the impact of situationally induced achievement goals on task performance. *Human Performance, 28,* 165–182. doi:10.1080/08959285.2015.1006772

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software, 36*(3), 1-48.

Vingilis, E., Wade, T. J., & Adlaf, E. (1998). What factors predict student self-rated physical health? *Journal of Adolescence, 21,* 83–97. doi:10.1006/jado.1997.0131

Vogl, K., & Preckel, F. (2014). Full-time ability grouping of gifted students: Impacts on social self-concept and school-related attitudes. *Gifted Child Quarterly, 58,* 51-68. doi: 10.1177/0016986213513795.

Wang, J., &Wang, X. (2012). *Structural Equation Modeling: Applications using Mplus: Methods and Applications*. West Sussex, UK: Higher Education Press.

Wang, Q., & Pomerantz, E. M. (2009). The motivational landscape of early adolescence in the United States and China: A longitudinal investigation. *Child Development, 80,* 1272–1287. doi:10.1111/j.1467-8624.2009.01331.x

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*, 427-450. doi:10.1007/BF02294627.

Watt, H. M. (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,* 1556-1574. doi:10.1111/j.1467-8624.2004.00757.x

Way, N., & Robinson, M. G. (2003). A longitudinal study of the effects of family, friends, and school experiences on the psychological adjustment of ethnic minority, low-SES adolescents. *Journal of Adolescent Research, 18,* 324–346. doi:10.1177/0743558403018004001

Wentzel, K. R. (1996). Social and academic motivation in middle school. *The Journal of Early Adolescence, 16*, 390–406. doi:10.1177/0272431696016004002

Wigfield, A., & Eccles, J. S. (2002). Students' motivation during the middle school years. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors in education* (pp. 159–184). San Diego, CA: Academic Press.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91-120). San Diego, California: Academic. doi:10.1016/b978-012750053-9/50006-1

Wigfield, A., & Wagner, A. L. (2005). Competence, motivation, and identity development during adolescence. In A. Elliott and C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 222-239). New York: Guilford Press.

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). Chapter 3: the development of children's motivation in school contexts. *Review of research in education, 23*, 73–118. doi: 10.3102/0091732X023001073

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., 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,* 451–469. doi:10.1037/0022-0663.89.3.451

Wirthwein, L., Sparfeldt, J. R., Pinquart, M., Wegerer, J., & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review, 10,* 66–89. doi:10.1016/j.edurev.2013.07.001

Wouters, S., Colpin, H., Van Damme, J., De Laet, S., Verschueren, K. (2013). Early adolescents' academic self-concept formation: Do classmates or friends matter most? *Learning & Individual Differences,27,* 193-200. doi: 10.1016/j.lindif.2013.09.002

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell: ACER Press.

Yeager, D. S., Lee, H. Y., & Dahl, R. E. (2017). Competence and motivation during adolescence. In Elliot, A. J., Dweck, C. S., & Yeager, D. S (Eds.). *Competence and Motivation during Adolescence. Handbook of Competence and Motivation: Theory and Application* (pp. 431-448). New York: Guildford Press.

Zeidner, M., & Schleyer, E. J. (1998). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology, 24,* 305-329. doi: 10.1006/ceps.1998.0985.

Zimmer-Gembeck, M. J., Chipuer, H. M., Hanisch, M., Creed, P. A., & McGregor, L. (2006). Relationships at school and stage-environment fit as resources for adolescent engagement and achievement. *Journal of Adolescence, 29,* 911–933. doi:10.1016/j.adolescence.2006.04.008

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 91,* 241–250. doi:10.1037/0022-0663.91.2.241

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe bzw. unerlaubte Hilfsmittel angefertigt, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Trier, den 3.09.2019