The Contributions of Knowledge and Stress to Academic Achievement:

A Meta-Analysis, a Latent Transition Analysis, and a Latent State-Trait Model

Dissertation

submitted in fulfillment of the requirements for the degree of
doctor rerum naturalium (Dr. rer. nat.) in Psychology
at the University of Trier

by

Maja Sophie Flaig

Supervised by

Prof. Dr. Michael Schneider
Prof. Dr. Jan Hofer

Submitted on

August 1, 2018

Place of Dissertation

Trier, Germany

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Professor Michael Schneider for his guidance throughout the time of my Ph.D. I felt blessed for having him as advisor of my thesis and related research because of his patience, continuous support, and immense knowledge. Besides my advisor, I would like to thank Professor Jan Hofer for being part of my thesis committee.

Furthermore, I would like to thank the following university staff, colleagues, and fellow doctoral students: Ursula Weiland for her administrative support, Dr. Peter Aymanns for his kind encouragements and interest in my work, Tobias Heltemes, Sevim Nuraydin, and Dr. Bianca Simonsmeier-Martin: Thank you for all your encouragements, help, clever insights, boundless support, the excellent food, and all the fun we had in and outside the office.
I am also very grateful to the following students: Michelle Hartl and Jana Schindler who supported the data collection and preparation of my research projects.

I consider myself very fortunate of counting the following persons to my fellow doctoral students and of course, friends: Sven Gunkel, Camilla Haake, Sophie Hegemann, Johanna Kranz, Hanna Merk, Simon Schulze, and Natalie Tröller: Thank you for your fantastic ideas, your cooperation, and your emotional support during the last months. *Whereas* many of you contributed to my work in one or the other way, I would like to give special thanks to Tobias Heltemes and Natalie Tröller, who did a wonderful job on proof reading *this* dissertation.

I want to thank my family and friends for being there for me. Mama and Papa, I thank you for always believing in me, and supporting me in every possible way. I am also grateful to my friends who reminded me to stay soft even when things got a bit hairy. In the truest sense of the word, this also applies to my dear writing buddy *Kala*, who was just purrfectly herself.

Finally, I am more than grateful to my partner: Tobi, I would like to say thank you for being extra patient, supportive, encouraging, and loving whenever I needed it.

Thanks for all your encouragements!

Abstract

Academic achievement is a central outcome in educational research, both in and outside higher education, has direct effects on individual's professional and financial prospects and a high individual and public return on investment. Theories comprise cognitive as well as non-cognitive influences on achievement. Two examples frequently investigated in empirical research are knowledge (as a cognitive determinant) and stress (as a non-cognitive determinant) of achievement. However, knowledge and stress are not stable, what raises questions as to how temporal dynamics in knowledge on the one hand and stress on the other contribute to achievement. To study these contributions in the present doctoral dissertation, I used meta-analysis, latent profile transition analysis, and latent state-trait analysis. The results support the idea of knowledge acquisition as a cumulative and long-term process that forms the basis for academic achievement and conceptual change as an important mechanism for the acquisition of knowledge in higher education. Moreover, the findings suggest that students' stress experiences in higher education are subject to stable, trait-like influences, as well as situational and/or interactional, state-like influences which are differentially related to achievement and health. The results imply that investigating the causal networks between knowledge, stress, and academic achievement is a promising strategy for better understanding academic achievement in higher education. For this purpose, future studies should use longitudinal designs, randomized controlled trials, and meta-analytical techniques. Potential practical applications include taking account of students' prior knowledge in higher education teaching and decreasing stress among higher education students.

Table of Contents

Introduction

Academic achievement is an important factor for the prosperity of individuals and societies. It has immediate as well as long-term benefits for individuals (Tata, 1999), and is a significant determinant of employment and professional success (OECD, 2000; Roth, BeVier, III, & Schippmann, 1996; Roth & Clarke, 1998), health (Furnée, Groot, & van Den Brink, 2008; Groot & Van Den Brink, 2007), and individual well-being (Witter, Okun, Stock, & Haring, 1984). Individual academic achievement also contributes to common welfare as it facilitates participation in society, improves self-directed and lifelong learning, and has a high public return on investment, particularly in higher education (OECD, 2000). Investing in a higher education degree will yield a public net gain that exceeds the costs by a factor of two to three (OECD, 2012, p. 171). And for the individual, completing a higher education program will be rewarded with a net gain of USD 110,000 to USD 162,000 in total, which is 60% larger than the gain from lower levels of education (OECD, 2012, p. 166). Completing a higher education program can also have indirect positive effects on individuals' lives as graduates tend to exercise more, are more involved in educational activities with their kin, and participate more actively in the social and political life (Ma, Pender, & Welch, 2016).

For many decades researchers have been interested in predicting and promoting students' academic achievement throughout all levels of education, which has yielded many insights into cognitive and non-cognitive determinants of academic achievement (Winne & Nesbit, 2010). For example, studying students' cognitive prerequisites for school achievement has laid the foundations for the development of standardized psychometric instruments which have been important instruments for selecting students for educational programs for more than a century (Neisser et al., 1996). Large-scale assessment studies have set international standards for education, provided insights into differences in academic achievement within and between countries and their determinants, and have inspired changes in educational policies (Breakspear, 2012; Grek, 2009; Neumann, Fischer, & Kauertz, 2010; Schwippert, 2007; Twist, 2012). And synthesizing techniques, such as meta-analysis, have yielded benchmarks for judging the effectiveness of educational innovations, and provided insights into other determinants of academic achievement, such as student characteristics and context factors (Hattie, 2009). In sum, prior research has established that academic achievement is a multi-faceted construct that is the end product of a complex causal network of instructional methods, learning processes, stable and dynamic factors. Yet, many of these factors have remained unstudied, so that it is unclear how they contribute to academic achievement.

The current dissertation focuses on knowledge as a cognitive determinant and stress as a non-cognitive determinant of academic achievement, particularly in higher education. These two determinants are important in practice and have been attracting widespread interest in educational research. However, little evidence is available on how knowledge and stress contribute to learning and achievement in higher education. The current dissertation closes this gap by providing insights from three empirical studies. It is divided into four main sections: the theoretical background (Chapters 1 to 7), the presentation of the three empirical studies (Chapters 8 to 10), the general discussion of the insights gained from these three studies, their methodological implications, practical implications, recommendations for future research (Chapter 11), and the conclusion (Chapter 12). Specifically, Study 1 (Chapter 8) provides insights into the cumulative and long-term nature of knowledge acquisition processes as it estimates the contribution of prior knowledge to later learning outcomes both in and outside higher education. Study 2 (Chapter 9) investigates the role of conceptual change processes for the acquisition of knowledge in higher education and their relation to academic achievement. Study 3 (Chapter 10) investigates whether higher education students' stress experiences can be described in terms of a stable, trait factor and situational, state factors, and whether state and trait stress are differentially related to academic achievement.

Overall, academic achievement is a significant resource for the prosperity of societies and has powerful impacts on individual lives (Furnée et al., 2008; OECD, 2000; Tata, 1999). Therefore, providing empirical evidence for predicting and promoting academic achievement is a vital aspect of educational research. Knowledge and stress are two powerful determinants of academic achievement that have attracted much attention in recent years, yet, many issues remain unsettled, in particular in the field of higher education. The current dissertation tries to close this gap by studying the contributions of knowledge and stress to academic achievement in higher education, providing insights from three empirical studies, and discussing their implications for theory and practice.

1. Higher Education

The higher education sector encompasses a remarkable variety of educational programs which are attracting the interest of an increasing numbers of learners. These programs follow the successful completion of secondary education and prepare graduates for entering the labor market (UNESCO, 2012, p. 47) with overall positive occupational prospects (OECD, 2012). Therefore, an enormous growth in the higher education sector has been reported recently. Since 1995, when only about 20% of the population of young adults (< 30 years old) graduated from higher education programs, graduation rates have increased to more than 50% in OECD countries on average (OECD, 2012, p. 62f.).

There is a large variety of higher education programs worldwide, yet different programs can be characterized by common educational aims. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) has therefore published an international standard for the classification of educational programs (International Standard Classification of Education, ISCED 2011; UNESCO, 2012). Higher education programs are organized on the ISCED levels 5 to 8 and include “short-cycle tertiary education, Bachelor’s or equivalent level, Master’s or equivalent level, and doctoral or equivalent level, respectively” (UNESCO, 2012, p. 46). Programs at ISCED level 5 are rather practically- and vocationally-focused, aim at conveying professional knowledge, skills and competencies, usually with a duration of less than three years, and prepare for direct entry into the labor market. In contrast, programs at ISCED level 6, often termed *Bachelor’s programs*, provide “qualifications for entry into advanced research [programs] and professions with high requirements in knowledge and skills” (OECD, 2012, p. 64). Bachelor’s programs aim at conveying “intermediate academic and/or professional knowledge, skills and competencies”, usually last three to four years, and lead to first degrees (UNESCO, 2012, p. 51).

Bachelor’s types of programs are among the most common programs in higher education worldwide, as it is estimated that about 39% of the under 30 year old adults in higher education are about to obtain a bachelor’s degree (OECD, 2012). According to the ISCED system, students who hold a first or bachelor’s degree are qualified to apply for programs that lead to second degrees, usually Master’s or equivalent programs (ISCED level 7), that “provide participants with advanced academic and/or professional knowledge, skills and competencies” with a duration of one to four years (UNESCO, 2012, p. 55). Students in master’s programs form the second largest population of young adults in higher education, as 15% of that population are expected to obtain a master’s degree in their respective study programs based on the 2010 statistics (OECD, 2012). Finally, ISCED level 8 represents degrees at the doctorate level, which aim at conveying more advanced academic education and research qualification with a usual duration of three years (UNESCO, 2012), accounting for about 2% of the graduations from higher education on average (OECD, 2012).

Because of the growing significance of higher education, the current dissertation focuses on academic achievement in higher education in particular. However, investigating academic achievement in general has a long tradition in educational research and a number of findings from studies on school achievement may pertain to academic achievement in higher education, too. Thus, in the following chapters, I recourse to evidence collected in *and* outside higher education in order to explain how knowledge and stress contribute to academic achievement. In the following chapter, I will give a brief overview of research on academic achievement, present a theoretical model of achievement development, and identify different indicators of academic achievement in higher education.

2. Achievement

2.1 Description of the Construct

The term *achievement* refers to the attainment of a specific target or goal (Nugent, 2013). In the academic context, these goals are related to the learning objectives of educational programs (UNESCO, 2012, p. 14). Therefore, academic achievement can be defined as "performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university [. . .]" (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014).

The study of academic achievement is a vital aspect of educational research and has led to an enormous literature base with the result that scanning all available evidence on academic achievement has been estimated to "require three decades of full-time work" (Winne & Nesbit, 2010, p. 654). As early as the beginning of the twentieth century, scholars have studied academic achievement, for example to evaluate the fairness of grading practices (Edgeworth, 1888), or to predict school achievement. For example, already in the early 1900s Alfred Binet constructed the first intelligence test to predict whether or not a child would succeed in school (Neisser et al., 1996). In recent decades, large-scale assessments in which school achievement is compared across countries have played a key role in educational research (Rindermann, 2007). These studies encompass assessments of students' reading, mathematics, and science literacy, and include the *Programme for International Student Assessment* (PISA; OECD, 2017), *Trends in International Mathematics and Science Study* (TIMSS; Mullis, Martin, Foy, & Arora, 2012) and the *Progress in International Reading Literacy Study* (PIRLS; Mullis, Martin, Foy, & Drucker, 2012). Insights from these studies have impacted on educational policies (e.g., Breakspear, 2012; Grek, 2009) in many nations and have led to an advancement of research and practice on teaching (Neumann, Fischer, & Kauertz, 2010; Schwippert, 2007; Twist, 2012).

Key questions in research on academic achievement concern the identification of effective teaching methods and other aspects that foster successful learning, such as characteristics of teaching contexts, instructors and students. A famous endeavor in educational research has been undertaken by John Hattie who collected and synthesized more than 800 meta-analyses relating to achievement throughout all educational levels, including primary, secondary, and tertiary education. From this vast literature base, he identified 138 influences on academic achievement relating to students, teachers, curricula, teaching methods, students' homes, and schools, and put them in a rank order according to the magnitude of their effect sizes. In general, most effects were positive, indicating an improvement in students' achievement when implementing specific teaching methods, curricula, and educational innovations. However, there was substantial variation in the magnitude of effect sizes, indicating that some teaching methods were much more effective to improve students' academic achievement than others with an average effect size of Cohen's d = 0.4.

For this reason, Hattie (2009) suggests to set a new bar for judging the effectiveness of educational innovations, that is, the average effect size from his synthesis of more than 800 meta-analysis. He argues:

> Setting the bar at zero is absurd. If we set the bar at zero and then ask that teachers and schools "improve achievement", we have set a very very low bar indeed. No wonder every teacher can claim that they are making a difference; no wonder we can find many answers as to how to enhance achievement; no wonder every child improves. […], it is easy to find programs that make a difference. Raising achievement that is enhancing learning beyond an effect size of d = 0.0 is so low a bar as to be dangerous and is most certainly misleading. (p.16)

In sum, predicting and promoting students' academic achievement has been a fundamental issue in educational research for many decades. Numerous studies have established that academic achievement is the result of many influences, including cognitive, and non-cognitive student characteristics, instructional methods, and educational policies. However, it would be misleading to conclude that changing any of these factors is effective to enhance academic achievement as some factors may have considerably stronger effects on achievement than others. In the following, I review research on academic achievement and its determinants in the field of higher education specifically.

2.2 Academic Achievement in Higher Education

Even though most studies on academic achievement focus on achievement in schools outside higher education, in recent years there has been great advance in research on academic achievement and teaching in higher education. Numerous handbooks draw on these findings to inform teaching practices in higher education (A. W. Bates & Poole, 2003; Bligh, 2002;

Fry, Ketteridge, & Marshall, 2008; Ramsden, 2003) and researchers have devoted much time and effort to summarize the findings from single studies in meta-analyses (e.g., Kuncel, Hezlett, & Ones, 2004; Poropat, 2009; Richardson, Abraham, & Bond, 2012; Robbins et al., 2004). A recent review transferred Hattie’s approach to the field of higher education and synthesized the results of 38 meta-analyses of achievement and its correlates in higher education (M. Schneider & Preckel, 2017). One hundred and five instruction-related and student-related correlates of achievement emerged from a data base including 3,330 effect sizes and almost 2 million learners. In line with Hattie’s findings, the results suggest an overall close and positive relation between instructional methods and student achievement. Specifically, it was found that teachers can foster academic achievement by paying attention to the implementation of teaching methods on a micro level and oftentimes small changes may be very effective, such as encouraging class attendance, using advance organizers, and asking open-ended questions. Besides academic achievement can be fostered by good teaching practices, several student characteristics were found to impact on learning and achievement in higher education. Meta-analyses found that cognitive abilities, personality dispositions, and self-regulation strategies are important predictors of academic achievement (e.g., Kuncel et al., 2004; Poropat, 2009; Richardson et al., 2012). When comparing the results of several meta-analyses, the strongest effect sizes were found for cognitive abilities and prior achievement, medium to strong effect sizes for learning strategies, and small effect sizes for students’ personality dispositions and background characteristics (M. Schneider & Preckel, 2017).

Two ongoing debates concern the roles of prior performance and stress for students’ academic achievement. Prior knowledge is considered to be the strongest predictor for learning outcomes in theory (Ausubel, 1968; Duncan et al., 2007; Kuncel, Hezlett, & Ones, 2001; Schuler, Funke, & Baron-Boldt, 1990) and in practice, as standardized achievement tests are used to select students who have the highest chances of being successful in higher education programs on the basis of the general and domain-specific skills, knowledge and expertise they have acquired during their educational careers (Kuncel & Hezlett, 2007). Stress has recently raised great concerns as a risk factor for low academic achievement, in particular inside higher education , where students often face multiple challenges, including, the transition to university, being in a different country, financial issues, examinations, high workload, and meeting deadlines (Robotham, 2008). Recently, stress levels of students in higher education programs have risen in different countries (A. K. Bates & Bourke, 2016; Bewick, Gill, Mulhearn, Barkham, & Hill, 2008; Lust, Ehlinger, & Golden, 2015; Techniker

Krankenkasse, 2015a). A possible explanation for the widespread interest in stress and prior knowledge may be that they affect learning processes in multiple ways. Both, prior knowledge (Dochy, 1990; Dochy, Segers, & Buehl, 1999; Ormrod, 2011, p. 231) and stress (Joëls, Pu, Wiegert, Oitzl, & Krugers, 2006; Lupien, Maheu, Tu, Fiocco, & Schramek, 2007; Vogel, Fernández, Joëls, & Schwabe, 2016) directly affect memory processes involved in learning, such as the encoding, storage and retrieval of information in long-term memory. Stress also indirectly affects learning processes as it alters the ways learners interact with their learning environment.

In summary, academic achievement is an active field of research with a plethora of empirical studies on school achievement and comparatively fewer studies conducted inside higher education. Prior research inside and outside higher education has identified many determinants of academic achievement, (e.g., Hattie, 2009; M. Schneider & Preckel, 2017). Two determinants of academic achievement that are currently attracting widespread interest in educational research and practice are prior knowledge and stress as they affect academic achievement in multiple ways. Thus, the current dissertation focuses on the effects of prior knowledge and stress on academic achievement in higher education. In the following subchapter, I introduce a theoretical model that is a heuristic framework to study academic achievement in and outside higher education. It represents the theoretical foundation of the current dissertation.

2.3 A Theoretical Model of Achievement Development

Theoretical frameworks of academic achievement organize the plethora of aspects affecting academic achievement identified in previous studies and are useful to develop and test hypotheses about learning and achievement in and outside higher education. In the following, I introduce the *L*earner-environment *I*nteraction in *K*nowledg*E* construction (LIKE) Framework, (see Figure 1) as the theoretical foundation of my dissertation. The LIKE Framework, which was developed as part of Study 1 is an overarching framework integrating findings from empirical research and helps organizing the many aspects rooted in learners and instruction that have been associated with academic achievement in schools and higher education institutions, such as learning environments, teaching methods, students' motivation and abilities. It is in line with theoretical frameworks that have been published earlier as regards to structure and contents (e.g., Biggs, 1993; Kuncel et al., 2004; Schrader & Helmke, 2015; Ullén, Hambrick, & Mosing, 2015). Yet, LIKE goes beyond these frameworks as it integrates many different aspects of academic achievement and clarifies in which ways they affect learning.

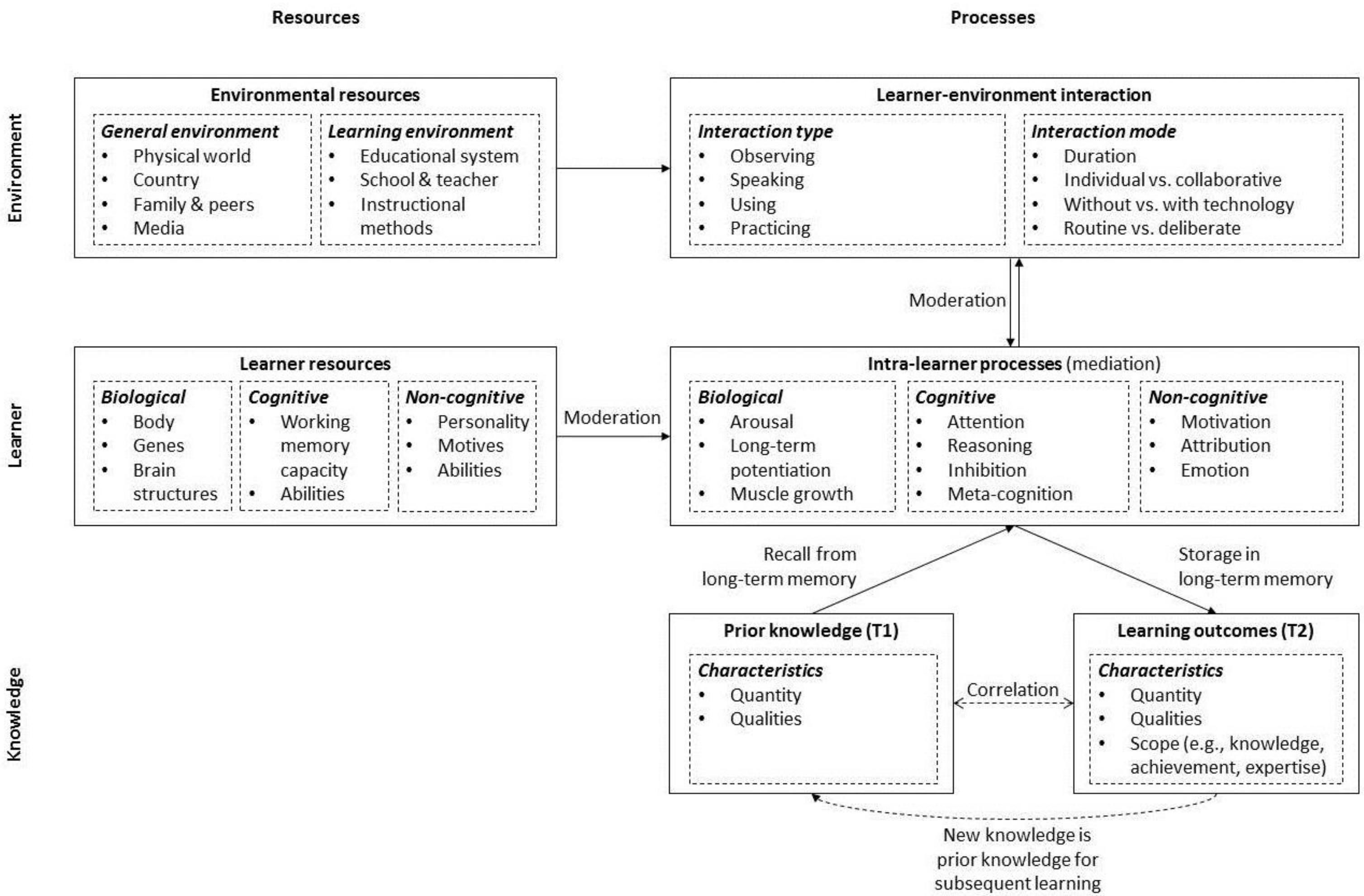

Figure 1. Learner-environment Interaction in KnowledgE construction (LIKE) Framework. The effect of prior knowledge on learning outcomes is mediated by intra-learner processes that are directly moderated by learner resources and the learner-environment interaction and indirectly moderated by environmental resources.

LIKE has several key characteristics that are described in the following in more detail. First, in line with Steinmayr et al. (2014), academic achievement is conceptualized as the learning outcomes (Figure 1, bottom right) indicating to which degree students have achieved the learning objectives of the units of instruction, courses, or educational programs they have attended (see Chapter 2.4 for a description of the different indicators of academic achievement in higher education). Second, in line with Kuncel and colleagues (2004), learning outcomes are conceptualized as a function of (prior) declarative and procedural knowledge on the one hand and other achievement-relevant factors on the other hand. This takes into account that academic tasks in higher education are complex and require not only domain-specific knowledge, but also a set of academic and non-academic skills and competencies. For example, when writing term papers, preparing oral presentations or studying for examinations, higher education students need to be persistent and regulate conflicting behaviors in order to achieve optimal results (Kuncel et al., 2004). Third, academic achievement is conceptualized as an interaction between stable factors and dynamic processes, that is, an interaction between deliberate practice, personality dispositions, and cognitive abilities as described in the Gene-Environment Interaction model of expert performance by Ullén et al. (2015). In the LIKE framework in Figure 1, stable resources are depicted on the left, while dynamic processes are depicted on the right side. For example, cognitive learner resources, such as working memory capacity, impact on intra-learner processes, such as reasoning. Fourth, academic achievement is conceptualized as result of learner resources, environmental resources, instruction, and learning processes, in line with theoretical frameworks that focus on the teaching-learning context in educational systems (e.g., Biggs, 1993; Schrader & Helmke, 2015). In the LIKE framework learner and environmental resources are organized separately at the bottom and top parts of Figure 1. Fifth, LIKE does not only specify important determinants of academic achievement but also how these determinants affect academic achievement. Prior knowledge is hypothesized to affect cognitive, non-cognitive, and biological intra-learner processes, such as attention. In turn, these intra-learner processes can be facilitated or hampered by learner resources, such as personality dispositions, as well as the interactions between learners and their learning environments, such as individual or collaborative learning. In statistical terms, academic achievement is hypothesized to be a moderated mediation (Preacher, Rucker, & Hayes, 2007) between prior knowledge, intra-learner processes, learner resources and the interplay between learners and their learning environment.

In summary, LIKE organizes many influences on academic achievement in an overarching framework and specifies how these factors, i.e., environmental resources, learner resources and learning processes interact in the development of knowledge and academic achievement. In the following subchapter, I present the different indicators that are currently used to assess academic achievement in higher education. Following this, in Chapters 3 to 5 I provide more detailed reviews of research on the relationships between prior knowledge and academic achievement, as well as stress and academic achievement.

2.4 Measures of Achievement

When studying academic achievement in higher education, there is a wide range of different indicators, such as domain-specific tests of knowledge, skills, and competencies, grades and grade point averages, degree attainment, standardized achievement tests, ratings of student performance, letters of recommendations, and others (e.g., Kuncel & Hezlett, 2007; Kuncel et al., 2001; Kuncel, Kochevar, & Ones, 2014; Kuncel, Wee, Serafin, & Hezlett, 2010). Some of these indicators have great practical importance, such as standardized achievement tests used for admission decisions, grades, and degrees. However, even though the attainment of a specific degree is an important goal in higher education, degree attainment has limitations for studying academic achievement. Degree attainment is a function of many behaviors, which can be subsumed under academic persistence, and other factors, some of which might even be beyond students‘ control (Kuncel et al., 2001), including contextual influences, social influences, social engagement and academic engagement (Robbins et al., 2004). Considering degree attainment, in the ISCED 2011 it is stated (UNESCO, 2012):

> [T]here is no direct relationship between education [programs] or qualifications and actual educational achievement. The education [programs] that an individual has participated in or has successfully completed are, at best, only an approximation of the skills, knowledge and competencies mastered at the time of completion. (p. 8)

Grades and standardized achievement tests on the other hand are directly related to the skills, knowledge, and competencies that are the objectives of educational programs and have extensively been used in educational research. This has several reasons, including the practical importance of grades and standardized achievement tests, their predictive value for many important life outcomes, and their universal significance in the educational system. Grades and grade point averages (GPA) provide a summative evaluation of the degree to which students have accomplished instructional goals that are part of a curriculum (Steinmayr et al., 2014). They are inherent features of most educational systems and institutions (Allen, 2005) and key variables in educational research (Kuncel, Credé, & Thomas, 2005; Robbins et al., 2004). Standardized achievement tests include tests of general and domain-specific cognitive abilities, knowledge, skills and competencies, and are used for admissions to undergraduate, graduate, and doctorate programs in various disciplines and many institutions all over the world (see Atkinson & Geiser, 2009; Kuncel & Hezlett, 2007, for a review). An example of a widely used set of standardized achievement tests is the Graduate Record Examinations (GRE®) General and the GRE Subject Tests. The GRE Tests assess undergraduate students' levels of general reasoning, critical thinking and writing skills, and their expertise in Biology, Chemistry, English Literature, Mathematics, Physics, and Psychology (Educational Testing Service, 2017, p. 3).

Standardized achievement tests have great practical significance, as they show substantial correlations with various indicators of graduate student performance, including GPA, exam scores, faculty ratings, and more distal outcomes, such as degree completion and research productivity, in law, medicine, natural science and social science programs as found by several meta-analyses (e.g., Kuncel, Credé, & Thomas, 2007; Kuncel & Hezlett, 2007; Kuncel et al., 2001; R. L. Linn & Hastings, 1984). Because of these high correlations, results from standardized tests increase the percentage of correct decisions in admission procedures by 5 to more than 30 %, depending on the base rate of potentially successful applicants and the selectiveness of the school. For example, in a review of nine meta-analyses by Kuncel and Hezlett (2007), standardized achievement tests accounted for a high amount of the variance in domain-specific academic knowledge (16 to 33%), considering these examinations can take place months or years after the admission decision. Despite this positive evaluation of criterion validity, some concerns have been raised regarding the fairness of standardized achievement tests. A meta-analysis found that women's grades in college and graduate schools are underpredicted based on standardized admissions tests. However, the effect was generally weak (0.24 grade points compared to males with the same test results), and weaker in graduate admission than in college admission tests (Fischer, Schult, & Hell, 2013), suggesting that test fairness increases with increasing expertise. Coaching or test preparation has been shown to yield enhance students' test performance, but the effects seem to be marginal (Kuncel & Hezlett, 2007), indicating rather high test-retest reliability.

Grades and GPAs are of utmost importance for students, institutions, and academic programs as they are often used in admission decisions (Tata, 1999), affect student motivation (Pulfrey, Buchs, & Butera, 2011), and ability appraisals (Jussim, 1989; Jussim & Eccles, 1992). Furthermore, they also have great utility for research, as they predict many important life outcomes, such as wage and health (Borghans, Golsteyn, Heckman, & Humphries, 2016), and can easily be assessed using records or even self-reports. Yet, self-reported grades should be used with caution as over- and under-reporting is common and self-reported GPAs can be biased, especially for students with low GPAs and ability. Inside higher education, self-reported grades are slightly more reliable (r = .90) than self-reported grades in high school (r=.82), as found in meta-analysis (Kuncel et al., 2005). Recently, much research outside higher education has focused on investigating whether grades are generally valid measures of students' knowledge, skills, and competencies. Some researchers have expressed strong doubts about the construct validity of grades (e.g., Allen, 2005), because teachers often do not only consider the quality of students' performance when assigning grades. As a consequence, grades become multidimensional measures of performance quality, aptitude, effort, compliance, and attitudes (Stiggins, 2001) which has been established in studies outside higher education (e.g., Bowers, 2011). Inside higher education, more standardized grading practices and a greater personal distance between teachers and students may reduce the multidimensionality of grades, still other factors rooted in the higher education system may flaw the usefulness of grades for educational research (Brookhart et al., 2016). For example, different grading standards and practices inside higher education limit the comparability of grades across institutions and countries. Moreover, higher education students nowadays receive more favorable grades than 25 years ago. Whether this is due to actual higher achievement or other factors, such as changing roles of students in the educational system, is still open for discussion.

Assessments of procedural and declarative knowledge, skills and competencies have been used to assess factual knowledge, correct concepts, and misconceptions in order to estimate academic achievement in science and social programs (e.g., Hieggelke, Maloney, Van Heuvelen, & O'Kuma, 2001; Nehm & Reilly, 2007; A. K. Taylor & Kowalski, 2004). Even though assessments of knowledge, skills, and competencies directly assess the degree to which students have accomplished learning goals in higher education programs (UNESCO, 2012), they are rarely used to study academic achievement for several reasons. First, the use of domain-specific knowledge tests is often hampered by the fact that published, standardized instruments are not available or impracticable in the current study setting (e.g., Nehm & Reilly, 2007). Thus, researchers who are interested in students' declarative knowledge in a specific field often have to develop the tests on their own or obtain scores from curricular tests or comprehensive examinations provided by higher education institutions. Second, the psychometric quality of these measures is often flawed or unknown, as they were either newly or not specifically developed for research purposes. Third, developing assessments of knowledge that meet high standards of psychometric quality is a complex process. An example of such a test stems from a study of Rolfhus and Ackerman (1999). The authors developed a battery of knowledge tests covering traditional academic domains, such as history, art, social sciences, and natural sciences on a higher education level. In an extensive validation process, they constructed 20 tests for different academic domains which they optimized until they arrived at a multifactorial test with high internal consistencies for four major knowledge domains. Finally, a reason for the rare use of assessments of domain-specific knowledge may be that many researchers are interested in less detailed, but more global summary measures of academic achievement provided by grades and GPA (Kuncel et al., 2005).

In sum, there are many different indicators of academic achievement in higher education which reflect different aspects of academic achievement. Whereas assessments of domain-specific knowledge are directly related to specific learning objectives of higher education programs, they have rarely been used in research. On the other hand, grades, GPA, and standardized achievement tests, which are summary measures of academic achievement often over several months or even years of instruction, have been used frequently in educational research. In the following, I present findings from research investigating the role of prior knowledge for learning and achievement.

3. Prior knowledge as Predictor of Achievement

Prior knowledge and prior performance are often seen as strong, if not the most important predictors for later performance (Ausubel, 1968; Duncan et al., 2007; Kuncel et al., 2001; Schuler et al., 1990), both in theory and practice. Yet, research has shown that prior knowledge which is necessarily incomplete and can even be incorrect, can yield facilitating effects (Committee on Developments in the Science of Learning, 2004; Dochy, 1990; Dochy et al., 1999; Hambrick & Engle, 2002; Ormrod, 2011; W. Schneider, 1993; Stern, 2001), as well as hindering effects for later learning (Bilalić, McLeod, & Gobet, 2008b; M. Fisher & Keil, 2016; Vosniadou, Vamvakoussi, & Skopeliti, 2008). Across various educational levels, facilitating effects of prior knowledge have been demonstrated in numerous studies and prior knowledge has often explained up to half of the variance in later learning outcomes (Dochy, De Rijdt, & Dyck, 2002). Inside higher education, students who were more knowledgeable before instruction also showed more advanced knowledge after instruction in psychology (Thompson & Zamboanga, 2003), physics (H. Hudson & Rottmann, 1981), introductory science courses (Hailikari, Katajavuori, & Lindblom-Ylanne, 2008), medicine (Kerfoot et al., 2011), biology (Lawson, 1983), and mathematics (Hailikari, Nevgi, & Komulainen, 2008). On the other hand, prior knowledge can also hinder learning, lead to overconfidence and underachievement (Bilalić et al., 2008b; M. Fisher & Keil, 2016; Vosniadou et al., 2008). Misconceptions can hinder learning as they lead to misinterpretations of new information and incorrect conclusions without learners' awareness (Vosniadou et al., 2008). Hindering effects because of students' misconceptions can be expected inside higher education, as it was found that students in various fields and at various levels, still hold many misconceptions which are pervasive (Bishop & Anderson, 1990; Brumby, 1984; Dagher & BouJaoude, 1997; Gregory & Ellis, 2009; Nehm & Reilly, 2007; Shtulman & Calabi, 2013; Sundberg & Dini, 1993) and related to lower academic achievement (Kuhle, Barber, & Bristol, 2009; McCutcheon, Hanson, Apperson, & Wynn, 1992).

Theoretical work has focused on explaining how prior knowledge facilitates or hinders learning by identifying the underlying cognitive processes of knowledge acquisition, i.e., how learners encode, store and retrieve information in their long-term memories (Dochy, 1990; Dochy et al., 1999; Ormrod, 2011, p. 231; see also "intra-learner processes", Figure 1). Theoretical positions have pointed to several different pathways trough which prior knowledge might affect learning, as prior knowledge might provide structures for interpretation of new information, enable elaboration, reduce working memory load, guide attention, and facilitate retrieval of relevant information (Dochy, 1990). Facilitating effects of prior knowledge are therefore principally expected when prior knowledge is available, accessible, correct, and well-structured. For example, research on differences in expert and novice performance suggests that expertise leads to better organized structures of knowledge in long-term memory which enable efficient chunking mechanisms that facilitate the encoding of new and the retrieval of old information (Committee on Developments in the Science of Learning, 2004). Second, prior knowledge affects the storage of new information as it provides possible links for new information to information already stored in long-term memory, enabling meaningful learning, for example through elaboration (Ormrod, 2011, p. 231). However, even correct prior knowledge can sometimes hinder learning as it has a strong effect on attention and thus, can make learners inflexible in their choice of problem solving strategies (Bilalić et al., 2008b).

Yet, empirical studies have also shown that the effect of prior knowledge on learning outcomes is subject to multiple other influencing factors, enrooted both in the learner and the learning environment. Therefore, in the LIKE framework (Figure 1), the effect of prior knowledge on later learning outcomes is depicted by means of a mediated moderation. According to this model, intra-learner processes that mediate the facilitating or hindering effects of prior knowledge on later learning are in turn moderated by learner resources and the interaction between learners and the learning environment. Thus, instruction can enhance learning by optimizing the accordance between learners' prior knowledge and the learning situation (Dochy, 1990), which may lead to using different instructional strategies for learners with high versus low prior knowledge. Furthermore, non-cognitive factors, such as motivational goals, interests, and self-efficacy beliefs have been shown to determine students' acquisition of conceptual knowledge in science (Pintrich, Marx, & Boyle, 1993). Inside higher education, research has shown that the effect of prior knowledge on learning outcomes highly depends on instructional strategies, learning tasks and learning activities. Students who are more knowledgeable before instruction may only perform better than their less knowledgeable peers when they can make use of their prior knowledge, for example when the problem solving tasks are well-structured, as shown in a study with undergraduate students in an Educational Technology course (Y. Lee & Nelson, 2005). This could also explain why in some studies in the field of higher education, the effects of prior knowledge on academic achievement were small (4-5% explained variance) as indicated by students' grades and GPAs (Griggs & Jackson, 1988; Plant, Ericsson, Hill, & Asberg, 2005; Xu, Villafane, & Lewis, 2013). Furthermore, some instructional strategies may compensate for limited prior knowledge. Results from studies conducted inside higher education suggest that learners with limited prior knowledge need more structure and feedback in order to achieve good results, for example, through collaborative and program-controlled learning activities (Dahlström, 2012; Gay, 1986; Krause, Stark, & Mandl, 2009).

In theory and practice prior knowledge is considered one of the strongest, if not *the* strongest predictor for later learning and achievement. Several cognitive mechanisms involved in learning can explain why prior knowledge has such a strong impact. Yet, the strength and the direction of the impact of prior knowledge on later learning is subject to a complex interplay between many factors: the amount and content of prior knowledge, characteristics of learners and their learning environment. As shown in the LIKE framework (Figure 1), this effect is embedded in a wider context of learning and instruction. Environmental and learner resources can affect the ways students (are able to) make use of their prior knowledge by affecting how students interact with their learning environments and how they process new information. This raises the question to which extent prior knowledge as a single factor affects learning, and under which circumstances this effect is positive, negative or zero. Study 1 of this dissertation aimed at closing this gap by meta-analyzing the overall net effect of prior knowledge on learning outcomes and investigating how characteristics pertaining to knowledge, the learner, and the learning environment moderate this relationship. These results allow to identify mediating and moderating mechanisms of the influence of prior knowledge on learning that can be studied in more depth in future studies.

Study 1 is in line with knowledge acquisition as a cumulative process and examines the bivariate relationship between prior knowledge and learning outcomes. Yet, in many cases prior knowledge is incompatible with new information and a fundamental restructuring of prior knowledge is necessary for learning (diSessa, Gillespie, & Esterly, 2004; Shtulman & Valcarcel, 2012; Vosniadou & Brewer, 1992). Consequently, there is a need to examine discontinuous learning processes and the multivariate relation between incorrect and correct prior knowledge and learning outcomes. Thus, a second study on prior knowledge was conducted in which (prior) knowledge was assessed in a multivariate way, and learning trajectories were investigated longitudinally. In the following chapter, the rationale for conducting Study 2 is explained in more detail.

4. Multivariate Trajectories of Knowledge Leading to Achievement

Research on conceptual change has shown that univariate assessments of knowledge are insufficient in analyzing how learners' initial, naïve ideas develop towards an academic understanding in various domains outside higher education (Edelsbrunner, Schalk, Schumacher, & Stern, 2015; Kainulainen, McMullen, & Lehtinen, 2017; McMullen, Laakkonen, Hannula-Sormunen, & Lehtinen, 2015; M. Schneider & Hardy, 2013). Specifically, univariate assessments of knowledge are limited as they can only take account of quantitative changes in learners' knowledge but not qualitative changes (Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2017). Qualitative changes in knowledge have been demonstrated for learners in various domains (e.g., Ioannides & Vosniadou, 2002; Rinne, Ye, & Jordan, 2017; M. Schneider & Hardy, 2013; Vosniadou & Brewer, 1992). For example in physics, young learners were found to show fundamental shifts in their explanatory frameworks of natural phenomena (Ioannides & Vosniadou, 2002; Vosniadou & Brewer, 1992). And also inside higher education, such fundamental shifts are to be expected, as research showed that students' naïve ideas are not in line with academic theories in various disciplines (Bishop & Anderson, 1990; Brumby, 1984; Dagher & BouJaoude, 1997; Gregory & Ellis, 2009; Kuhle et al., 2009; McCutcheon et al., 1992; Nehm & Reilly, 2007; Shtulman & Calabi, 2013; Sundberg & Dini, 1993). However, until today studies of multivariate trajectories of learners' knowledge were confined to educational situations outside higher education only.

From studies conducted outside higher education, several general characteristics of the development of conceptual knowledge can be outlined. First of all, in learning situations in which new ideas are incompatible with learners' initial ideas, successful learners arrive at restructuring their prior knowledge using various learning mechanisms, which are subsumed under the term *conceptual change* (M. Schneider, Vamvakoussi, & Van Dooren, 2012). In conceptual change research there is an ongoing debate about the nature of these initial ideas, which have been proposed to be coherent and theory-like by some researchers (e.g., Vosniadou et al., 2008), whereas others have stressed the diverse and fragmented nature of naïve knowledge (e.g., diSessa et al., 2004). There is no definite conclusion on the issue yet, as there is empirical evidence for coherence (e.g., Gómez, Benarroch, & Marín, 2006; Hatano & Inagaki, 1994; Vosniadou & Brewer, 1992; Wellman & Gelman, 1992) as well as fragmentation (e.g., diSessa et al., 2004; Harrison, Grayson, & Treagust, 1999; Izsak, 2005). Regardless of the theoretical perspective on coherence versus fragmentation, these studies have several implications for the study of conceptual knowledge development. First, conceptual knowledge consists of multiple elements, e.g., correct concepts, everyday observations, and misconceptions. Second, these elements are organized in greater structures of knowledge which can be coherent and theory-like as well as fragmented. Third, structures of conceptual knowledge can differ largely between persons as well as within persons at different points in time. Finally, learning leads to knowledge restructuring which yields learning trajectories that can involve sudden and discontinuous changes. In accounting for these characteristics, traditional approaches that assess knowledge in a univariate way and analyze the data by means of the general linear model are limited (Hickendorff et al., 2017).

Hence, when interested in studying multivariate trajectories of knowledge and accounting for the heterogeneity among learners, multivariate assessments of knowledge and mixture models are needed. One set of suitable statistical methods for longitudinal mixture modeling is latent class or latent profile transition analysis. (Hickendorff et al., 2017). These analyses allow researchers to identify developmental pathways of conceptual knowledge over time which can be interpreted as sequences of concepts that learners follow through during their learning process. By using latent transition analysis, researchers can determine the number of distinct profiles or classes of conceptual knowledge, the number of developmental pathways between these profiles, and the number of learners that are assigned to these profiles and pathways. To date, there is only a small number of studies applying this approach with data collected outside higher education only, investigating children's understanding of physics and mathematics concepts (Edelsbrunner et al., 2015; Kainulainen et al., 2017; McMullen et al., 2015; M. Schneider & Hardy, 2013). These studies used a multivariate approach to assess different qualities of conceptual knowledge, i.e., the degree to which learners endorsed scientific concepts, everyday conceptions, and misconceptions (e.g., M. Schneider & Hardy, 2013). The combination of the scores on these scales at each time point was then used to identify subgroups of learners with similar patterns of knowledge. Over time, the learners displayed changes in their patterns of knowledge which were modeled as learning trajectories between the subgroups of knowledge in latent class and latent profile transition analyses (Hickendorff et al., 2017). Three key findings emerged from these studies. First, despite there was considerable heterogeneity among learners in the overall samples, a relatively small number of four to six homogenous subgroups of learners could be identified. In these groups, learners showed similar patterns of conceptual knowledge. Second, a relative small number of learning trajectories accounted for the transitions between the different knowledge profiles over time. Third, these pathways were mostly in accordance with a developmental ordering from learners' initial conceptions to more advanced conceptions and scientific understanding of the learning domain and indicated conceptual change processes.

In summary, multivariate assessments of knowledge and longitudinal mixture models have yielded important insights about how conceptual knowledge develops over time from an initial, naïve understanding to a more advanced understanding of a content domain that is in line with academic theories and principles. Multivariate assessments of conceptual knowledge, i.e., separate scores for misconceptions, everyday conceptions, and scientific concepts, latent class or latent profile transition analysis allow to study discontinuous learning trajectories between prior knowledge and later learning outcomes. Yet, studies using this approach have been exclusively conducted outside higher education. This is a drawback as higher education students have shown to hold pervasive misconceptions in various fields, for example, mechanics (Clement, 1982), evolution (Shtulman & Calabi, 2013), and psychology (Hughes, Lyddy, & Lambe, 2013; A. K. Taylor & Kowalski, 2004) that can hamper academic understanding and achievement. Study 2 closes this gap by investigating multivariate trajectories of knowledge in psychology students using latent profile transition analysis, investigating whether these learning trajectories reflect conceptual change processes, and how different profiles of conceptual knowledge relate to students' grades.

Although prior knowledge is a strong predictor for academic achievement, it is not the only factor that is important in learning (see LIKE framework, Figure 1). In particular, non-cognitive predictors of learning, such as motivational beliefs (Pintrich et al., 1993), self-regulation strategies and personality characteristics (Poropat, 2009; Richardson et al., 2012), but also stress (Robotham & Julian, 2006), have recently been receiving increased interest in educational research. Study 3 was conducted in order to investigate how stress as a non-cognitive predictor affects academic achievement in higher education, as stress is supposed to affect learning in multiple ways via cognitive and non-cognitive processes.

5. Achievement and Stress

Stress has long been considered a natural part of studying in higher education (Whitman, 1985), before recently, concerning numbers of students with high levels of stress have put stress in higher education into a new perspective, (A. K. Bates & Bourke, 2016; Bewick et al., 2008; Lust et al., 2015; Techniker Krankenkasse, 2015a). Whereas high levels of stress can be adaptive in the short run, findings from endocrinological studies suggest that repeated elevations of stress hormones can lead to maladaptive responses (Juster, McEwen, & Lupien, 2010). For example, stressors that are present over a longer period of time, traumatic events, or a sequence of many smaller challenges can lead to suppressed immunity and a higher risk of infections (Segerstrom & Miller, 2004). In line with these findings, excessive stress was also associated with depression, anxiety, cardiovascular problems, obesity, and substance abuse (De Kloet, Joëls, & Holsboer, 2005; Deasy, Coughlan, Pironom, Jourdan, & Mannix-McNamara, 2014; Wang, Lesage, Schmitz, & Drapeau, 2008). Furthermore, stress negatively affects performance in the workforce (Gilboa, Shirom, Fried, & Cooper, 2008; L. W. Hunter & Thatcher, 2007) and in education, both inside higher education (Cotton, Dollard, & De Jonge, 2002; Pluut, Curşeu, & Ilies, 2015; Richardson et al., 2012) and outside higher education (Evans, Li, & Whipple, 2013; Evans &

Schamberg, 2009; Pungello, Kupersmidt, Burchinal, & Patterson, 1996; Schwartz, Gorman, Nakamoto, & Toblin, 2005).

In higher education, there are many demands that may cause stress in students, e.g., the transition to university, being in a different country, financial issues, examinations, high workload, and meeting deadlines (Robotham, 2008). While some students have no problems in adjusting to these demands, others experience high levels of stress. In line with this, a number of theoretical models has emphasized that stress is not the mere occurrence of demands and challenges but rather the individual's interpretation of these demands and their coping resources that causes stress, including the transactional model of stress (Folkman, 2011; Lazarus & Folkman, 1984), the job demands-resources model (Bakker & Demerouti, 2007), and the conservation of resources model (Hobfoll, 1989). Specifically, when individuals appraise the demands placed upon them as taxing, exceeding, or endangering their resources and well-being, they will experience stress (Folkman, 2011; Lazarus & Folkman, 1984). Stress can include different reactions on the physiological, emotional, behavioral, and cognitive level. For example, stress reactions include changes in cortisol secretion (Doane, Chen, Sladek, Van Lenten, & Granger, 2015), fear, anger, or guilt (Gadzella, 1994), alcohol consumption (Pritchard, Wilson, & Yamnitz, 2007), and impaired cognitive performance (Joëls et al., 2006). One pathway to reduce stress is therefore changing the individual's interpretation of the stressful situation. For example, interpreting demands as manageable challenges and adopting a "stress-is-enhancing mindset" can help students to decrease their levels of stress and achieve an optimal level of arousal (Crum, Salovey, & Achor, 2013).

Stress can affect academic achievement through multiple pathways, including direct effects on cognitive processes involved in learning such as memory encoding and retrieval, (Joëls et al., 2006; Lupien et al., 2007; Vogel et al., 2016), as well as indirect effects on (non-cognitive) learner resources and the learner-environment interaction (see LIKE framework, Figure 1). Students who experience mild stress in the situation of learning may show enhanced memory for the learning material (Joëls et al., 2006; Lupien et al., 2007), whereas students who experience moderate to high levels of stress before an exam usually show impaired retrieval from long-term memory, less flexible, creative thinking and problem-solving (Vogel & Schwabe, 2016). Stress may also hamper deeper conceptual understanding or conceptual change processes as stress disrupts the integration of new information into prior knowledge (Vogel & Schwabe, 2016) and deciding whether pieces of information pertain to naïve or academic theories (Merz, Dietsch, &

Schneider, 2016). Psychosocial stress experienced before learning may also impair transfer as shown in a laboratory study with healthy adults (Dandolo & Schwabe, 2016).

Indirect effects of stress on academic achievement can occur because stress draws on learner resources and alters the learner-environment-interaction. For example, infrequent class attendance which is closely related to students' academic achievement, (d=0.98; M. Schneider & Preckel, 2017), may result from poor time management, which has often been associated with stress (Abouserie, 1994; Misra & McKean, 2000; Nonis, Hudson, Logan, & Ford, 1998; Ross, Niebling, & Heckert, 1999). Stressed students report more negative social interactions (Edwards, Hershberger, Russell, & Markert, 2001), which could be both, an effect and cause of academic stress, e.g., due to competition (Gadzella, 1994). Yet, a positive social climate in classes, characterized by respect and interactions that promote learning, is one of the most important sources for academic achievement (M. Schneider & Preckel, 2017).

Acute and chronic stress have been found to negatively impact on health and performance in many populations, including higher education students. Stress-related impairments in learning and memory processes and low academic achievement in higher education can be explained by multiple pathways. Direct pathways include impairments of cognitive processes involved in knowledge acquisition and retrieval which have been shown to be hampered by acute stress. Indirect pathways include shortage of learner resources and obstructive interactions between learners and their learning environments which may occur more frequently under chronic stress. Unfortunately, studies inside higher education have not investigated the differential effects of acute and chronic stress on academic achievement. Thus, the aim of Study 3 was to investigate acute, state stress, and chronic, trait stress, separately by applying a latent state-trait model. The question was to which degree state and trait stress predict students' final semester grade point averages, and related outcomes of health and well-being in a sample of higher education students from various programs.

6. Methodological Aspects

Academic achievement is influenced by many factors (see LIKE framework, Figure 1), including characteristics of the instructions and instructors, student behaviors, characteristics, and contextual factors (M. Schneider & Preckel, 2017). In my dissertation I focus on two of these factors, i.e., prior knowledge and stress, which are of great practical and theoretical relevance in the field of higher education. Thus, as part my dissertation, three studies were conducted in order to study how prior knowledge and stress affect academic achievement in higher education. Three different methodological approaches were used which are introduced in the following three subchapters. Knowledge acquisition as a cumulative process was investigated in in Study 1 in which the effect of pretest knowledge on posttest knowledge and achievement was analyzed meta-analytically, synthesizing findings from a wide range of knowledge domains, types of instruction, and educational levels. In Study 2, discontinuous trajectories of knowledge were investigated in a single domain of knowledge inside higher education by applying a multivariate, mixture modeling approach in a longitudinal design. In Study 3, differential effects of experienced acute, state and chronic, trait stress on academic achievement were investigated using latent state-trait analyses in two data sets of higher education students. In the following, meta-analysis, latent profile transition analysis, and latent state-trait analysis are described in more detail.

6.1 Meta-Analysis

Meta-analysis is a powerful method to synthesize findings from the great quantity of studies conducted on issues in educational research. The term meta-analysis was proposed by Glass (1976, p. 3) to emphasize that meta-analysis is “the analysis of analyses”; a statistical procedure that synthesizes (inconsistent) findings from primary studies in order to answer a specific research question (Cheung, 2015). Compared to other synthesizing techniques, such as narrative reviews, meta-analyses can handle a much larger number of studies and provide an objective, quantitative synthesis of the effect sizes (Glass, 1976). Furthermore, meta-analysis can correct for flaws in the methodological quality in primary studies, e.g., measurement errors, dichotomization, and statistically account for heterogeneity in primary studies that arises from methodological and sample differences, e.g., different study designs, age of the participants in the sample (Viechtbauer, 2010).

Meta-analysis is an umbrella term for several statistical models, which make different assumptions of the observed effect sizes in the primary studies and the true effect sizes on the population level. Meta-analytical models mainly vary in three dimensions (Cheung, 2015): (1) whether one true effect size is assumed versus several true effect sizes are assumed on the population level, (2) whether one observed effect size is contributed versus several observed effect sizes are contributed by each of the primary studies, (3) whether one is interested in estimating average effect sizes versus modeling relations among the effect sizes. In the following, these three dimensions are described in more detail. The first dimension pertains to the nature of the true effect size on the population level, i.e., whether there is one effect size versus several. When applying a fixed effect model, it is assumed that there is only one true, or common effect size in the population, e.g., one true correlation ρ between prior knowledge and post-test knowledge with $SD_\rho = 0$ (Borenstein, Hedges, Higgins, & Rothstein, 2010; Schmidt & Hunter, 2015). Equation 1 describes the fixed effect model in mathematical terms. The observed effect y_i in the i-th study (i=1,…k) included in the meta-analysis can be decomposed into the true effect θ_i and the sampling error e_i (Viechtbauer, 2010):

$y_i = \theta_i + e_i$ (Equation 1)

Thus, as shown in Equation 1, each of the primary studies is an estimate of the true effect size and differences between the effect size estimates y_i are only due to measurement error. This is a relatively conservative assumption as the primary studies are usually not conducted using identical methods and samples. However, differences in the methodology may introduce some degree of heterogeneity among the true effects, e.g., the true correlation ρ between prior knowledge and posttest knowledge varies depending on the knowledge domain $SD_\rho > 0$ (Schmidt & Hunter, 2015). This heterogeneity can be treated as purely random using a random effects model (Viechtbauer, 2010) which is a generalization of the fixed effect model described in Equation 1. As shown below in Equation 2, in the random effects model the true effects are decomposed into a mean effect μ representing the average true effect, and $u_i \sim N(0, \tau^2)$ representing the heterogeneity variance (Cheung, 2015).

$\theta_i = \mu + u_i$ (Equation 2)

Thus, an advantage of the random effects model over the fixed effect model is that the heterogeneity among the true effects can be estimated whereas this cannot be done under a fixed effect model. When applying a random effects model, it is possible that the heterogeneity among the effect sizes is zero, i.e., the effect sizes are homogenous and $\tau^2 = 0$. In this case the true effect equals the average effect, $\mu = \theta_i$, which is consistent with a fixed effect model (Schmidt & Hunter, 2015; Viechtbauer, 2010).

The second dimension pertains to the number of effect sizes that is extracted from the primary studies. Meta-analyses that extract only one effect size per primary study are called *univariate,* whereas extracting more than one effect size leads to a *multivariate meta-analysis* (Cheung, 2015). For example, primary studies could report correlations between prior knowledge and posttest for several knowledge measures, measurement occasions, or treatment groups (Hedges, Tipton, & Johnson, 2010). In this case, the assumption of independence among the effect sizes is not tenable anymore and should be taken into account when computing the meta-analytical effect size. Different methods are available for several types of dependence (Cheung, 2015). The simplest case is that there is dependence among the effect sizes due to sampling error because several correlations (or other effect size estimates) are calculated using the same participants. This type of dependence can be handled in a fixed effect model by estimating the degree of dependence for each study. In the second case, dependence among the effect sizes occurs because the true effects are correlated on the population level. For example, studies with a larger true effect on one knowledge measure also show a larger true effect on other knowledge measures because the two measures capture related concepts. This type of dependence can only be treated by a random effects model in which the dependence is represented by the variance component of u_i (see Equations 2). In the third case, the effect sizes are dependent because they are nested hierarchically, however unlike in the other two cases, the degree of dependence is unknown. While in the first and second case, multivariate meta-analysis is suitable, this is not possible with unknown dependence in the third case (Cheung, 2014). In this case, a three-level model can be used in order to account for the dependence among the effect sizes (Konstantopoulos, 2011; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013).

The third dimension concerns the structure of the true effect, i.e., whether one is interested in simply estimating average effects or modeling the relationships among them. Meta-analytic structural equation model (MASEM) combines the techniques of meta-analysis and structural equation modeling and can be used to test hypotheses of the structure of effect sizes in contrast to simply estimating the average effect size as introduced above (Cheung, 2015). By the use of MASEM, the covariance matrices from the primary studies are pooled, the hypothesized model is fitted to the data and the model fits are estimated. This technique can be very helpful in explaining different findings from single studies, and finding the best fitting theoretical model.

Conducting a meta-analysis involves several stages from formulating a research question to synthesizing and reporting the findings (Cooper, Hedges, & Valentine, 2009). At the initial stage of a meta-analysis, the researcher formulates a problem or question, which is either the (1) description of an event or association between variables, or the (2) examination of change in one or more variables over time. Based on this research question, the constructs of interest are defined and in the second stage, the literature research is performed. The literature search can include published, peer-reviewed research only, but often it is advisable to also include unpublished research, i.e., *grey literature*, to avoid publication bias in the direction and magnitude of the effect sizes. In the third stage, the primary studies that have been identified in the literature research are coded based on the rules defined in the coding sheet and coding manual. Any study feature that is not coded in this stage cannot be considered in the analysis. There are no definite standards for which study features should be coded, but it is recommended that not only information pertaining to the research question is coded but also information pertaining to the quality of the primary studies, e.g., the internal validity, and validity of the constructs used in the study, such as study design and reliability measures. In the next stage the actual meta-analysis is conducted by synthesizing the findings from the primary studies in a univariate, multivariate or three-level meta-analytical model (Cheung, 2015). Bias in the effect size estimates can be corrected before the meta-analytical integration, for example, to correct for dichotomization, unreliability, and range restriction (Schmidt & Hunter, 2015). Next, the results are synthesized and interpreted. At this stage it is also important to analyze whether and to which degree there is publication bias among the effect sizes that are integrated in the meta-analysis and how it affects the meta-analytical results. Popular methods to analyze publication bias are analyzing findings from published and unpublished research separately and comparing the resulting effect sizes, plotting the effect sizes against sampling size or sampling error in a funnel plot, and estimating the number of missing effect sizes using the trim-and-fill method (Schmidt & Hunter, 2015). Finally, the results from the latter stage should be summarized and presented using tables and or graphics, for example, using forest plots (Cooper et al., 2009).

In sum, meta-analysis is a research tool that has many benefits over other synthesizing techniques such as narrative reviews as it provides systematic, objective, and quantitative synthesis answers to research questions. Compared to primary studies, meta-analyses mostly yield more accurate and better generalizable estimates of the effect sizes (Cheung, 2015; Schmidt & Hunter, 2015). By summarizing findings from primary studies meta-analytically, new insights can be gained into the relations between variables, which can foster theoretical and practical advances. Also, meta-analyses uncover gaps in the current research literature and flaws in primary studies which can be addressed by future studies (Cooper et al., 2009). Careful application of meta-analytic models is necessary in order to account for possible biases in the effect size estimates.

While Study 1 uses meta-analysis to answer questions about the average effect of prior knowledge on learning outcomes, the research question in Study 2 regarding multivariate trajectories of knowledge over time requires a different methodological approach. As introduced in Chapter 4, multivariate assessments of knowledge and longitudinal latent variable mixture modeling can be used to study trajectories of knowledge on a more detailed level. In the following, I introduce the methodological aspects that pertain to latent class or profile analysis and their longitudinal extensions.

6.2 Latent Class Analysis, Latent Profile Analysis and their Longitudinal Extensions

Classifying individuals with and without reading disability, modeling changes in peer victimization in schools, and identifying students with misconceptions and fragmented knowledge - there are many cases in educational research where there is unobserved heterogeneity in the data (B. O. Muthén, 2001). Latent variable mixture modeling summarizes statistical models that can be used to account for this unobserved heterogeneity by classifying individuals based on a set of indicator variables on a categorical latent variable. Latent class analysis (LCA) uses categorical indicator variables which measure a construct in a cross-sectional study design, while latent profile analysis (LPA) is used for continuous indicators (Hickendorff et al., 2017). LCA and LPA models have two types of model parameters: latent class probabilities and conditional item probabilities (Hickendorff et al., 2017; B. O. Muthén, 2001). Class probabilities refer to the number or percentage of individuals that is assigned to a latent class, i.e., the size or prevalence of the latent class. Conditional item probabilities refer to the measurement part of the model that specifies how the class indicators are related to the latent class variable. Item probabilities can then be used to estimate the class membership for each individual using posterior class probabilities, similar to estimating the factor scores for individuals in factor analysis (B. O. Muthén, 2001; Nylund, 2007). As probabilities are used, class membership is not fixed but fractional, and individuals can have non-zero probabilities in several classes. For example, an individual can have posterior probabilities of .80, .15, and .05 for being assigned to classes 1, 2, and 3, respectively (Nylund, 2007). In this case, the most likely class membership for this individual is class 1.

LCA, LPA and their longitudinal extensions are largely data driven procedures, however, in the process of analysis the researcher has to make at least three crucial decisions concerning (1) the measurement model, (2) the number of latent classes, (3) the inclusion of covariates and distal outcomes, and (4) allowed transitions between classes or profiles over time in longitudinal models. First, the researcher has to decide by which indicators the latent classes or profiles shall be estimated. In longitudinal models, when using latent class transition analysis (LCTA) or latent profile transition analysis (LPTA), the researcher must also decide whether changes in the measurement model can be expected over time or whether measurement invariance is assumed. Item probabilities in LCTA, and means or means and variances in LPTA can be constrained equal across time to ensure that the meaning of latent classes does not change over time. Constraining the means of a profile at T1 and T2 to be equal allows researchers to interpret the profile the same way, for example as misconceptions profile. Second, the researcher has to decide about the number of latent classes that is to be estimated. This issue is related to the first one, as in LCTA and LPTA it is recommended to estimate the classes or profiles simultaneously for all measurement points when measurement invariance is assumed, whereas in an unconstrained measurement model, the classes or profiles can be estimated separately for each measurement point (Hickendorff et al., 2017).

There is no single criterion for selecting the best fitting model, but a simulation study has shown that the Bayesian information criterion and the bootstrap likelihood ratio test perform well under varying conditions (Nylund, Asparouhov, & Muthén, 2007). Usually, the information criteria, Akaike information criterion (AIC), Bayesian information criterion (BIC), and the sample-size adjusted BIC are plotted on the y axis of a diagram, with consecutively estimated models of one class to n classes on the x-axis. While typically, the AIC is reported to decrease the more classes are estimated by the model, the BIC levels out or shows a steep increase toward a specific solution, indicating worsening model fit (e.g., Kainulainen et al., 2017). Additionally, the bootstrap likelihood ratio test can be used in LCA and LPA to test whether there is a statistically significant increase in model fit between a model with k-1 and k classes or profiles (Nylund et al., 2007), otherwise the more parsimonious model should be chosen. However, in some cases there are theoretical considerations that suggest a specific model choice, for example choosing the five-class solution instead of the four class solution because of an interesting additional profile (e.g., M. Schneider & Hardy, 2013).

The third important model choice which should be taken by the researcher before specifying the LCA, LPA, LCTA, or LPTA model, is whether or not the researcher wants to relate the classes to covariates, concurrent outcomes, or distal outcomes confer (Clark & Muthén, 2009; Collins & Lanza, 2010; Lanza, Tan, & Bray, 2013). Covariates can be used to predict initial class sizes and transition probabilities in longitudinal models; concurrent and distal outcomes on the other hand are predicted by class membership. Covariates can be included in a one-step procedure or in a three-step procedure (Hickendorff et al., 2017) which has several implications for model interpretation. Fourth, researchers have to decide which transitions will be allowed in an LCTA or LPTA model. In most models first-order effects are assumed, for example by predicting the class or profile membership at T2 from the T1 memberships, specifically, estimating the conditional probabilities for the T2 memberships given T1 memberships the transition (B. O. Muthén & Asparouhov, 2011) but it is also possible to specify higher-order effects, for example from T1 classes or profiles on T3 classes or profiles, when lasting direct effects are assumed (Nylund, 2007).

Latent variable mixture models have several advantages compared to other statistical techniques, yet there are some potential pitfalls that have to be taken into account when conducting these analyses. One advantage that applies to all statistical models that use latent variable is, that measurement error is accounted for. Furthermore, model fits can be calculated and compared if the number of parameters to be estimated is smaller than the number of given parameters. This enables researchers to test different plausible models, compare their model fits and choose the best fitting model based on empirical and theoretical considerations. However, a pitfall of latent variable mixture modeling is that it is an explorative procedure, which can limit the generalizability of findings if the researchers do not implement strategies to validate their latent class or profile solutions (Hickendorff et al., 2017). Another pitfall lies in the process of model estimation as the likelihood which is estimated by the expectation maximization (EM) algorithm in Mplus can yield several local maxima (B. O. Muthén, 2001), and consequently, ambiguous class or profile solutions. For this reason it is recommended to use several starting values, increase the number of random starts and iterations, and monitor whether the best log likelihood values are replicated several times, particularly in complex models (Hickendorff et al., 2017). Additionally, model complexity can be reduced by decreasing the number of parameters to be estimated, e.g., by model constraints, or by increasing the number of given information, e.g., collecting more data. Finally, LCA, LPA, LCTA, and LCPA are only feasible with rather large sample sizes that exceed at least a minimum of $N = 70$, and better comprise more than a hundred participants (Wurpts & Geiser, 2014). Yet, these numbers are only recommendations and there is no definite rule for choosing the sample size. A more sophisticated but also time-consuming approach is to conduct a Monte Carlo Study to decide on the sample size (L. K. Muthén & Muthén, 2002).

Overall, latent class, latent profile analysis, and their longitudinal extensions are useful for many questions in educational research as they can take account of unobserved heterogeneity in the data. Advantages of these models include that latent variables are estimated free of measurement error and structural equation modeling allows to test and compare several models. Researchers who want to use these analyses should bear in mind that the procedures are explorative and should also include theoretical considerations when deciding on the number of latent classes or profiles.

Whereas the multivariate profiles of knowledge and discontinuous learning trajectories over time were the focus of Study 2, Study 3 investigated longitudinal dynamics of stress and their relations to academic achievement. In the following, I introduce latent state-trait analysis as a methodological approach to study the temporal dynamics of students' stress levels in Study 3.

6.3 Latent State-Trait Analysis

Longitudinal dynamics of psychological constructs have become the focus of attention in contemporary research in the social sciences, for example when studying depression, anxiety, stress, and life satisfaction (Geiser et al., 2015; Steyer, Mayer, Geiser, & Cole, 2015). One powerful statistical tool to investigate these temporal dynamics is latent state trait (LST) analysis which is rooted in latent state-trait theory (Schmitt & Steyer, 1993; Steyer, 1999; Steyer et al., 2015). Latent state-trait theory overcomes some limits imposed by classical test theory on the measurement of psychological traits by allowing to investigate time-dependent changes in a measured psychological construct. In LST theory it is possible to distinguish between four sources of time-dependent changes: trait changes, situational variation, changes in the measurement instruments, and measurement error (Steyer et al., 2015). For example, in clinical research one is often interested in long lasting, irreversible changes, such as changes in trait depressiveness in intervention studies. By the means of LST analysis, these changes can be modeled while taking situational fluctuations of depressiveness into account which occur due to situational differences, are relatively short in duration, and reversible (Geiser et al., 2015). Other studies primarily focus on these situational fluctuations, while assuming trait stability, for example studies on moods, physiological stress levels, and experienced emotions.

State and trait variables can be implemented in LST models by the means of a hierarchical factor model, with the state variables as first order factors and a trait variable that is superimposed on the state factors as a second order factor. Additionally, one or several method factors can be included into the model that are common to all measurement occasions as first order trait factors. Statistically, such a hierarchical LST model requires that there are at least two indicators per construct at each measurement occasion (i= 1,…, m; m≥2), and that there are at least two measurement occasions (k=1,…, n; n≥2). In this model, the observed variables Y_{ik} (for all i=1,…, m and k=1,…, n) are all decomposed the following way (Schmitt & Steyer, 1993). The observed variable $Y_{ik}$is decomposed into a true score and measurement error:

$Y_{ik} = \tau_{ik} + \epsilon_{ik}$ (Equation 3)

The true score τ_{12} is decomposed into the latent state variable η_k and a method factor $\xi_i$which is a common trait variable:

$\tau_{ik} = \beta_{ik}\eta_k + \kappa_{ik}\xi_i$ (Equation 4)

The latent state variable η_k is decomposed into the latent trait factor ξ and a state residual factor ζ_k:

$\eta_k = \gamma_k\xi + \zeta_k$ (Equation 5)

When Equation 5 is inserted in Equation 4, and Equation 4 is inserted in Equation 3, the complete decomposition of the observed variable Y_{ik} is described by:

$Y_{ik} = \beta_{ik}(\gamma_k\xi + \zeta_k) + \kappa_{ik}\xi_i + \epsilon_{ik}$ (Equation 6)

The structure of the model implies that the covariances among all latent variables are zero, as well as the covariances among the latent variables, the error terms, and state residuals. The variances of the observed variables can be decomposed into variance attributed to state effects, trait effects, method effects, and error variance. From these variances, four coefficients can be estimated: *reliability*, *common consistency*, *method specificity*, and *occasion specificity*. Reliability refers to the proportion of variance of the observed variables that is due to variance in the true scores, which is denoted by:

$Rel(Y_{ik}) = \frac{Var\ (\tau_{ik})}{Var(Y_{ik})}$ (Equation 7)

Common consistency refers to the proportion of variance in the observed variables that is due to stable trait differences, which is denoted by:

$Con(Y_{ik}) = \frac{\beta^2{}_{ik}\gamma^2{}_{ik}Var\ (\xi)}{Var(Y_{ik})}$ (Equation 8)

Method specificity refers to the proportion of variance that is due to stable differences in the specific method of measurement, which is denoted by:

$MetSpe(Y_{ik}) = \frac{\kappa^2{}_{ik}Var\ (\xi_i)}{Var(Y_{ik})}$ (Equation 9)

And finally, occasion specificity refers to the proportion of variance that is due to systematic differences in the situation of measurement and the person x situation interaction, denoted by:

$OccSpe(Y_{ik}) = \frac{\beta^2{}_{ik}Var\ (\zeta_k)}{Var(Y_{ik})}$ (Equation 10)

Pitfalls of LST analysis include measurement non-invariance and small designs (Geiser et al., 2015). Measurement non-invariance occurs when the scores of individuals of different groups or at different measurement occasions do not lie on the same scale which can cause artefactual mean differences or correlations with external variables (Widaman & Reise, 1997). Thus, measurement non-invariance in latent state-trait analysis has far-reaching consequences for the interpretation of the longitudinal dynamics of the construct measured (Geiser et al., 2015). Testing for state variability and trait change processes in a meaningful way is only possible when the measurement instruments do not change over time. If the latent trait factor loadings and intercepts vary over time, they may mask changes in the means and variances of the latent trait over time. Therefore, establishing measurement invariance at least on the latent trait factors is crucial to establish a state variability process. Measurement invariance can be tested statistically by estimating a series of models that increase restrictions on the parameters of the measurement model (Widaman & Reise, 1997). The fit indices of these models can then be compared and the level of measurement invariance can be established. Geiser et al. (2015) recommended that additionally to invariance testing, the mean structure of the latent states should be included into the model, which is also necessary in order to test for strong measurement invariance. Estimating latent state means can uncover trait change processes that otherwise would be overlooked. Small designs should be avoided if possible, as these result in a lack of power when testing assumptions. It is recommended that at least three indicators at four or more time points should be used.

In sum, LST analysis is a research tool that can be used to provide insights into the temporal dynamics of psychological constructs in longitudinal designs, such as anxiety, well-being, or stress by separating sources of variance that are attributable to state effects, trait effects, method effects, and measurement error. LST analysis has almost universal applications in psychological research, as psychological assessments never take place in a “situational vacuum and we always measure persons in situations” (R. Steyer, 1999, p. 392). Potential pitfalls include measurement non-invariance and small designs, which can lead to bias in the coefficients of key interest in LST analysis, i.e., consistency and occasion specificity, and low power when testing for model assumptions (Geiser et al., 2015). Especially if invariance assumptions are not met, it is advisable to test alternative models, for example, single trait models.

In the following chapter, I provide an overview of the three studies that were conducted as part of the current dissertation. Next, the three studies are presented in Chapters 8 to 10.

7. The Current Dissertation

The importance of academic achievement for individuals and the society has been documented by numerous studies (Furnée et al., 2008; OECD, 2000; Roth et al., 1996; Roth & Clarke, 1998) and much research in recent years has focused on ways to predict and promote academic achievement in practice. From a theoretical point of view, high achieving students inside and outside higher education interact with their learning environments in a way which optimizes intra-learner processes on the biological, cognitive, and non-cognitive level, for example by increasing long-term potentiation, attention, reasoning, and motivation (see LIKE framework, Figure 1). Inside higher education, prior knowledge and stress are considered two important influences on these learning processes. Prior knowledge as well as stress have been found to affect many factors that enable learning and academic achievement, such as attention, memory encoding and recall, learning strategies, academic self-efficacy, and motivation (Dochy et al., 2002; Hailikari, Nevgi, et al., 2008; Joëls et al., 2006; LePine, Podsakoff, & LePine, 2005; Merz et al., 2016; Vogel & Schwabe, 2016).

Despite numerous investigations have established the significance of prior knowledge and stress for learning in higher education, there is a lack of meta-analysis and longitudinal studies of prior knowledge and stress which are the focus of the current dissertation. Specifically, there remains a need for clarifying to which degree prior knowledge facilitates learning, as prior knowledge can yield facilitating as well as hindering effects, and null effects depending on the nature of prior knowledge, instructional strategies, learner characteristics, and so on. Furthermore, there is a gap in our knowledge concerning the developmental trajectories of students' knowledge in higher education, specifically, how correct and incorrect elements of knowledge in a domain evolve over time and whether these knowledge trajectories reflect conceptual change processes. Finally, previous work has established differential effects of acute, state stress and chronic, trait stress for cognition and achievement, but no study has investigated how state and trait stress relate to students' academic achievement in higher education. The current dissertation aimed at closing these gaps in the research literature by conduction three studies: 1) a meta-analysis of prior knowledge and later learning outcomes, 2) a longitudinal study that provided fine-grained analyses of multivariate knowledge trajectories over the course of two academic years and 3) a longitudinal investigation of experienced stress over the course of one semester using a latent state-trait model.

The first study (Chapter 8) is a meta-analysis that investigated the effects of prior knowledge on later learning outcomes. There is much evidence that prior knowledge is one of the strongest positive predictors for learning (e.g., Dochy et al., 2002) but it is not clear whether this effect is robust for different types of knowledges, content domains, educational levels and instructional methods. For example, prior knowledge can sometimes hinder learning as it might be incorrect or lead to inflexible problem solving strategies (Bilalić et al., 2008b; M. Fisher & Keil, 2016; Vosniadou et al., 2008). Furthermore, some instructional strategies may compensate for low prior knowledge, leading to greater learning gains in students' with low prior knowledge (Dahlström, 2012; Gay, 1986; Krause et al., 2009). Thus, the meta-analysis investigated the magnitude and direction of the overall net effect of prior knowledge on later learning across different domains and educational levels. Moderator analyses were used to investigate aspects of knowledge, content domains, and educational situations in which prior knowledge has stronger or weaker effects on learning outcomes.

Whereas in the first study, the bivariate relationship between prior knowledge and learning outcomes was considered, the second study (Chapter 9) focused on the multivariate relationship between prior knowledge and learning outcomes, developmental trajectories of knowledge over time and whether these trajectories reflect conceptual change processes. It was investigated how conceptual knowledge of human memory evolves among undergraduate students in psychology over the course of their first two academic years. It was hypothesized that similar to learning in K-12 education, there is evidence for misconceptions and fragmented profiles of knowledge at the beginning of instruction and a development towards better integrated and scientifically correct knowledge over the first two academic years in a sample of undergraduates in psychology. The different profiles of conceptual knowledge were supposed to reflect differences in achievement, indicated by the grades students achieved in an exam of human memory, learning, motivation, and emotion.

The third study (Chapter 10) investigated how stress experienced during a short period of time (i.e., several weeks at the end of semester) and stress experienced over a prolonged period of time (i.e., the whole semester) relates to students' academic achievement at the end of the respective semester. Therefore, in a first step it was investigated whether university students' stress levels can be described by means of a latent state-trait model in two datasets, including a pre-study with two measurement points over the course of one academic year, and the actual study which was conducted longitudinally at ten measurement points over the course of one semester. In a second step students' state and trait stress in the second data set were correlated with students' GPA at the end of the semester.

8. Study 1: Domain-Specific Prior Knowledge and Learning: A Meta-Analysis

8.1 Abstract

Domain-specific prior knowledge has been hypothesized to be one of the strongest positive determinants of learning, but it might sometimes also hinder learning (e.g., misconceptions or in negative transfer). This meta-analysis integrated empirical findings on the relation between the amount of domain-specific prior knowledge and subsequent cognitive learning outcomes (i.e., knowledge and achievement). We included 240 articles reporting 4327 effect sizes obtained with 62,129 participants. Almost all the studies investigated the correlation between prior knowledge and knowledge at posttest, which was high overall ($r^+ = .521$, 95% CI [.491, .550]) and for many moderator levels (e.g., countries and domains). Randomized controlled trials demonstrated the causality of this relation. Surprisingly few studies investigated the correlation between prior knowledge and knowledge gains over time, leading to a low statistical power of these analyses. Descriptively, there was a compensatory effect (negative correlation) for instruction with low cognitive demands and a Matthew effect (positive correlation) for instruction with high cognitive demands and, but only the former reached statistical significance. Overall, the high stability of individual differences in knowledge supports theories emphasizing the accumulative long-term nature of knowledge acquisition. More randomized controlled trials investigating knowledge gains are needed. Generally, several processes (e.g., encoding and elaboration) mediate the effect of prior knowledge on learning. The moderators of prior-knowledge effects need to be interpreted in terms of which mediating processes they affect. The discussion elaborates on these issues and offers a framework for further research on learning through knowledge acquisition.

8.2 Theoretical Background

The *knowledge-is-power hypothesis* (Hambrick & Engle, 2002; Möhring, Schroeders, & Wilhelm, 2018) states that "the most important single factor influencing learning is what the learner knows already" (Ausubel, 1968, p. vi). Over the past decades, many researchers have investigated this assumption. For example, memory researchers have demonstrated that the content of long-term memory affects how new information is processed in working memory and encoded in long-term memory (Baddeley, Eysenck, & Anderson, 2009). Cognitive scientists have devised models of these mechanisms (Anderson et al., 2004; Gopnik & Wellman, 2012; Laird, 2012). Developmental psychologists of the past (Piaget, 1971; Vygotsky, 1978) and the present (Case, 1992; Siegler, 1996; Wellman & Gelman, 1992) have investigated the role of knowledge in cognitive development over the lifespan. Educational psychologists have incorporated prior knowledge as a central component in theories on academic achievement (Thompson & Zamboanga, 2003), expert performance (Ericsson & Charness, 1994), transfer (Barnett & Ceci, 2002; Singley & Anderson, 1989; Thorndike & Woodworth, 1901), multimedia learning (Mayer, 2001), and conceptual change (diSessa, 2008; Vosniadou, 1994). Surprisingly, this plethora of research has left central questions open: How strong is the overall mean effect of prior knowledge on learning? How variable are the effect sizes? Are there systematic variations in the effect sizes that can be explained by moderator variables?

For the purpose of the present investigation, we define *knowledge* as information stored in memory (e.g., Anderson, 1983; Weinert, 1999). In line with how the term knowledge is typically used in the psychological and educational research literature, this definition includes declarative knowledge about abstract and relational concepts (Goldwater & Schalk, 2016) and more isolated facts (M. Schneider & Grabner, 2012) as well as procedural knowledge about how to solve problems (Anderson et al., 2004). It also includes scientifically incorrect misconceptions as well as scientifically correct concepts (Shtulman & Valcarcel, 2012; Smith III., diSessa, & Roschelle, 1993). Knowledge is *domain specific* when it relates to the key principles in a domain, for example, the concept of equivalence in mathematics or the concept of force in physics (Carey & Spelke, 1994; Wellman & Gelman, 1992). Domain-specific knowledge is sometimes also termed content knowledge (Chi & Ceci, 1987) and has been described as a component of academic achievement, expertise, and similar cognitive learning outcomes (Gobet, 2005; J. E. Hunter, 1986; OECD, 2016; Steinmayr et al., 2014). We define *prior knowledge* as the knowledge available in a person's long-term memory at the onset of learning (cf. Dochy & Alexander, 1995).

These definitions have two implications for the current study. First, they imply that investigations of the influence of prior knowledge on subsequent learning require an assessment or manipulation of knowledge before learning (e.g., a pretest) and an assessment of the outcomes after learning (i.e., a posttest). Only then can researchers relate the differences in prior knowledge before learning to the outcomes after learning or the knowledge gains. Second, if prior knowledge is knowledge before learning, it is necessarily incomplete or fully or partly incorrect.

Otherwise, no learning of something new and relevant would be possible. This partly incorrect or incomplete nature of prior knowledge raises the question whether prior knowledge always facilitates or sometimes hinders further learning.

8.2.1 Processes Mediating Positive and Negative Effects of Prior Knowledge on Learning

Previous studies have found that prior knowledge can positively affect learning mediated through some mental processes (Dochy et al., 1999; Hambrick & Engle, 2002; W. Schneider, 1993; Stern, 2001), but can negatively affect learning mediated through other mental processes (Bilalić, McLeod, & Gobet, 2008a; M. Fisher & Keil, 2016; Luchins & Luchins, 1987; Vosniadou, 2008). For example, prior knowledge can positively affect learning because it guides learners' attention (e.g., Tanaka, Kiyokawa, Yamada, Dienes, & Shigemasu, 2008; Yu, Zhong, & Fricker, 2012) and facilitates the interpretation and encoding of new information (Kintsch, 1994; van Kesteren, Rijpkema, Ruiter, Morris, & Fernández, 2014). It also allows for the bundling of new information into chunks that can efficiently be memorized, processed, and retrieved (Chase & Simon, 1973b; Ericsson, Chase, & Faloon, 1980; Gobet et al., 2001). Prior knowledge about the effectivity and efficiency of problem-solving strategies enables more exploration, goal-directed behavior, and the construction of more advanced new strategies (M. Schneider, Rittle-Johnson, & Star, 2011; Siegler, 1996).

Prior knowledge can negatively affect learning mediated through several other cognitive processes. For example, misconceptions and incomplete correct knowledge in a domain (e.g., that the surface of the earth looks flat in everyday life) can give rise to incorrect conclusions (e.g., the earth is a disc; Vosniadou & Brewer, 1992) that hamper further learning. Learners with high correct prior knowledge in a domain tend to pay selective attention to the features of a situation that have been relevant for solving problems in the past. This can induce perceptual biases (Hecht & Proffitt, 1995; Lewandowsky & Kirsner, 2000) or prevent learners from looking for new and better problem solutions (Einstellung effect; Bilalić, McLeod, & Gobet, 2010; Luchins & Luchins, 1959). The extended practice necessary to automatize procedural knowledge is another possible cause of inflexible behavior (A. Johnson, 2003; Müller, 1999). Having more knowledge elements about a topic increases the probability of intrusions or interferences involving these elements in the same domain (Arkes & Freedman, 1994; Castel, McCabe, Roedinger, & Heitman, 2007) and negative transfer to other domains (Woltz, Gardner, & Bell, 2000). For example, children's highly automatized and correct knowledge about whole numbers can interfere with learning about fractions, which look similar but differ in important mathematical characteristics, such as their density (Siegler, Fazio, Bailey, & Zhou, 2013).

The mechanisms mediating the positive or negative effects of prior knowledge on learning do not mutually exclude each other. They might work in parallel and interact with each other. This interplay raises questions about the combined net effect of prior knowledge on learning. Is it positive, negative, or close to zero because positive and negative influences cancel each other out?

8.2.2 The Compensatory Effect and the Matthew Effect

The effect of prior knowledge does not only imply how much learners with high or low prior knowledge will know in the future, but it also implies how the variability of knowledge in a group of learners will change over time. When students with low prior knowledge acquire more new knowledge than their peers, learning reduces the preexisting differences between students. This narrowing of the gap between more and less knowledgeable learners over time has been termed the *compensatory effect* (Schroeders, Schipolowski, Zettler, Golle, & Wilhelm, 2016). Figure 2 (bottom left) depicts an example of a compensatory effect by showing five fictitious persons' amounts of knowledge before and after learning.

In contrast, when students with higher prior knowledge acquire more new knowledge than their peers, learning amplifies the preexisting differences between students. This widening of the gap between more and less knowledgeable learners over time has been termed the *Matthew effect* (e.g., Stanovich, 1986; Walberg & Tsai, 1983). Figure 2 (top left) depicts an example of the Matthew effect. From an educational point of view, the Matthew effect has the disadvantage that the increased heterogeneity in knowledge can be associated with social and ethnic inequality between students, thus hampering instruction (cf. Baumert, Nagy, & Lehmann, 2012). So far, the compensatory effect and the Matthew effect have only rarely been investigated in studies on knowledge acquisition. Most previous studies on these effects have used academic achievement measures but yielded heterogeneous results as to when Matthew effects or compensatory effects occur and how strong they are (Baumert et al., 2012; Duff, Tomblin, & Catts, 2015; Pfost, Dörfler, & Artelt, 2011; Schroeders et al., 2016).

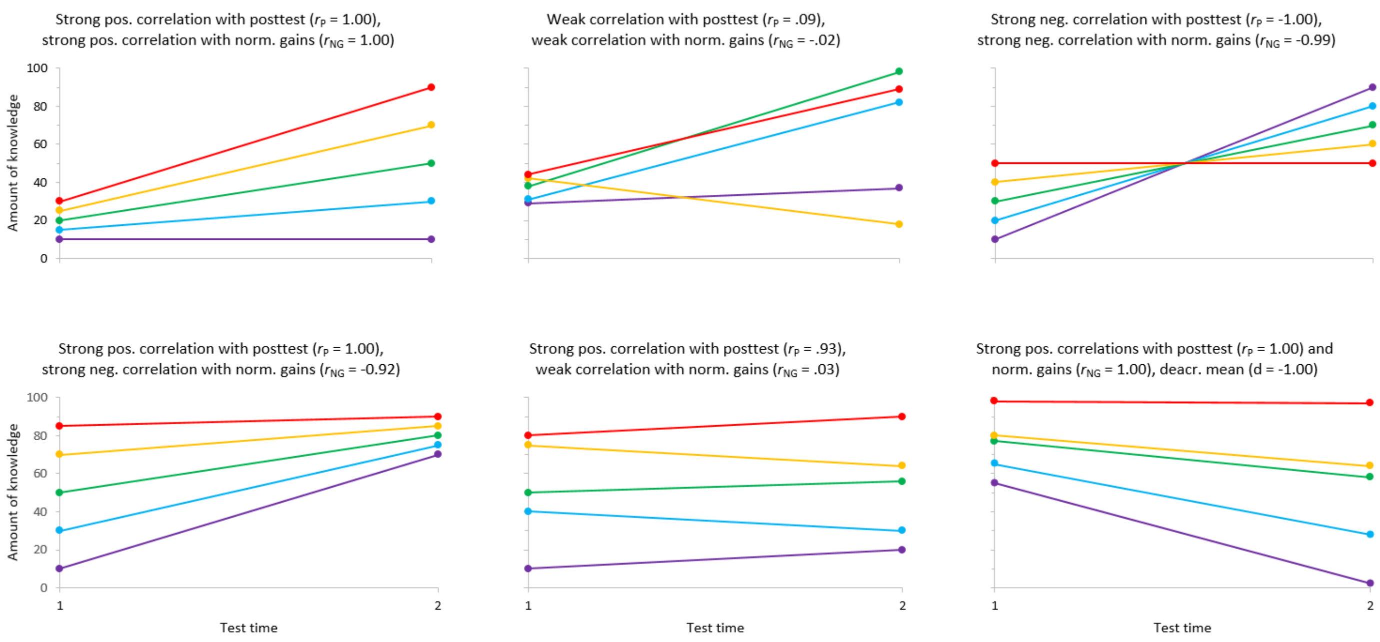

Figure 2. Six different sets of correlations of prior knowledge (T1) with posttest knowledge (T2) and normalized knowledge gains, visualized using five fictitious persons' amounts of knowledge as examples.

8.2.3 Prior Knowledge as a Predictor of Posttest Knowledge, Absolute Knowledge Gains, or Normalized Knowledge Gains

Most previous studies on prior knowledge reported the correlation r_P between prior knowledge and posttest knowledge. Only a few studies reported the correlation r_G between prior knowledge and the knowledge gains from pretest to posttest. In the present meta-analysis, we show that the two types of correlations capture different aspects of prior-knowledge effects on learning and differ empirically.

The correlation r_P of prior knowledge with posttest knowledge indicates to what extent individual differences between learners' amounts of knowledge remain stable from before to after learning. A strong positive correlation implies that the rank order of learners with respect to the amount of their knowledge remains relatively unchanged. The correlation r_P can thus be used to predict how well a learner will perform compared to other learners.

In contrast, the correlation r_G between prior knowledge and the knowledge gains indicates to what extent learners with a high amount of prior knowledge have smaller knowledge gains (r_G <

0; compensatory effect) or larger knowledge gains ($r_G > 0$; Matthew effect) than their peers. The correlation can thus be used to predict to what extent learners with low prior knowledge fall behind more knowledgeable learners or catch up with them. Knowledge gains can either be computed as absolute gain scores (i.e., *AG = posttest score – pretest score*) or as normalized gain scores (i.e., NG = (posttest score – pretest score) / (scale maximum – pretest score); Hake, 1998). Absolute gain scores have the disadvantage that learners with high prior knowledge have less room on the scale for improvements than learners with low prior knowledge. Normalized gain scores account for this problem by dividing the absolute learning gain by the maximal possible learning gain (Hake, 1998). In the following, we denote the correlation between prior knowledge and absolute knowledge gains as r_{AG} and the correlation between prior knowledge and normalized knowledge gains as r_{NG}.

The correlations of prior knowledge with posttest knowledge and with knowledge gains are partly independent of each other. Figure 2 visualizes this for posttest knowledge and normalized gains. It is possible that both correlations (r_P and r_{NG}) are strongly positive (top left), close to zero (top middle), or strongly negative (top right), but it is also possible that the correlation with posttest knowledge is strongly positive, whereas the correlation with knowledge gains is strongly negative (bottom left) or close to zero (bottom middle). The same holds true and the graphs would look similar for absolute knowledge gains. The correlations r_P, r_{AG}, and r_{NG} do not indicate whether the sample, on average, gained knowledge or lost knowledge because correlations are invariant to transitions of the mean. For example, strong positive correlations r_P and r_{NG} can go along with pretest-posttest increases (Figure 2, top left) and pretest-posttest decreases (bottom right) of the average amount of knowledge in the sample. Due to the differences of the correlations of prior knowledge with posttest knowledge, absolute knowledge gains, and normalized knowledge gains, we analyzed the three correlations separately in our meta-analysis.

8.2.4 Possible Moderating Influences

Systematic differences between the studies on prior knowledge might explain why some studies found stronger correlations between prior knowledge and posttest knowledge or knowledge gains than others did. In the following sections, we introduce three groups of possible moderators: (1) knowledge characteristics, (2) learner characteristics, and (3) environmental characteristics.

Knowledge Characteristics. A characteristic of knowledge that potentially moderates the relation between prior knowledge and learning is the *type of knowledge*. A widely used distinction between knowledge types is the one between declarative and procedural knowledge (Anderson et al., 2004). Declarative knowledge is verbalizable and explicit. Two educationally relevant subtypes of declarative knowledge are fact knowledge (e.g., 5 + 2 = 7) (M. Schneider & Grabner, 2012), which tends to consist of isolated pieces, and conceptual knowledge, which tends to be more abstract and relational (e.g., the principle of commutativity in mathematics) (Goldwater & Schalk, 2016). Procedural knowledge consists of automatized and implicit rules that specify which operators help to reach a goal. It includes knowledge about cognitive operators, for example, how to add two numbers, and motoric procedural knowledge, for example, how to jump (A. Johnson, 2003; Rittle-Johnson, Schneider, & Star, 2015). A second characteristic of knowledge that might moderate the relation between prior knowledge and learning is the *content domain*. Content domains can differ in how closely and systematically elements of knowledge are interrelated. Prior knowledge might more strongly facilitate learning in domains with higher degrees of interrelatedness than in domains with lower degrees of interrelatedness. A third potential moderator is the *similarity* of prior knowledge and posttest knowledge. The more similar prior knowledge is to new knowledge, the easier it might be for a learner to use the prior knowledge to learn the new knowledge (Medin, Goldstone, & Gentner, 1993). This similarity is multidimensional. Among the dimensions on which two pieces of knowledge can be more or less similar are knowledge domain, physical context, temporal context, functional context, social context, and modality of assessment (Barnett & Ceci, 2002).

Learner Characteristics. Possible moderators on the level of learner characteristics are age and educational level. *Age* is a potential moderator because it relates to differences in cognitive processing and learning (e.g., executive functions; Lejeune, Desmottes, Catale, & Meulemans, 2015; Steinberg, 2014). Furthermore, the importance of prior knowledge might differ between learners' *educational levels* (kindergarten, primary school, secondary school, higher education, etc.), which vary in learner age, required prior knowledge, and, partly, instructional methods.

Environmental characteristics. Environmental characteristics might be further moderators of the relation between prior knowledge and learning. Instructional interventions designed in ways to activate, explicate, and address prior knowledge likely strengthen the relation between prior knowledge and learning (Baumert et al., 2010; Pressley et al., 1992), as students often fail to spontaneously activate prior knowledge when it would be relevant (Renkl, Mandl, & Gruber, 1996). For studies in which the learners participated in an instruction intervention, we coded the setting (e.g., school instruction), duration, cognitive demands, and instructional methods of the intervention. We expected a particularly strong moderation effect of cognitive demands. Instruction with high cognitive demands requires learners to understand conceptual connections, analyze, justify, explain, and draw conclusions (Stein & Smith, 1998). This is easier for learners with high prior knowledge than for learners with low prior knowledge (Kirschner, Sweller, & Clark, 2006). Accordingly, we expected that instruction with high cognitive demands would amplify preexisting differences in prior knowledge leading to strong positive correlations between prior knowledge, posttest knowledge, and knowledge gains. Another possible environmental moderator is the *country*, as countries differ in their educational systems, curricula, and cultural patterns of teaching (Hiebert, Stigler, & Manaster, 1999, p. 196). Many other knowledge, learner, and environmental characteristics proved to be important in single studies, but could not be included in our meta-analysis because the number of respective studies was too small.

8.2.5 The Present Study

In sum, the concept of prior knowledge is widely used in many areas of psychological research, but the average strength of the effect of the amount of domain-specific prior knowledge on learning is still unknown. We therefore conducted a meta-analysis on this question. To maximize the generalizability of our findings, we did not restrict our analysis to any content domains, age groups, or institutional contexts. We included all studies using domain-specific prior knowledge to predict knowledge or achievement at posttest or knowledge gains from pretest to posttest. We included only knowledge, and not achievement, as an independent variable because we were interested in the effects of prior knowledge, which is related to but not identical with achievement. We included both knowledge and achievement as dependent variables because we were interested in the broad effects of prior knowledge on cognitive learning outcomes. In the remainder of the manuscript, for simplicity, we used the term *posttest knowledge* to refer to knowledge and achievement at posttest.

The present meta-analysis followed four research questions. Our first research question related to how much empirical evidence is available about the relation between prior knowledge and both posttest knowledge and knowledge gains, and about the causality of these relations. As prior knowledge has widely been used throughout psychology and education for decades, we expected to find a large amount of relevant empirical evidence.

Our second research question was about how prior knowledge relates to posttest knowledge. We hypothesized a strong positive correlation r_P between prior knowledge and posttest knowledge because learners' knowledge in a domain typically accumulates over months or years. This long-term nature of knowledge acquisition is emphasized, for example, by theories of conceptual change, conceptual development, and the acquisition of expert performance (diSessa et al., 2004; Keil, 1996; Siegler & Svetina, 2002; Ullén et al., 2015; Vosniadou, 1994). Accumulated individual differences grown over such extended time periods tend to be highly stable and hard to change by short learning interventions (Ericsson, Krampe, & Tesch-Römer, 1993; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). Thus, we expected that these differences could be observed in a similar way at pretest and at posttest, leading to a strong positive correlation r_P.

Our third research question concerned how prior knowledge predicts pretest-posttest knowledge gains (i.e., whether evidence for a compensatory effect or a Matthew effect exists). We expected to find a weak positive correlation r_{NG} between prior knowledge and normalized knowledge gains because the amount of knowledge gained during a learning phase depends on a multitude of instruction characteristics, teaching characteristics, and learner characteristics (e.g., Ormrod, 2011). Since prior knowledge is just one of these variables, it can only be expected to explain a part of the variance in learners' knowledge gains. We expected the correlation r_{AG} between prior knowledge and absolute knowledge gains to be even lower than r_{NG} because absolute gains penalize high prior knowledge and can thus underestimate the true association strength (Hake, 1998).

Our fourth research question related to how strongly the effects of prior knowledge on posttest knowledge and on knowledge gains are modulated by characteristics of the knowledge, the learner, and the environment. As explained, there are reasons to expect such moderating effects.

8.3 Method

8.3.1 Literature Search

Figure 3 summarizes the literature search process and inclusion criteria. In May 2015, we searched the title, abstract, and keywords of all articles in the literature database PsycINFO. The search string was designed to not only include studies explicitly using the term prior knowledge but also other studies investigating knowledge in designs with at least two measurement points.

The search string was as follows: *(pre-test OR post-test OR pretest OR posttest OR pre test OR post test OR longitudinal OR repeated measure* OR measurement point* OR prior knowledge) AND knowledge*. After limiting the results to empirical quantitative studies conducted with non-disordered participants, written in the English language, and published in a peer-reviewed journal, the search provided 4572 hits (for more details, see Figure 3). We updated the search two years later in May 2017. This search provided additional 890 hits, resulting in 5462 hits in total. We accounted for publication bias by using visual and statistical methods (described in the results section) instead of including unpublished studies. The quality of unpublished studies is hard to assess, and researchers usually obtain only a nonrepresentative set of unpublished studies, which does not increase the quality of the meta-analytic results (Ferguson & Brannick, 2011).

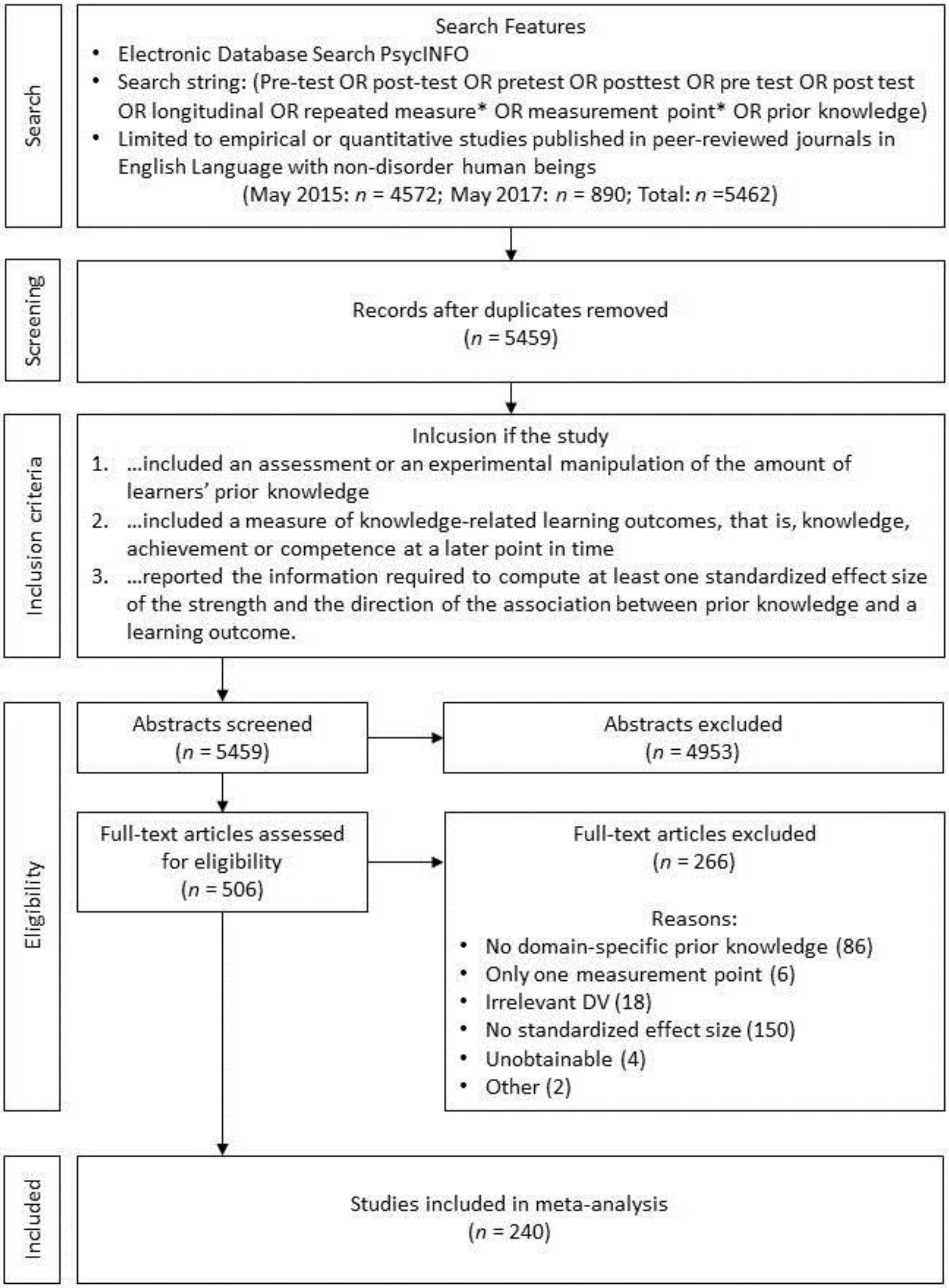

Figure 3. Flow chart for the literature search.

8.3.2 Inclusion of Studies

The three inclusion criteria were as follows: (1) The study included an assessment or an experimental manipulation of the amount of learners' domain-specific prior knowledge as defined in the introduction. To facilitate the interpretation of the results, we excluded studies that manipulated the activation rather than the amount of prior knowledge (e.g., Amadieu et al., 2015) or which compared learning with familiar versus unfamiliar materials (e.g., Badham, Hay, Foxon, Kaur, & Maylor, 2016). We only included objective quantitative measures of domain-specific prior knowledge and excluded self-assessments, composite scores from more than one domain, and measures of crystallized intelligence, abilities, achievement, or meta-cognitive knowledge. (2) The study included a measure of knowledge or achievement after learning or of the knowledge gains made from one measurement point to another. Studies in which the domain of the prior knowledge test and the learning outcome test differed were included. We also included composite measures of learning outcomes in more than one domain (e.g., GPA). We excluded learner self-ratings of their learning outcomes. (3) The study reported the information required to compute at least one standardized effect size of the strength and the direction of the association between prior knowledge and posttest knowledge, absolute knowledge gains, or normalized knowledge gains.

After removing three duplicates, we screened all remaining 5459 titles and abstracts and excluded studies that were obviously not relevant for our meta-analysis. For all other studies, we inspected the full texts. The first author screened all studies for inclusion. The fourth author independently screened a random sample of 115 studies for inclusion. The absolute intercoder agreement was 83%. Disagreements were resolved through discussion. The studies included in the meta-analysis are listed in Appendix A (SM1).

8.3.3 Data Coding

The coding rules were documented in a manual, which was used for coder training. The first author coded the information from all included studies. The second author independently coded the information for 547 effect sizes from 86 randomly chosen studies. The absolute inter-coder agreement was 90% of all double-coded values. Inter-coder disagreements were resolved through discussion. Appendix A (SM2) provides a list of all coded variables and their levels, respectively.

For studies with more than two measurement points and for longitudinal studies, we coded one effect size for each possible pair of measurement points. We coded the similarity of prior knowledge and posttest knowledge using the taxonomy proposed by Barnett and Ceci (2002). To ease coding, we only distinguished between two levels of each dimension (similarity high vs. similarity low). For example, we coded the similarity of the temporal context as high when prior knowledge and learning outcomes were assessed on the same day and as low when they were assessed on different days. In addition to Barnett and Ceci's dimensions, we coded the dimension knowledge type (e.g., high similarity for declarative knowledge at T1 and at T2, low similarity for declarative knowledge at T1 and procedural knowledge at T2). We coded the cognitive demands of the intervention based on the authors' description. We coded this moderator only for studies including at least one condition with higher cognitive demands and one with lower cognitive demands and when the conditions were explicitly described by the authors as having higher or lower cognitive demands relative to each other. When coding the instructional methods, we used the labels employed by the authors as our main criterion. For example, we only coded the employment of problem-based learning when the authors of the original publication used the term problem-based learning to describe their instructional intervention. We coded the country only for studies investigating knowledge in the context of school instruction.

8.3.4 Preparation of Effect Sizes

Appendix A (SM3) lists the equations for the preparation of the effect sizes for the meta-analytic integration. We used Pearson correlations (*r*) as the dependent variable of our meta-analysis because the majority of the included studies reported correlational results. Whenever a study compared a high- and a low-prior knowledge group and reported the group means and standard deviations for at least two measurement points, we calculated the absolute and the normalized gain scores from this information. Absolute knowledge gains were computed as *posttest score – pretest score*, and normalized knowledge gains were computed as (*posttest score – pretest score*) / (*scale maximum – pretest score*) (Hake, 1998). We then computed group differences in these gain scores as Cohen's *d* values, respectively. For studies reporting group differences for the posttest only, we computed Cohen's *d* for this difference. We then converted the Cohen's *d*s to Pearson correlations. Unless specified otherwise, all correlations were coded so that a positive value indicated that a higher quantity of (correct or incorrect) prior knowledge went along with a higher quantity of (correct or incorrect) posttest knowledge or positive pretest-posttest gains. The correlations were subjected to a Fisher's Z_r-transformation to approximate a normal sampling distribution (Lipsey & Wilson, 2001) before the meta-analytic integration.

Only when the original studies reported the reliabilities of the measures, we corrected the correlations for measurement error, which would otherwise make the correlation between two constructs appear smaller than it actually is (Schmidt & Hunter, 2015, p. 112). Some researchers measured prior knowledge as a continuous variable and then dichotomized the participants into a high- and a low-prior knowledge group. We corrected for this loss of information (i.e., loss of variance) using the formula given by Schmidt and Hunter (2015, p. 134). We winsorized the corrected correlations and variances to less extreme values when they exceeded a value above -1 or 1 due to attenuation. Throughout the manuscript, we use the symbol r^+ to refer to the corrected effect sizes.

8.3.5 Statistical Analysis

Outliers. Before we performed the meta-analysis, we identified outliers through Cook's values (Cook & Weisberg, 1982; Viechtbauer & Cheung, 2010) using the *metafor* package (Viechtbauer, 2010) in R (R Core Team, 2014). We excluded two outliers found with posttest knowledge (Lonigan, Burgees, & Anthony, 2000; Ree, Carretta, & Teachout, 1995), two with absolute knowledge gains (Zacharia, Loizou, & Papaevripidou, 2012) and one with normalized knowledge gains (Zacharia et al., 2012).

Publication Bias. We visually and statistically tested for publication bias using funnel plots, the trim and fill method(Duval & Tweedie, 2000), and Egger regressions for random-effects models (Egger, Smith, Schneider, & Minder, 1997). These tests were conducted using the *metafor* package (Viechtbauer, 2010) in R.

Meta-analytic integration. The formulas used for the meta-analytic integration of the effect sizes are provided in Appendix A (SM3). The majority of the included studies reported several effect sizes, for example, for various dependent measures or measurement points. These effect sizes are statistically dependent and thus violate a central assumption of classical meta-analytical models. To handle statistically dependent effect sizes, we employed robust variance estimation (Hedges et al., 2010; Tanner-Smith, Tipton, & Polanin, 2016). Given the expected heterogeneity, we used random effects models for the meta-analytic integration (cf. Raudenbush, 2009). The mean effect sizes and meta-regression models were estimated using a weighted least squares approach (cf. Hedges et al., 2010; Tanner-Smith & Tipton, 2014). The statistical analyses were performed using the *robumeta* package (Z. Fisher & Tipton, 2014) in R.

Moderators. We separately computed the mean effect size for every level within a moderator category. For levels with degrees of freedom smaller than 4, we only reported the mean, but not the confidence interval, as the results were not trustworthy due to the small number of observations (Z. Fisher & Tipton, 2014). These levels were also excluded from the significance tests of the moderators. Before performing the moderator analyses, continuous moderator variables were log-transformed to obtain normal distributions of the variables. Categorical moderators were dummy-coded using the moderator level with the lowest effect size as the reference level. Unless stated otherwise, each moderator was tested in a separate analysis. The moderators were entered as predictors in regression models for the prediction of the effect sizes (see SOM 3 for details). For dummy coded variables, each predictor indicated whether the coded moderator level significantly differed from the reference level. We computed the overall significance for each regression analysis using the Wald-test function of the *clubSandwich* package (Pustejovsky, 2017; Tanner-Smith et al., 2016) in R. We computed the overall proportion of explained variance R^2 for each regression model as described in the SOM 3.

8.4 Results

8.4.1 Research Question 1: Available Empirical Evidence

Characteristics of the Included Studies. The 240 included articles report findings from 335 independent samples and 62,129 participants in total. The studies presented or allowed the computation of 4327 effect sizes. The publication year ranged from 1965 to 2017, with a median of 2010. Of the 240 articles, 60% were published within the last 10 years (2008–2017), indicating an increasing interest in prior knowledge. Appendix A (SM2) lists the frequencies of the levels of the moderator variables in the included studies. Twenty-eight countries are represented in the meta-analysis. Most studies reported data from North America (45%), followed by Europe and the Middle East (32%), Asia (11%), and Australia/New Zealand (1%). The time between the measurement of prior knowledge (T1) and the measurement of the learning outcomes (T2) varied between 0 and 3780 days, with a median of 360 days. The learners' sample mean age ranged from 0.63 to 33.25 years with an overall mean of 11.13 years (SD = 7.02).

Characteristics of the Included Effect Sizes. A large number of studies reported the correlation between prior knowledge and posttest knowledge. This relation had been investigated in 235 included studies, reporting 4223 effect sizes ranging between $r_p^+ = -.635$ and $r_p^+ = .995$. Of these, 44 effect sizes from nine studies were obtained in randomized controlled trials (RCTs). We categorized a study as RCT when the participants were randomized into at least two groups and the levels of prior knowledge were manipulated to differ between these groups. For 28 studies with 1305 effect sizes, it was possible to control for intelligence using partial correlations.

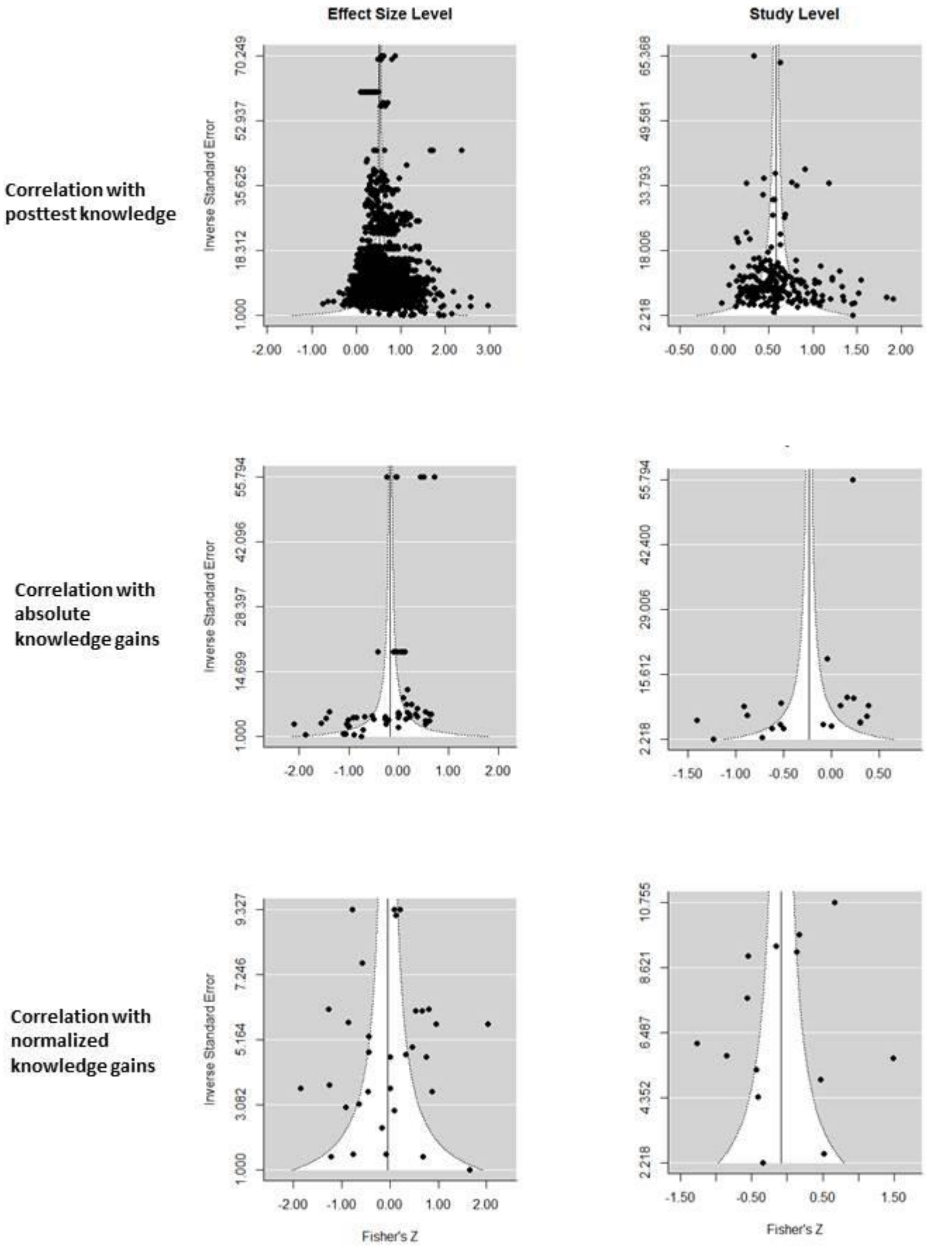

Figure 4. Funnel plots of the inverse of standard error and the effect size (corrected for artifacts and here transformed to Fisher's Z) for the effect-size level (left) and the study level of aggregation (right) for posttest knowledge (top), absolute knowledge gains (middle), and normalized knowledge gains (bottom).

In contrast, the association between prior knowledge and knowledge gains had been investigated in only a few studies, none of which was an RCT or allowed controlling for intelligence. The correlation between prior knowledge and absolute knowledge gains could be computed for 20 studies with 62 effect sizes ranging from $r_{AG}^+ = -.971$ to $r_{AG}^+ = .726$. The correlation between prior knowledge and normalized knowledge gains could be computed for 14 studies with 33 effect sizes ranging from $r_{NG}^+ = -.952$ to $r_{NG}^+ = .966$. Thus, overall, there is a strong asymmetry in terms of how much empirical evidence is available about prior-knowledge effects on posttest knowledge and about prior-knowledge effects on knowledge gains.

No evidence for a publication bias, that is, an underrepresentation of effect sizes close to zero, was found in our database. The effect sizes found with posttest knowledge, absolute knowledge gains, and normalized knowledge gains varied symmetrically around their common means, both for the individual effect sizes (Figure 4, left column) and for the study-average effect sizes (Figure 4, right column). Neither Egger regressions nor the trim-and-fill method (Duval & Tweedie, 2000) indicated any underrepresentation of small effect sizes in any of the six cases depicted in Figure 4.

8.4.2 Research Question 2: Correlation of Prior Knowledge with Posttest Knowledge

Table 1 shows the average effect sizes separately for studies using prior knowledge to predict posttest knowledge, absolute knowledge gains, or normalized knowledge gains. With respect to our second research question (i.e., how strongly prior knowledge is associated with posttest knowledge), we found a positive and statistically significant correlation with $r_P^+ = .525$. As expected, the correlation was strong according to the standards set by Cohen (1992), indicating a high stability of individual differences in knowledge from before to after learning. The I^2 index of 93.53 indicated a large amount of heterogeneity, implying that third variables moderate the extent of this stability.

The connection between prior knowledge and posttest knowledge is causal, as indicated by the significant positive effect size of $r_P^+ = .308$ found in the RCTs. The correlation was significantly lower in the RCTs than in other study designs, but this difference explained only a variance proportion of $R^2 = .007$ of the effect sizes. Controlling the correlation for intelligence across the 28 studies in which this was possible did not lead to a statistically significant decrease ($z = 1.23$, $p = .219$). Therefore, the association of prior knowledge with posttest knowledge is causal and cannot entirely be attributed to a confounding influence of intelligence.

Table 1

Overall Meta-Analytic Results for the Correlations of Prior Knowledge with Posttest-Knowledge, Absolute Knowledge Gains, and Normalized Knowledge Gains

	Correlation with posttest knowledge						Correlation with absolute knowledge gains						Correlation with normalized knowledge gains					
	j	k	r_P^+	CI r_P^+ 95%	τ^2	I^2	j	k	r_{AG}^+	CI r_{AG}^+ 95%	τ^2	I^2	j	k	r_{NG}^+	CI r_{NG}^+ 95%	τ^2	I^2
Studies allowing the computation of at least one of the three types of effect sizes	235	4223	.525	[.495, .553]	.099	93.53	20	62	-.210	[-.437, .043]	.380	96.99	14	33	-.083	[-.478, .342]	.580	94.83
Studies allowing the computation	12	30	.760	[.603, .864]	.362	90.09	12	30	-.398	[-.713, .050]	1.20	96.78	12	30	.123	[-.352, .556]	1.05	96.34

of all three types of effect sizes																		
Randomized controlled trial																		
No	226	4179	.532	[.502, .561]	.099	93.70	20	62	-.210	[-.437, .043]	.380	96.99	14	33	-.083	[-.478, .342]	.580	94.83
Yes	9	44	.308	[.186, .420]	.038	60.61	0	0	-	-	-	-	0	0	-	-	-	-
Controlling for intelligence																		
Before controlling	28	1305	.518	[0.464, 0.568]	.061	90.31	0	0	-	-	-	-	0	0	-	-	-	-
After controlling	28	1305	.482	[.424, .535]	.072	91.75	0	0	-	-	-	-	0	0	-	-	-	-

8.4.3 Research Question 3: Correlation of Prior Knowledge with Knowledge Gains

With respect to our third research question (i.e., how strongly individual differences in prior knowledge are associated with individual differences in knowledge gains), we expected to find weak positive effects. On the group level, we found strong pretest-posttest knowledge gains of Cohen's $d = 1.62$. However, the empirical findings with respect to the individual differences were inconclusive. The correlations between prior knowledge and gain scores were descriptively negative, but not significantly different from zero for absolute knowledge gains ($r_{AG}^{+} = -.210$) and for normalized knowledge gains ($r_{NG}^{+} = -.083$). The 95% confidence intervals around these means were extremely large because only 20 studies reported the correlation between prior knowledge and absolute knowledge gains (or the information required to compute this correlation), and a mere 14 studies reported the correlation between prior knowledge and normalized knowledge gains (or the information required to compute this correlation). The large confidence intervals make it impossible to draw any definitive conclusions about how strongly prior knowledge correlates with knowledge gains. Due to a lack of relevant empirical studies, the causality of the prior knowledge effects on knowledge gains and the influence of intelligence could not be investigated in our meta-analysis.

As hypothesized, the correlation with prior knowledge was highest for posttest knowledge as a dependent variable (r_{P}^{+}), lower for normalized knowledge gains (r_{NG}^{+}), and lowest for absolute knowledge gains (r_{AG}^{+}). The mean correlations found with posttest knowledge and absolute knowledge gains differed significantly ($z = 6.07$, $p < .001$). The same was true for the mean correlations found with posttest knowledge and normalized knowledge gains ($z = 3.64$, $p < .001$). In contrast, the mean correlations found with absolute and normalized knowledge gains did not differ significantly ($z = 0.58$, $p = .562$).

We repeated the analyses with only those 12 studies that reported sufficient information to compute and compare all three types of dependent variables, thus holding study characteristics constant over the three types of dependent variables (see Table 1). This increased the differences and led to mean effect sizes ranging from $r_{P}^{+} = .760$ for posttest knowledge over $r_{NG}^{+} = .123$ for normalized knowledge gains to $r_{AG}^{+} = -.398$ for absolute knowledge gains. These results indicate that, independently of other study characteristics, the correlations of prior knowledge with posttest knowledge and with knowledge gains differ statistically, capture partly independent aspects of learning, and need to be analyzed separately. In the following subsections, we thus report the moderator analyses separately for posttest knowledge and knowledge gains.

8.4.4 Research Question 4: Moderators of the Correlations

Moderators of the correlation with posttest knowledge. Our fourth research question concerned how strongly the effect of prior knowledge on posttest knowledge and on knowledge gains is moderated by knowledge-related, learner-related, and environment-related variables. Table 3 shows the overall mean correlation between prior knowledge and posttest knowledge and how this overall effect was moderated by third variables. When the levels of moderators were coded (see SOM2), but are not listed in Table 3, this indicates that no study or only one study reported the respective level of the moderator. If the confidence intervals or results of tests for moderator effects are not listed in Table 3, this indicates that the available evidence was too limited to permit the analysis. The moderator levels for which the evidence was so limited that no confidence intervals could be computed were excluded from the significance tests performed by the meta-regression moderator analyses.

Knowledge-related moderators. Among the knowledge-related moderators, knowledge type had a significant effect and explained a variance proportion of R^2 = .037 of the correlations between prior knowledge and posttest knowledge. No moderator effects were found for the knowledge subtype, the broad content area, and the content domain. The similarity of the prior knowledge measure and the posttest measure significantly moderated the correlation between prior knowledge and posttest knowledge. Separate bivariate regressions showed that the correlation was significantly related to the similarity of the dimensions knowledge type (R^2 = .031), content domain (R^2 = .036), social context of the assessment (R^2 = .008), and modality of the assessment (R^2 = .026). In all of these cases, the correlation was stronger for a high similarity and lower, but still significantly greater than zero for a low similarity. When all similarity dimensions were simultaneously entered as predictors into a multiple regression, they explained a variance proportion of R^2 = .021. This relatively small value might be due to the fact that the similarity of two pieces of knowledge is difficult to quantify.

Learner-related moderators. Participant age did not moderate the correlation, but educational level was a significant moderator (R^2 = .059). The association between prior knowledge and posttest knowledge was significantly stronger for primary education than for higher education.

Environment-related moderators. Country, intervention, and intervention duration did not moderate the correlation. The cognitive demands of the interventions significantly moderated the correlation between prior knowledge and posttest knowledge (R^2 = .056). The correlation was lower for interventions with lower cognitive demands and higher for interventions with higher cognitive demands. Only one of the nine instructional methods listed in Table 3 moderated the correlation: Interventions including problem-based learning had a smaller correlation than interventions not employing this instructional method (R^2 = .012). All the instructional methods entered together as predictors in a multiple regression had no statistically significant effect.

Methodological moderators. We additionally conducted the following exploratory analyses of methodological moderators. The correlation between prior knowledge and posttest knowledge was independent of whether group differences or individual differences in prior knowledge were used to predict posttest knowledge. The correlation increased with the numbers of items used in the prior knowledge test (R^2 = .036) and the posttest (R^2 = .025). A likely explanation is that longer tests typically increase the precision of measurement by reducing the random noise in the data. The item response format and whether the learning outcome test was a posttest (directly after the intervention) or a retention test (at a later time after the posttest) did not moderate the correlation. The correlations of prior knowledge and posttest knowledge were significantly higher when the exact same test was used at both measurement points than when different tests were used (R^2 = .149). Whether prior knowledge was used to predict a knowledge measure or an achievement measure and whether it was used to predict knowledge in one domain or multiple domains at posttest did not make a statistically significant difference.

***Differences between correct and incorrect knowledge*.** In the analyses reported in Tables 1 and 2, we tested how the amount of prior knowledge predicted the amount of subsequent knowledge and achievement, irrespective of the correctness of this knowledge. Here, we tested whether the correctness of knowledge (i.e., misconceptions and error scores vs. correct concepts and solution rates) influenced the direction of the correlation. We gave all scores quantifying the amount of incorrect knowledge a negative sign and all scores quantifying the extent of correct knowledge a positive sign and recoded all the effect sizes accordingly. Table 2 displays the results of the analyses. As expected, the amount of correct prior knowledge positively correlated with the amount of correct posttest knowledge and negatively correlated with the amount of incorrect posttest knowledge. The amount of incorrect prior knowledge positively correlated with the amount of incorrect posttest knowledge and negatively correlated with the amount of correct posttest knowledge. The correctness of knowledge explained a variance proportion of R^2 = .190 of the effect sizes. We repeated the analysis with the absolute values of the effect sizes, thus ignoring the signs and comparing only the strength of the correlations. This analysis did not indicate significant differences between the four correlations. Thus, the correlations of correct and incorrect prior knowledge with correct and incorrect posttest knowledge differed in their directions, but not in their strengths.

Table 2

Meta-Analytic Results for Measures of Correct vs. Incorrect Prior Knowledge and Posttest Knowledge (see main text for details)

	j	k	r_P^+	CI r_P^+ 95%	τ^2	I^2	Moderator	
							Sign.	R^2
Overall	228	4149	.507	[.475, .538]	.128	95.00		
Measures							**	.190
Correct prior knowledge; correct posttest knowledge	225	4037	.524	[.494, .553]	.118	94.63	**	
Correct prior knowledge; incorrect posttest knowledge	9	82	-.557	[-.746, -.287]	.156	90.98	Ref	
Incorrect prior knowledge; correct posttest knowledge	6	18	-.457	[-.587, -.303]	.023	66.22	**	
Incorrect prior knowledge; incorrect posttest knowledge	4	12	.578	-	-	-		

Moderators of the correlation with knowledge gains. We had planned to conduct the same moderator analyses for absolute and normalized knowledge gains as we had done for posttest knowledge. However, only a few studies reported the information required to compute the correlation between prior knowledge and absolute or normalized knowledge gains. These few effect sizes had a high degree of heterogeneity. Consequently, only a few moderator analyses were possible, the statistical power of these analyses was low, the resulting confidence intervals were large, and almost no moderator effect reached statistical significance. SOM 4 lists the mean values for the levels of the moderators that could be coded and the results of the significance tests that were possible despite the limited amount of data.

The only moderator that had significant effects on the correlations with both absolute and normalized knowledge gains was the cognitive demands of the intervention. The correlation was lower for interventions with low cognitive demands and higher for interventions with high cognitive demands. This difference explained variance proportions of R^2 = .281 for absolute knowledge gains and R^2 = .366 for normalized knowledge gains. These moderator effects were extremely strong (cf. J. Cohen, 1992), which explains why the effect sizes found with knowledge gains were so heterogeneous and close to zero and why the results of other moderator analyses not accounting for the cognitive demands of the intervention did not reach statistical significance. Additionally, the correlation between prior knowledge and absolute knowledge gains was higher when one test was used to distinguish high- and low prior knowledge learners and a second test to assess the knowledge gains as compared to only one test being used for both purposes (R^2 = .110).

Table 3

Mean Effect Sizes and Moderator Analyses for the Correlation of Prior Knowledge with Posttest Knowledge

	j	k	r_P^+	CI r_P^+ 95%	τ^2	I^2	Moderator	
							Sign.	R^2
Overall	235	4223	.525	[.495, .553]	.099	93.53	-	-
Knowledge characteristics								
Knowledge type[a]							*	.037
Declarative	159	2125	.513	[.478, .548]	.205	95.68	Ref	
Procedural	10	51	.645	[.468, .774]	.117	91.83	*ns*	
Declarative and procedural mixed	74	645	.619	[.566, .667]	.104	95.51	**	
Knowledge subtype[a]							*ns*	.016
Declarative: facts	41	268	.580	[.507, .645]	.153	94.28	*ns*	
Declarative: concepts	128	1244	.528	[.487, .566]	.220	96.21	Ref	
Procedural: cognitive skill	10	51	.645	[.468, .774]	.117	91.83	*ns*	
Broad content area[a]							*ns*	.002
STEM	87	578	.566	[.509, .619]	.112	943.81	*ns*	
Language	99	3055	.540	[.504, .575]	.072	92.11	Ref	
Humanities	5	19	.355	-	-	-	-	
Social sciences	17	107	.570	[.390, .708]	.197	93.69	*ns*	

Health sciences	2	9	.439	-	-	-	-	
Sports	5	54	.533	-	-	-	-	
Content domain[a]							*ns*	.022
Mathematics	37	416	.601	[.521, .670]	.087	94.23	**	
Physics	16	55	.546	[.418, .653]	.139	93.85	*	
Chemistry	6	20	.352	[.149, .528]	.054	97.42	Ref	
Biology	21	62	.624	[.474, .739]	.250	91.51	**	
Geosciences	6	19	.482	[.002, .782]	.245	92.83	*ns*	
Medicine and nursing	2	9	.439	-	-	-	-	
Psychology	14	99	.546	[.401, .665]	.150	92.41	*	
First language	94	2715	.537	[.499, .573]	.076	92.20	**	
Second language	10	53	.710	[.610, .789]	.076	87.52	**	
Sports	5	54	.533	-	-	-	-	
History	3	10	.457	-	-	-	-	
Other	3	20	.503	-	-	-	-	
Similarity of prior knowledge and learning outcome								
Similarity of the knowledge type							**	.031
Low	117	1398	.441	[.409, .472]	.053	91.39	Ref	
High	210	2821	.547	[.515, .578]	.162	95.63	**	

Similarity of the content domain							**	.036
Low	72	688	.419	[.364, .472]	.044	89.82	Ref	
High	209	3535	.548	[.518, .578]	.103	94.01	**	
Similarity of the physical context							*ns*	.000
Low	7	94	.566	[.373, .712]	.230	96.40	*ns*	
High	228	4071	.528	[.498, .558]	.134	95.30	Ref	
Similarity of the temporal context							*ns*	.000
Low	184	3974	.529	[.496, .560]	.096	94.16	*ns*	
High	55	248	.515	[.441, .581]	.129	88.13	Ref	
Time between assessments	171	3908					*ns*	.000
Similarity of the social context							*	.008
Low	15	185	.451	[.357, .535]	.037	88.43	Ref	
High	222	3909	.533	[.502, .561]	.112	93.30	**	
Similarity of the modality							**	.026
Low	46	650	.437	[.381, .490]	.084	92.86	Ref	
High	205	3064	.538	[.505, .570]	.155	95.56	**	
Multivariate effect of all similarity dimensions	204	3397					*	.021

Learner characteristics								
Age	136	3362					*ns*	.000
Educational level							**	.059
Daycare	5	139	.377	-	-	-	-	
Kindergarten/preschool	57	1790	.458	[.411, .504]	.060	90.16	*ns*	
Primary education	71	1676	.584	[.542, .624]	.064	93.80	**	
Secondary education	49	228	.543	[.455, .621]	.125	92.87	*ns*	
Higher education	57	358	.457	[.392, .518]	.108	90.46	Ref	
Continued education	5	22	.753	-	-	-	-	
Several	3	10	.743	-	-	-	-	
Environmental characteristics								
Intervention								
Intervention Setting							*ns*	.001
No intervention	14	150	.487	[.356, .599]	.073	83.16	Ref	
School instruction only	78	2772	.511	[.461, .557]	.106	96.17	*ns*	
School instruction and other intervention	57	867	.513	[.451, .570]	.075	92.49	*ns*	
Other intervention only	85	397	.548	[.489, .602]	.145	89.04	*ns*	
Intervention Duration							*ns*	.008

0–2 hours	73	352	.515	[.452, .573]	.141	88.93	*ns*	
2–24 hours	7	29	.497	[.197, .711]	.152	83.31	Ref	
2–7 days	5	16	.516	-	-	-	-	
> 1 week	57	889	.561	[.496, .621]	.086	93.60	*ns*	
Cognitive demands of intervention							*	.056
Lower	39	100	.447	[.329; .551]	.162	94.42	Ref	
Higher	38	96	.656	[.543; .745]	.281	90.06	**	
Instructional methods in intervention								
Written instruction							*ns*	.007
No	23	226	.571	[.464, .662]	.171	90.93	*ns*	
Yes	101	463	.519	[.469, .565]	.101	87.04	Ref	
Oral instruction							*ns*	.016
No	85	495	.504	[.452, .553]	.106	85.98	Ref	
Yes	36	253	.584	[.495, .660]	.163	92.23	*ns*	
Multimedia instruction							*ns*	.012
No	70	489	.496	[.438, .550]	.098	85.44	Ref	
Yes	43	149	.585	[.508, .653]	.138	87.66	*ns*	
Practice							*ns*	.006
No	75	440	.516	[.457, .570]	.114	87.43	Ref	

Yes	47	302	.563	[.485, .633]	.147	91.91	*ns*	
Constructive activities							*ns*	.017
No	81	501	.504	[.451, .555]	.098	83.64	Ref	
Yes	45	158	.608	[.516, .685]	.207	92.22	*ns*	
Technology							*ns*	.018
No	78	521	.505	[.449, .557]	.102	86.72	Ref	
Yes	41	127	.630	[.525, .717]	.223	91.94	*ns*	
Feedback							*ns*	.024
No	94	561	.520	[.469, .569]	.124	87.71	Ref	
Yes	18	65	.640	[.476, .761]	.207	91.74	*ns*	
Collaborative learning							*ns*	.020
No	98	569	.514	[.466, .559]	.104	87.16	Ref	
Yes	22	87	.627	[.478, .742]	.228	92.65	*ns*	
Problem-based learning							*	.012
No	96	583	.550	[.495, .600]	.132	88.22	*	
Yes	17	93	.451	[.371, .524]	.073	87.03	Ref	
Multivariate effect of all instructional methods	98	582					*ns*	.088
Country[b]							*ns*	.088
Australia	3	29	.375	-	-	-	-	

Belgium	2	80	.231	-	-	-	-	
Canada	5	115	.715	[.526, .837]	.142	91.90	**	
Peoples Republic of China	5	300	.482	[.401, .556]	.108	93.55	**	
Hong Kong	6	168	.380	[.274, .477]	.071	92.80	Ref	
Taiwan	3	9	.675	-	-	-		
Denmark	2	7	.371	-	-	-		
United Kingdom	12	627	.508	[.408, .595]	.074	89.96	*	
Finland	8	322	.425	[.262, .566]	.088	92.45	*ns*	
France	2	122	.333	-	-	-	-	
Germany	9	253	.491	[.223, .690]	.141	97.91	*ns*	
Israel	4	49	.471	-	-	-	-	
Netherlands	5	163	.643	[.180, .872]	.216	96.45	*ns*	
New Zealand	2	42	.814	-	-	-	-	
Norway	3	172	.557	-	-	-	-	
Spain	2	5	.467	-	-	-	-	
USA	52	904	.518	[.450, .580]	.123	96.45	**	
Methodological study characteristics								
Study design							*ns*	.002
Group differences in prior knowledge (quasi-experimental or	83	403	.549	[.488, .604]	.140	88.12	*ns*	

experimental design)								
Individual differences in prior knowledge (correlational design)	152	3820	.514	[.480, .547]	.093	94.84	Ref	
Number of items in prior knowledge measure	186	3053					**	.036
Number of items in learning outcome measure	189	3088					*	.025
Response format[a]							*ns*	.019
Open	101	1368	.521	[.472, .567]	.124	92.85	Ref	
Fill-in	7	34	.632	[.562, .693]	.015	58.05	*	
Single or multiple choice	66	508	.556	[.501, .606]	.096	90.52	*ns*	
Other	6	30	.528	[.312, .692]	.117	94.85	*ns*	
Various	18	99	.619	[.506, .711]	.166	95.51	*	
Retention test							*ns*	.000
No	235	4159	.524	[.495, .553]	.099	93.57	*ns*	
Yes	13	64	.507	[.366, .626]	.077	83.03	Ref	
Same test for prior knowledge and learning outcome							**	.149
No	201	3606	.473	[.444, .501]	.070	91.27	Ref	
Yes	104	615	.681	[.644, .714]	.180	96.68	**	
Measures at T2								

Outcome at T2							*ns*	.002
Knowledge	228	4169	.528	[.498, .556]	.079	91.63		
Achievement	10	54	.438	[.209, .621]	.352	99.14		
Domain specificity							*ns*	.007
Specific domain	232	4179	.520	[.491, .548]	.081	92.11		
Various domains	8	44	.591	[.171, .831]	.302	98.95		

[a] Only for effects in which the moderator had the same level at T1 and T2. For effects with differing moderator levels, see the results category “Similarity of prior knowledge and learning outcome.”
[b] Only coded for studies including school instruction.
Note. * $p < .05$; ** $p < .01$; *ns* = nonsignificant; ref: used as reference category in dummy coding. Missing mean effect sizes, confidence intervals, heterogeneities, p values, and R^2 values indicate that the respective analysis was not possible because the number of effect sizes was too small to permit the analysis. Moderation analyses were only calculated for the levels of the moderating variables where the confidence intervals are presented, respectively.

8.5 Discussion

In this meta-analysis, we estimated the effect of the amount of domain-specific prior knowledge on learning. We first discuss the main findings regarding the four research questions, followed by the limitations of our study, and the implications of the findings.

8.5.1 Main Findings

Available empirical evidence. Our first research question concerned the amount of available empirical evidence related to the relation between prior knowledge and learning. We found a strong imbalance between a high number of studies investigating the correlation between prior knowledge and posttest knowledge and a low number of studies investigating the correlation between prior knowledge and knowledge gains. Even though we scanned 5462 studies for inclusion, we could include only 20 studies with information on absolute knowledge gains and 14 studies with information on normalized knowledge gains. None of these studies was a randomized controlled trial, and none allowed controlling for intelligence. Due to the small number of studies on knowledge gains and the high heterogeneity of their findings, the main effects and some moderator effects could only be estimated with low precision, and many moderator effects could not be tested at all. In sum, in spite of decades of research on knowledge acquisition, very little is known about how strongly prior knowledge causally affects the learners' increases in knowledge.

There are at least four explanations for the scarcity of studies on knowledge gains. First, a relatively large group of studies in our meta-analysis compared the effectivity of instructional conditions, measured knowledge only once, and then used it as covariate. Second, the included longitudinal studies frequently used different knowledge tests at different time points to keep the tests age appropriate so that no gain scores could be computed. Third, some studies reported multivariate analyses of knowledge changes over time, but failed to report the zero-order correlations between prior knowledge and subsequent knowledge required for the meta-analytic aggregation. Finally, gain scores were once thought to have undesirable statistical characteristics, for example, chronically low reliabilities. Newer studies (e.g., Zimmermann & Williams, 1998) found these conclusions to be based on unrealistic statistical assumptions (e.g., the same variance at pretest and posttest, which is often not the case, as shown in Figure 2). Under more realistic assumptions, gain scores can have acceptable validities and reliabilities (Maris, 1998; May & Hittner, 2010). These four reasons may have resulted in the lack of experiments investigating the causal effect on prior knowledge on knowledge gains, even though such studies are possible and needed.

Correlations of prior knowledge with posttest knowledge. Our second research question related to how strongly prior knowledge is associated with posttest knowledge, that is, how stable individual differences in knowledge are over time. Averaged over 4223 effect sizes from 235 studies, we found a high stability of $r_P^+ = .525$. Accordingly, the learners' rank order with respect to the amount of their knowledge tended to remain relatively unchanged between measurement points. The estimation had high precision, as indicated by the small 95% confidence interval ranging from .478 to .548.

The correlation between prior knowledge and posttest knowledge decreased only slightly and remained statistically significant when we analyzed the nine randomized controlled trials manipulating the amount of prior knowledge before a learning phase. This finding indicates a causal connection between prior knowledge and posttest knowledge. In line with this, the correlation between prior knowledge and posttest knowledge decreased only marginally when controlling for intelligence. Since intelligence influences the acquisition of prior knowledge and the acquisition of new knowledge (Schneider & McGrew, 2012, p. 123), it is surprising that controlling for intelligence did not more strongly reduce the effect sizes. One possible explanation is that many studies that included an intelligence measure were studies on language development that measured nonverbal intelligence (Alonzo, Yeomans-Maldonado, Murphy, & Bevens, 2016; Yeung, Ho, Chan, & Chung, 2016). Due to this mismatch of verbal knowledge assessments and nonverbal intelligence measures, the included studies might have underestimated the true correlation between knowledge and intelligence. This highlights the need for more differentiated research on the correlations between different types of knowledge and the subfactors of intelligence. Still, the randomized controlled trials in our meta-analysis show that prior knowledge affects posttest knowledge at least partly independently of intelligence.

Our finding of high stability for individual differences in knowledge matches the findings from studies on other cognitive learning outcomes, namely, expert performance and academic achievement. Like individual differences in knowledge, individual differences in expert performance have been found to be highly stable over time (Ericsson & Lehmann, 1996). Previous meta-analyses on achievement (Bretz Jr., 1989; P. A. Cohen, 1984; Duncan et al., 2007; Kuncel et al., 2001; La Paro & Pianta, 2000; Samson, Graue, Weinstein, & Wahlberg, 1984; Schuler et al., 1990; Trapmann, Hell, Weigand, & Schuler, 2007) have found correlations of prior achievement with later achievement ranging from r = .10 (Duncan et al., 2007) to r = .48 (La Paro & Pianta, 2000). Our average correlation found for knowledge lies slightly above this range, possibly because knowledge is a less heterogeneous concept than achievement and is thus easier to predict over time. A second explanation is that achievement is mostly measured in longitudinal studies with months or years between measurement points, whereas prior knowledge is also investigated in smaller intervention studies where pretest and posttest were just hours or days apart (e.g., Fyfe & Rittle-Johnson, 2016).

The meta-analysis revealed several moderators of the correlation between prior knowledge and posttest knowledge (research question 4). Statistically significant knowledge-related moderators were the knowledge type and the similarity of prior knowledge and knowledge to be learned with respect to their content domain, knowledge type, modality, and social context. The only significant learner-related moderator was educational level, where the correlation was higher for primary education than for higher education. Among the environment-related moderators, only the cognitive demands of the intervention and the use of problem-based learning reached statistical significance. The correlation was lower for lower cognitive demands than for higher cognitive demands and higher for instruction not including problem-based learning than for instruction including problem-based learning.

Overall, the correlation between prior knowledge and posttest knowledge proved to be high and stable. It did not differ between facts, concepts, and cognitive skills, nor between countries, content domains, and intervention settings (i.e., school instruction vs. other instruction). Prior knowledge predicted achievement as well as it predicted knowledge. Most statistically significant moderator effects were weak, and the correlation was significantly greater than zero for all levels of moderators for which confidence intervals could be estimated.

Correlation of prior knowledge with knowledge gains. Our third research question was about how strongly prior knowledge correlates with absolute and normalized knowledge gains. As expected, the correlation of prior knowledge with normalized knowledge gains was smaller than the correlation found for posttest knowledge. The small correlation for normalized knowledge gains can be explained by the fact that the knowledge gains made during an instructional intervention depend on many influences, prior knowledge just being one of them (Ormrod, 2011). The correlation for absolute knowledge gains was descriptively even lower than the one found for normalized knowledge gains, which demonstrates the fact that absolute knowledge gains implicitly discriminate against high prior knowledge and can thus lead to an underestimation of the correlation (Hake, 1998). Neither the correlation for absolute knowledge gains nor the correlation for normalized knowledge gains reached statistical significance.

One main reason for the lack of statistical significance of the correlation between prior knowledge and knowledge gains was the strong moderating effect of the cognitive demands of the intervention (research question 4). This moderator effect was statistically significant and explained the large variance proportions indicating that the correlations were significantly lower for instruction with lower cognitive demands relative to instruction with higher cognitive demands. In absolute terms, descriptively, we found a strong compensatory effect (i.e., a negative correlation) for low cognitive demands and a weak Matthew effect (i.e., a positive correlation) for high cognitive demands. Due to the small number of studies investigating knowledge gains and the resulting lack of statistical power, only the negative correlation between prior knowledge and absolute knowledge gains for low instructional support was statistically different from zero.

Figure 5 displays how the correlations of prior knowledge with posttest knowledge (see Table 3), the correlations of prior knowledge with normalized knowledge gains (see SOM 4), and the Cohen's ds for the pretest-posttest increases in knowledge look combined for all respective effect sizes in the meta-analysis (left), for instructional conditions with low cognitive demands (middle), and for instructional conditions with high cognitive demands (right). Our meta-analysis was not designed to compare the effectiveness of instruction with high versus low cognitive demands. Therefore, the Cohen's d values cannot be compared with each other and are not interpreted here.

A possible explanation for the significant moderation effect of the cognitive demands is that, for learners with high prior knowledge, high cognitive demands are desirable difficulties (E. L. Bjork & Bjork, 2014) that challenge them to elaborate on and extend their knowledge, whereas learners with low prior knowledge lack the prerequisites to do so. Conversely, learners with low prior knowledge can profit from instruction with low cognitive demands, because it leaves larger amounts of working memory capacity free, which can then be used to interpret and elaborate on the new knowledge. Learners with high prior knowledge might not profit from instruction with low cognitive demands either because they already know the content or because they need much less interpretation and elaboration to understand and encode the new knowledge (Kalyuga, Ayres, Chandler, & Sweller, 2003).

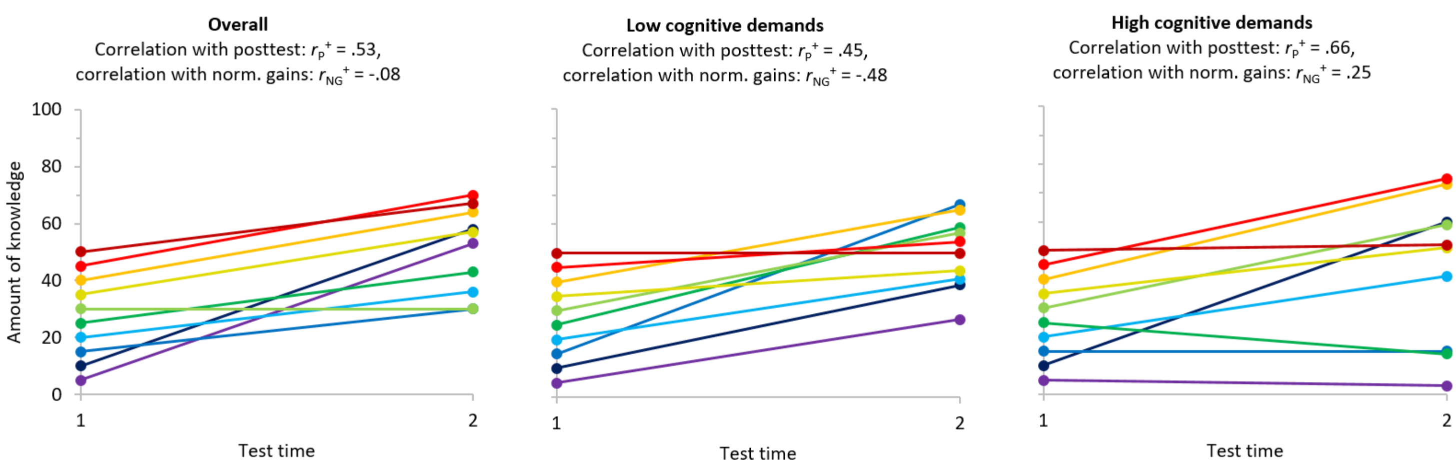

Figure 5. Correlations of prior knowledge (T1) with posttest knowledge (T2) and normalized knowledge gains found in the meta-analysis overall (left), for instruction with low cognitive demands (middle), and for instruction with high cognitive demands (right), visualized using 10 fictitious persons' amounts of knowledge as example

8.5.2 Limitations

This meta-analysis has at least three limitations. First, even though it included a broad selection of studies, some content domains (e.g., medicine), age groups (daycare), and countries (in Africa and Asia) are underrepresented. This limits the generalizability of our results. Second, even though we scanned 5462 search hits for inclusion in our meta-analysis, the field of research on knowledge acquisition is so broad that some relevant studies were surely not included in our meta-analysis. However, we see no reason to expect systematic differences between the 4327 included effect sizes and any possibly existing effect sizes that were not included. Third, we focused on the quantity of prior knowledge only, but sometimes the quality of knowledge might be more predictive for future learning. Qualitative changes in knowledge structures are usually investigated using interview data or multivariate statistical models, for example, latent transition analyses (Flaig et al., 2018; Hickendorff et al., 2017; M. Schneider & Hardy, 2013). However, the complex patterns of results from these studies cannot be meta-analyzed as easily as single effect sizes. Our meta-analysis might thus underestimate the size of the effect of prior knowledge on learning even though we already found high effect sizes.

8.5.3 Implications

The stability of individual differences in knowledge. One implication of our study is that individual differences in knowledge are highly stable over the course of learning. The effect size of r_P^+ = .525 found in our meta-analysis is equivalent to a Cohen's *d* of 1.23. Hattie (2009) reported a meta-analytic rank order of 138 variables associated with academic achievement, which did not include domain-specific prior knowledge. When included, it would be among the strongest three effect sizes of all 138 effects in the rank order. This demonstrates that domain-specific prior knowledge predicts individual differences in knowledge and achievement after learning better than almost all other variables, which supports the knowledge-is-power hypothesis (Hambrick & Engle, 2002; Möhring et al., 2018). Assessments of prior knowledge in school entrance tests or in formative assessments, for example, can thus provide valuable information to teachers, parents, and the learners themselves. The high stability of individual differences in domain-specific knowledge suggests that the differences aggregated over months or years of learning and thus are hard to change during short time periods, for example, in instructional interventions. This supports the theoretical approaches emphasizing the long-term nature of domain-specific knowledge acquisition, for example, learning-trajectory approaches (Clements &

Sarama, 2004) and theories of strategy change (Siegler & Svetina, 2002), skill building (Bailey, Duncan, Watts, Clements, & Sarama, in press), conceptual change or conceptual development (diSessa et al., 2004; Keil, 1996; Vosniadou, 1994), and the acquisition of expert performance (Ericsson & Charness, 1994; Ullén et al., 2015).

Differences between posttest knowledge and knowledge gains. One central implication of our study for further research is the necessity to analyze both the effect of prior knowledge on posttest knowledge and the effect of prior knowledge on knowledge gains. These two effects are partly statistically independent of each other, reflect different aspects of the learning process, and are empirically of different magnitudes. Most previous studies were limited in that they investigated correlations of prior knowledge and posttest knowledge. There is a need for randomized controlled trials of knowledge gains and longitudinal studies carefully controlling for confounding variables.

Domain specificity of knowledge. Prior knowledge had medium to strong effects on posttest knowledge in all investigated content domains. However, these effects were domain specific, that is, the effect sizes were higher when prior knowledge and posttest knowledge were from the same domain than when they were from different domains. This finding converges with the widespread notion that the beneficial effects of domain-specific knowledge are domain specific and that cross-domain transfer is difficult to achieve (Detterman & Sternberg, 1993; Hirschfeld & Gelman, 1994; Sala & Gobet, 2017). However, the meta-analytic correlations were still significantly greater than zero when there was some dissimilarity between prior knowledge and posttest knowledge. Among possible explanations of this finding are (a) near transfer , (b) confounding influence of third variables not controlled for (e.g., socio-economic status or meta-cognition) on prior knowledge and posttest knowledge, and (c) the difficulty of quantifying the similarity of knowledge measures validly (cf. Barnett & Ceci, 2002; Sala & Gobet, 2017).

Knowledge and achievement. The meta-analytic results indicate that prior knowledge predicts subsequent achievement as well as it predicts subsequent knowledge. The strong correlation of $r^+ = .438$ between knowledge before learning and achievement after learning demonstrates how closely the two constructs are related. However, the two constructs differ in that achievement measures assess the learning outcomes of instruction of months or years and usually include several sub-domains, subskills, or competencies (OECD, 2016; Steinmayr et al., 2014). In comparison, knowledge is a more homogeneous construct that can be changed within relatively short time frames by relatively simple interventions. This makes it easier to identify the sources of knowledge than it is to identify the sources of achievement in randomized controlled trials. Thus, it might be productive to trace achievement back to the underlying knowledge structures and to trace back these knowledge structures to the experiences and instructional practices that gave rise to their construction. In short, understanding knowledge acquisition can also help to better understand achievement.

Matthew effects and compensatory effects. Our results also shed light on the debate on the Matthew effect and the compensatory effect in learning. Some previous studies have found evidence in favor of a Matthew effect in learning (Duff et al., 2015; Pfost et al., 2011), whereas others found no such effect or even a compensatory effect (Baumert et al., 2012; Schroeders et al., 2016). Our moderator analyses can explain this heterogeneity (see Figure 5). Specifically, they show that the compensatory effect (a negative correlation between prior knowledge and knowledge gains) occurs mostly in instruction with low cognitive demands, in which students memorize facts, follow known procedures, and practice routine problems. For the Matthew effect, the results were inconclusive due to a lack of statistical power. Overall, teachers can influence the achievement gap between their students depending on how they design their instruction. The cognitive demands of the intervention moderated the effect sizes in our meta-analysis in a much stronger way than the actual instructional methods (oral instruction, collaborative learning, instructional technology, etc.). This demonstrates that cognitive demands are not inherent to instructional methods, but that each method can be implemented in cognitively more or less demanding ways.

Knowledge construction as moderated mediation. The studies synthesized in this meta-analysis focused on prior-knowledge effects and moderators of this effect. However, as explained in the introduction, there is more than one effect of prior knowledge on learning because prior knowledge affects a number of learning processes, each of which can positively or negatively affect the construction of new knowledge. For example, prior knowledge might affect attention, encoding, chunking, strategy discovery, problem-solving flexibility, the Einstellung effect, intrusions and interferences, and negative transfer (Bilalić et al., 2010; Dochy et al., 1999; M. Fisher & Keil, 2016; Hambrick & Engle, 2002; Luchins & Luchins, 1987; W. Schneider, 1993; Stern, 2001). Each of these processes might affect the learning outcomes in a different way. In this sense, these variables mediate the effect of prior knowledge on learning outcomes.

Since there is more than one effect of prior knowledge on learning outcomes, the moderators of prior-knowledge effects can only be understood when it is known which of the mediating processes they affect. For example, the use of problem-based learning as a teaching method might moderate how strongly attention mediates the effects of prior knowledge on learning outcomes. Alternatively, it might moderate how strongly elaboration mediates the effects of prior knowledge on learning. Tests of moderated mediations comprise an active field of ongoing methodological research (Preacher et al., 2007), and it might be productive to employ these tests in further research on prior-knowledge effects on learning. This could also help with the integration of the findings on prior-knowledge effects from different subdisciplines of the learning sciences because it allows researchers to categorize influencing variables as mediators (e.g., elaboration), direct moderators of the mediation (e.g., instructional methods stimulating elaboration), or indirect moderators of the mediation (e.g. teacher training affecting teachers' use of instructional methods). In the following section, we offer a framework for further research on knowledge acquisition, which incorporates this idea.

A Framework for Further Research on Knowledge Acquisition

The present meta-analysis demonstrates the need for the integration of the many findings on how prior knowledge affects learning. Researchers from different subdisciplines have used varieties of research paradigms in hundreds of studies. This flood of information makes it hard to detect relations, imbalances, and open questions. We propose the *Learner-enviroment Interaction in KnowledgE construction* (LIKE) framework. The framework is based on the insight that the influence of prior knowledge on learning is a case of a moderated mediation (Preacher et al., 2007), in which intra-learner processes mediate the effect of prior knowledge on cognitive learning outcomes, such as knowledge and achievement. Direct moderators of these mediating processes are learner resources and the learner-environment interaction, which is influenced by environmental resources as an indirect moderator.

Figure 1 shows the structure of the framework. The lists in each box provide examples and are not exhaustive. The framework is based on the following five assumptions:

1. In research on prior knowledge, it is useful to distinguish between the levels of knowledge (bottom level in Figure 1), the learner (middle), and the environment (top). It is also useful to distinguish between dynamic processes (right) and the comparably more stable resources (left) involved in knowledge construction. By *stable*, we mean that the resources usually change at a much slower rate than knowledge. Thus, their influence on knowledge construction processes is probably stronger than vice versa. This is indicated by the one-

headed arrows leading away from the resources, for example, from learner resources to intra-learner processes.

2. Of all the framework components, only intra-learner processes directly mediate the relation between prior knowledge and learning outcomes because only the individual learner can recall prior knowledge from long-term memory and can encode new knowledge in long-term memory.
3. Learner resources moderate the relation between prior knowledge and learning, but only to the extent to which they affect the processing of knowledge in the learner. In methodological terms, learner resources are direct moderators of the mediating processes between prior knowledge and learning outcomes.
4. Environmental resources (e.g., preprepared learning environments such as for school instruction) can affect the relation between prior knowledge and learning outcomes, but only to the extent to which the learner interacts with these environments and to the extent to which these interactions affect the processing of knowledge in the learner. Therefore, learning environments can only provide learning opportunities, but cannot guarantee what and how much is learned. Instead, the learning outcomes depend on how the learner adopts the opportunities . From a methodological view, environmental resources are indirect moderators, and the learner-environment interaction is a direct moderator of the mediation paths between prior knowledge and learning outcomes.
5. Knowledge is a component of achievement (OECD, 2016; Schrader & Helmke, 2015; Steinmayr et al., 2014) and related cognitive-learning outcomes (Ericsson & Charness, 1994). Therefore, prior knowledge and the other parts of the framework not only affect the acquisition of knowledge but also the acquisition of cognitive learning outcomes in general, including but not limited to achievement, competence, and expertise.

The LIKE framework describes how the use of prior knowledge is embedded in the wider context of learning and instruction. It is not meant to be a directly testable model of specific causal relations. Instead, it is a heuristic framework that can be used to plan studies and develop hypotheses about specific causal relations to explain empirical results and their interrelations, to summarize and integrate findings in reviews and meta-analyses, and to identify underresearched topics and open questions for future research. Whereas other frameworks have been published that partially included similar assumptions (e.g., Biggs, 1993, p. 19; Schrader & Helmke, 2015; Ullén et al., 2015), the defining key characteristic of the LIKE framework is that it describes the effect of prior knowledge on learning as moderated mediation, which clarifies the different ways in which variables influence learning. The relational structure of the framework demonstrates that none of these processes can be understood in isolation and emphasizes the importance of future randomized controlled trials and careful process analyses to investigate the causal processes mediating the effect of prior knowledge on learning and how these are moderated by knowledge, learner, and environmental characteristics.

8.6 Conclusion

Many studies have investigated the role of prior knowledge in learning, but most of them have only reported how prior knowledge correlates with posttest knowledge. Our meta-analysis demonstrates that this correlation is positive and strong, in general, as well as for many levels of knowledge-, learner-, and environment-related moderators. Thus, individual differences in knowledge are highly stable, so prior knowledge, measured before learning, has high predictive power for knowledge and achievement after learning. Randomized controlled trials indicate that this relation is causal rather than merely correlational. Only a few studies have investigated the association of prior knowledge with knowledge gains. Overall, this correlation was zero. However, it was significantly higher for interventions with high cognitive demands than for interventions with low cognitive demands. We found a compensatory effect (i.e., a narrowing gap between learners’ amounts of knowledge over time) only for instruction with low cognitive demands. Future research might benefit from using more randomized controlled trials to examine the effects of prior knowledge on knowledge gains and on posttest performance and to systematically evaluate the mechanisms that mediate and moderate the influence of prior knowledge on learning.

9. Study 2: Conceptual Change and Knowledge Integration as Learning Processes in Higher Education: A Latent Transition Analysis

9.1 Abstract

Conceptual change, that is, a restructuring of incompatible prior knowledge, is a well investigated learning mechanism in school children's acquisition of new concepts. An understanding of academic concepts is also a central learning goal of higher education. However, there is almost no research on conceptual change and knowledge integration in higher education. We tracked 137 undergraduate psychology students' concepts of human memory longitudinally over four semesters. A latent profile transition analysis (LPTA) showed that the students' development followed six transition paths between four knowledge profiles. These developmental pathways were well-ordered, indicated a general trend from fragmented knowledge to integrated scientific knowledge, and correlated with the students' university grades and with an additional test of memory understanding. The findings highlight the importance of conceptual change, in particular, knowledge integration, in higher education and exemplify the usefulness of LPTA for modeling individual differences in longitudinal changes of multidimensional knowledge structures.

9.2 Introduction

Students' understanding of academic concepts, for example, force in Physics, supply and demand in Economy, or human memory in Psychology, is a central learning goal of higher education. Conceptual understanding is a cornerstone of professional expertise (Tynjälä, 1999). It helps learners make predictions, explain observations, reason about the interrelations of facts, infer new knowledge, choose between alternative procedures, and construct new problem solving strategies (Goldstone & Kersten, 2003; Machery, 2010; Rittle-Johnson et al., 2015). Accordingly, the European Qualification Framework for Lifelong Learning lists "advanced knowledge of a field of work or study, involving a critical understanding of theories and principles" (European Commission, 2008, p. 12) as a central qualification for reaching a Bachelor's degree.

In many cases, conceptual change, that is, a restructuring of the learner's prior knowledge, is a necessary part of acquiring new conceptual knowledge (Carey, 1985; Posner, Strike, Hewson, & Gertzog, 1982; Vosniadou, 2008). Prior knowledge guides and constraints the interpretation of new knowledge and its encoding in memory. It often stems from observations and explanation attempts in everyday life outside formal instruction and thus can be incompatible with the scientific concepts to be learned (Carey, 1992). This explains why acquiring an understanding of academic concepts can be so difficult. Conceptual change is investigated by educational, developmental, cognitive, and philosophical scientists, who found the approach productive in content domains as diverse as physics, chemistry, biology, mathematics, medicine, and the social sciences (M. Schneider, Vamvakoussi, & Van Dooren, 2012; Vosniadou, 2008). Some of the past findings have direct implications for the design of effective learning environments, for example, school instruction (Duit, Treagust, & Widodo, 2008), professional development programs for teachers (e.g. Hewson, Tabachnick, Zeichner, & Lemberger, 1999), and projects to foster instructional quality in schools (Beeth et al., 2003).

Empirical research on conceptual change of students in higher education is virtually non-existent, despite the importance of academic concepts as learning goals in higher education. Almost all studies on conceptual change focused on students in K-12 schools or even younger children. In the present study, we examined to what extent conceptual change is still a relevant learning mechanism in higher education and whether it leads to similar developmental patterns in higher education as it does in K-12 school learning. We expected conceptual change to still be relevant in higher education, because learning by conceptual change has been described as a general human learning mechanism that is relevant independently of age group and content domain. For example, there are some similarities between children’s thinking processes when acquiring new concepts and scientists’ thinking processes when developing a new theory, which hints at an age-general importance of conceptual change (Gentner, 1997).

9.2.1 Knowledge Fragmentation and Integration

Research with school-aged children found that the fragmentation and integration of knowledge are two central component processes of conceptual change (M. Schneider & Hardy, 2013). Networks of conceptual knowledge in long-term memory can comprise types of elements, such as observations, hypotheses, explanations, analog mental models, mental images, category exemplars, and subjective theories (Goldstone & Kersten, 2003; Machery, 2010). These elements have been acquired in situation as diverse as conversations with peers, everyday-life observations, internet videos, school instruction, books, or movies. Learners do not always understand how these different and sometimes even conflicting pieces of knowledge relate to each other, and store them independently in long-term memory. This leads to fragmented knowledge.

Another source of knowledge fragmentation is the fact that storing correct concepts in long-term memory does not necessarily erase related misconceptions. Converging evidence from reaction times studies with sentence verification tasks (Potvin, Masson, Lafortune, & Cyr, 2015; Shtulman & Valcarcel, 2012), multiple choice tests (M. Schneider & Hardy, 2013), and interviews (diSessa et al., 2004) shows that naïve misconceptions and scientifically correct concepts or parts thereof can co-exist in long-term memory and do so frequently, not only in children, but also over the life-span (Shtulman & Harrington, 2016). Pieces of fragmented knowledge are activated dependent on the context (diSessa et al., 2004), so that learners do not necessarily realize when they hold pieces of knowledge in long-term memory that support or contradict each other.

Thus, the integration of pieces of knowledge into a coherent overarching knowledge structure is an important aim of instruction (M. C. Linn, 2006; M. Schneider, 2012). This can include connecting previously isolated pieces of knowledge in memory and subsuming previously unrelated concepts under a general principle. Understanding these relations can help students to differentiate better between normatively correct and incorrect ideas, thus, leading to more homogeneous and more correct knowledge. Teachers can stimulate knowledge integration by eliciting students' knowledge, adding new normative concepts, helping students to develop criteria for evaluating alternative conceptions, and by encouraging students to compare the alternatives and to sort out inadequate conceptions (M. C. Linn, 2006).

9.2.2 A Latent Profile Transition Analysis of Fragmented and Integrated Knowledge

Developmental patterns of co-existing pieces of knowledge can be investigated by latent profile transition analyses (LPTA), as demonstrated by three studies with school children on knowledge development in mathematics and science (Kainulainen et al., 2017; McMullen et al., 2015; M. Schneider & Hardy, 2013). M. Schneider and Hardy (2013) investigated third-graders' understanding of floating and sinking of objects in liquids. The children participated in several sessions of either (1) a constructivist learning environment with a high degree of instructional support given by the teacher, or (2) a constructivist learning environment with a low degree of instructional support, or (3) a baseline control group without instruction on the topic (see Hardy, Jonen, Möller, & Stern, 2006, for details of the interventions). Before and after the instruction as well as one year later the children indicated in a multiple choice test how strongly they agreed with a number of statements. Each statement described (a) a common misconception, (b) an everyday life explanation, which has some explanatory power in everyday life but does not hold up to systematic scientific evaluation, or (c) the relevant scientifically correct concepts. The three sum scores indicating how often each child agreed with misconceptions, everyday conceptions, or scientific concepts were used as indicators of latent profile memberships at the three measurement points in a latent profile transition model. The parameters of this model were estimated so that the similarity of the scores of persons grouped in the same latent profiles was maximized and the similarity of the scores of persons grouped in the different latent profiles was minimized (see Hickendorff et al., 2017, for methodological details). Thus, the latent profiles indicated groups of learners who had the same configuration of misconceptions, everyday conceptions, and scientific concepts. In addition to the profile characteristics and memberships at each measurement point, the model estimation also yielded the frequencies of the profile transitions over time. These were interpreted as pathways of conceptual change.

The model results in the study by M. Schneider and Hardy (2013) indicated five latent profiles. Some of these had mean score profiles that indicated integrated knowledge, that is, high profile mean scores for misconceptions only or for scientific concepts only. Other profiles indicated fragmented knowledge. In these profiles two or three of the scores for misconceptions, everyday concepts, or scientific concepts were higher than the sample mean. Most participants were on one of seven developmental pathways between these five profiles over time. These transition paths indicated a general trend from misconceptions and fragmented knowledge towards more correct and integrated knowledge. However, there were strong individual differences. Knowledge fragmentation increased on some paths and decreased on others. About 20 % of the children still had fragmented knowledge, that is, co-existing misconceptions, everyday concepts, and scientific concepts, even one year after participating in the constructivist learning environment. The instructional condition was related to the frequency of the transition paths. The constructivist learning environment with a high degree of instructional support led to the most integrated knowledge and the untreated control group to the least.

The other two studies using latent transition analyses to investigate knowledge development traced school students' knowledge of rational numbers over time (Kainulainen et al., 2017; McMullen et al., 2015). Similar to M. Schneider and Hardy's findings, students' knowledge of rational numbers was captured by a small number of knowledge profiles and systematic transition paths between these profiles over time, some of which could be interpreted in terms of conceptual change. However, the generalizability of these findings to other age groups and content domains remains unclear (cf. Edelsbrunner, Schalk, Schumacher, & Stern, 2018; McMullen, Van Hoof, Degrande, Verschaffel, & Van Dooren, 2018).

9.2.3 Is Conceptual Change Still Relevant in Higher Education?

In the current study we used latent profile transition analysis to investigate whether conceptual change and, more specifically, knowledge fragmentation and integration still can be found in higher education and whether their quality is related to the learning outcomes, that is, students' grades. There are at least three reasons to expect that this might not be the case. First, students in higher education successfully participated in school instruction, perhaps leading to a firm fundament of correct and integrated knowledge that higher education can build on. Second, students in higher education tend to have better developed meta-cognitive strategies than school children (Weil et al., 2013). Thus, they might be able to monitor, identify, and understand the confirmatory or contradictory relations between their prior knowledge and the knowledge to be learned making knowledge integration a trivial process. Finally, arguably, the content of higher education programs is more abstract than the content of school lessons, so that fewer interferences between prior knowledge from everyday life and the new knowledge may occur.

On the other side, there are also a number of reasons why conceptual change could still affect student learning in higher education. First, the structure and the functions of human memory do not change fundamentally between K-12 school and higher education. Adults' conceptual knowledge is still organized as a network, so restructuring this network might be necessary when prior knowledge and new knowledge are incompatible. Second, empirical research shows that university students still have many misconceptions, which are stable, hamper subsequent learning, and require restructuring (Merz et al., 2016; Shtulman & Valcarcel, 2012). Third, a recent meta-analysis on undergraduate science course innovations found that so-called conceptually-oriented learning tasks had a substantial effect on student achievement. Averaged over nine studies the effect size was $d = 0.47$ ($SD = 0.70$). The authors defined that tasks are conceptually oriented when they "elicit students' level of understanding of key science concepts, […] engage students in conceptual schemes within a topic rather than isolated facts, […] and engage students with real-world problems in creative ways that reflect a conceptually integrated understanding of the content" (Ruiz-Primo, Briggs, Iverson, Talbot, & Shepard, 2011, p. 1269).

Based on these findings we hypothesize that our latent profile transition analysis will find evidence for conceptual change and, in particular, for knowledge fragmentation and integration, in higher education.

9.2.4 Psychology Students' Concepts of Human Memory

We tested our hypotheses in a longitudinal study on Psychology students' understanding of human memory. Specifically, we assessed at four measurement points in the first four semesters of a Bachelor program in Psychology to what extent the students had the misconception of memory as a place for the static storage of information and to what extent they had the correct concept of memory as a dynamic system involving the construction and re-construction of knowledge (Lynn & McConkey, 1998). Examples of memory processes that modify the information to be stored are interference (R. A. Bjork, 1992), chunking (Gobet et al., 2001), and source monitoring (M. K. Johnson, Hashtroudi, & Lindsay, 1993).

Human memory is a complex causal system and one of the central constructs of Psychology (Baddeley et al., 2009). The search term *memory* has more than 190.000 hits in the literature database PsycInfo (November 2016). These articles are from fields as diverse as cognitive, educational, developmental, clinical, social, forensic, and biological psychology. An understanding of memory properties and its processes is also important in numerous psychological professions, and a flawed understanding of human memory in professionals who work in the clinical or legal context can have negative consequences (Garry, Loftus, & Brown, 1994). As far as we know, there is no standardized test of students' concepts of human memory. Therefore, we used a self-developed test in our study.

To assess the criterion validity of our new test, we additionally presented the participants with the implicit memory theory scale (IMTS; Niedźwieńska, Neckar, & Baran, 2007), which tests how skeptic individuals are with respect to the credibility of autobiographical memory. We expect that students who have a concept of human memory as a dynamic system that constructs and re-constructs information will be more skeptical with respect to the validity of autobiographic memories than other students.

9.2.5 The Current Study

In sum, conceptual understanding is a central learning goal of higher education. Yet there is almost no published research on conceptual change in higher education students. For this reason, we used a latent profile transition model with longitudinal data from university students in the current study. We constructed the measures so that the latent profiles can be interpreted in terms of knowledge fragmentation and integration and that any profile transition can be interpreted as signs of conceptual change. We tried to replicate M. Schneider and Hardy's (2013) findings on a general conceptual level, in particular, the trend from misconceptions to scientific concepts, the trend from fragmented to integrated knowledge, and the persistence of fragmented knowledge in some learners even after instruction.

Based on the previous findings of studies with school children we had the following hypotheses for our study in higher education: (1) Persons differ in their knowledge profiles and transition paths. However, the number of profiles and paths is relatively small. (2a) There are transitions between latent profiles differing in their knowledge profiles over time, giving evidence of conceptual change in higher education. (2b) The transitions reflect an overall developmental trend from profiles with higher scores of misconceptions towards profiles with higher scores of scientific concepts. (3a) At least one of the knowledge profiles indicates fragmented knowledge by indicating high agreement with mutually incompatible concepts. (3b) The profiles indicating fragmented knowledge will become less frequent but will not disappear completely over time. (4) The latent profiles differ in their mean scores on the Implicit Memory Theory Scale indicating that students' profiles of knowledge of human memory are related to how much trust they put in autobiographic memory. (5) The latent profiles differ in their grades on the human memory course showing that students' profiles of knowledge of human memory are related to grades in higher education.

9.3 Method

9.3.1 Participants

137 students enrolled in a Bachelor of Psychology program at a mid-sized university in a mid-sized German city participated at T1 in our study. Of these, N = 126 participated again at T2, N = 116 at T3 and N = 115 at T4. Almost all participants were German, and all were fluent speakers of German. At T1, the sample mean age was 20.4 (*min* = 18; *max* = 31) and all students were at the beginning of their first semester in the program. About 82 % of participants in the sample were female. This proportion is slightly higher than the proportion of all females in the program, which was 60-70 %. Participation in the study was anonymous and voluntary. To keep dropout at a minimum, participants were financially compensated with €25 per wave of the

longitudinal study and an additional completion bonus of €100 if they had participated in all waves.

The study was conducted in full accordance with the Declaration of Helsinki and the APA Ethics Code (American Psychological Association, 2002). Prior to their participation, students received an informed consent form including, among others, the following information (based on the APA Ethics code): (1) a statement on the purpose of the research as well as the expected study duration and procedures, (2) a statement that participation is voluntarily and that it may be terminated at any point; (3) a statement that there are no potential risks, discomfort or adverse effects with regard to their participation; (4) a statement that data is collected anonymously, and that even though some personal data (e.g., e-mail addresses) will be collected for organizational purposes, these data will not be used to identify individual participants and will be deleted as soon as possible.

9.3.2 Procedure

The students were tested longitudinally at four measurement points covering the first four semesters of the Bachelor in Psychology program. Data collection took place between November 2013 and May 2015. Baseline data collection (T1) took place during the first six weeks of the participants' first semester in the program, followed by three consecutive waves of measurement (T2, T3, T4) at the beginning of the second, third, and fourth semesters, respectively. Each wave consisted of two parts: A home module and a subsequent lab module. The home modules included several self-report measures and were completed online under unsupervised conditions before the respective lab sessions took place; answering the questionnaires took between 30 and 50 minutes (dependent on the wave). In the lab modules, knowledge and achievement tests were conducted in the university's computer labs. Groups of 1 to 25 participants were supervised by student experimenters and completed the tests individually and at their own pace. Each lab module took about 120 minutes.

9.3.3 Measures

Concepts of Human Memory. Conceptual understanding of human memory was assessed by a self-constructed multiple choice test at the beginning of each of the four lab modules. Previous studies (cf. M. Schneider & Hardy, 2013) have found that students' answers to multiple choice items are well in line with their answers to interview questions that aim to assess the same concepts. The tasks were presented at a computer screen and the answers were entered by mouse clicks. Each of the nine test tasks began with a written description of a situation in which a certain memory process is particularly important. These situations were derived from and similar to classical experiments in memory research (Brooks, 1967; Chase & Simon, 1973b; Ericsson et al., 1980; Goff & Roediger, 1998; Henik & Tzelgov, 1982; Hovland & Weiss, 1951; Loftus, 1975; Melcher & Schooler, 1996; W. Schneider, Körkel, & Weinert, 1989). In three situations, interference between elements in memory (R. A. Bjork, 1992) can be expected, in three tasks the chunking of elements into bundles of information (Gobet et al., 2001; M. K. Johnson et al., 1993), and in three tasks source monitoring mechanisms (M. K. Johnson et al., 1993). Each of the nine publications was cited between 181 and 3286 times, had been replicated repeatedly, and was (still) widely accepted by the scientific community.

In each of the nine tasks, the learners were presented with six statements about memory processes that might be relevant in this situation (see the example task in Figure 6 and Appendix B, SM6-SM8 for examples of all three types of tasks). There were three types of statements: (a) *Correct processing statements* described the memory process actually occurring in that situation as firmly established by empirical research, as described in the previous paragraph. These statements were compatible with the view of memory as a dynamic system involving the construction of knowledge. (b) *Incorrect processing statements* also describing dynamic knowledge construction processes in memory, however, only processes not occurring or being irrelevant in that specific situation as shown by established empirical research. Incorrect processing statements were partly correct (because they describe memory as a dynamic system) and partly incorrect (because they described processes not relevant in the respective situation). (c) *Static storage statements* described memory in that situation as a place for the static storage of information, that leaves the nature and the content of the stored information unchanged. They do not conform to the established empirical findings and the view of the scientific community and can be seen as misconception.

Wild horses

Two groups of children participate in the study.

The children in one of these groups (Group A) are twelve year-olds, who know little about horses in general.

The children in the other group (Group B) are eight year-olds, who know a lot about horses in general.

The two groups do not differ in terms of the children's intelligence and the number of boys and girls in the group.

All participants read a simple text about wild horses that teaches them what kinds of species exist, where they are found and what they need to live. Afterwards, the children are asked to answer questions assessing their knowledge and comprehension, based on the text from memory.

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect … Definitely correct
Group A will perform markedly better than Group B because they have four more years of practice in reading and remembering from texts than the younger children.	o o o o o o o
Group B will perform markedly worse than Group A, because the children are on a lower stage of cognitive development and therefore cannot process information so well.	o o o o o o o
Group B will perform markedly better than Group A because they have more prior knowledge and therefore can store information from the text in a more structured way.	o o o o o o o
Group A will perform markedly worse than Group B because the memory of twelve year-olds is partly impaired already due to hormonal changes during puberty.	o o o o o o o
Both groups will perform almost equally well because they have read the same text with the result that each person has memorized the same information.	o o o o o o o
There barely will be any difference between the two groups because their intelligence is similar and therefore they can process information almost equally well.	o o o o o o o

Figure 6. Example task in the test of concepts of human memory (in this figure translated from German to English). The correct answer is in bold print.

The participants evaluated each statement on a seven-point rating scale from definitely incorrect to definitely correct. In each task, only one memory process was relevant. Therefore, there was one correct processing statement, but three incorrect processing statements, and two static storage statements for each task. Half of the statements used a negative wording, e.g., "The two groups do not differ". The order of the nine tasks and of the six statements in each task was randomized separately for each person. We reversed the scores for the incorrect statements and then computed a separate sum score for static correct processing statements, incorrect processing statements, and static storage statements, respectively.

The test was not tailored to the content of the specific course. Human memory is referred to in many courses of the Bachelor Psychology program. So we developed the test to reflect students' accumulating knowledge of memory over several semesters and courses. Still the test content is more similar to the lecture "Human Memory" than to any other course in the students' program. Students typically attend this lecture during their first semester in the program.

Implicit Memory Theory Scale. At T3, we administered a German translation of the Implicit Memory Theory Scale (IMTS; Niedźwieńska et al., 2007) as part of the at-home module. One researcher from our lab translated it from English into German. To check for the validity of the translation, a second researcher from our lab translated the German items back into English without knowing the original English items. In case of disagreements between the two English versions, the German items were modified by both researchers together. The IMTS is a standardized instrument to test how individuals judge the credibility of autobiographical memory. Low skepticism is associated with misconceptions of human memory, such as the belief that memory is a static storage and that recollection is an accurate representation of real events, whereas high skepticism is associated with the belief that memories are reconstructed during recall and that memories are prone to qualitative changes. The global original scale has good psychometric properties (Niedźwieńska et al., 2007). In the original study, the internal consistency was good with a Cronbach's alpha of .83 and the retest reliability was also good, with a value of .81 for a two-week interval and .74 for a seven-month interval. The IMTS proved to be sensitive to knowledge differences between psychologists and non-psychologists, as well as to knowledge differences between psychology students before and after instruction on autobiographical memory. We report the mean and standard deviation of the IMTS scores in our sample in Table 4.

Grades. At T4, the university administration sent us the grades for all the exams the participants in our study had completed so far. In the German grading system, grades range from 1 to 5, with smaller numbers indicating better performance. The students are relatively free in choosing the sequence of their exams, so the types and numbers of completed exams varied between participants. For our analysis, we only used the grades of a course in general psychology, which covers the topics *human memory*, *learning*, *motivation* and *emotion*, and thus, is the best proxy for human memory competence. At T4 82 % of the sample had completed the respective exam. The mean and standard deviation of the grades is reported in Table 4

9.3.4 Statistical Analysis

In the latent profile transition model, we specified a latent profile variable for each of the four measurement points. Each latent profile variable had the sum scores for correct processing, incorrect processing, and static storage at the respective measurement point as indicators of the persons' profile memberships. The profile means of the indicator variables were constrained to be equal over time, so that fewer parameters had to be estimated and that the profiles had the same interpretation at all four measurement points. The number of participants in each profile and the variance of the indicator variables was allowed to differ between measurement points. The profile membership at each time point was used as a predictor of the profile membership at the respective next time point, so that intercepts and regression weights of three multinomial logistic regressions were estimated. MPlus uses these parameters to compute transition probabilities and transition paths between the latent profiles over time in the model. Overall, the model has 63 free parameters. These are the 12 means and 12 variances of the class indicators (i.e. for the three indicators at the four time points), 12 regression intercepts (i.e. for four minus one latent classes at four points in time), and 27 regression weights (for regressions of four minus one profiles at one time point to four minus one profiles at the respective next time point times three pairs of time points).

The data were prepared for structural equation modelling using SPSS IBM® Version 23. The latent profile transition model was analyzed in MPlus 7.31 (B. O. Muthén & Muthén, 1998-2012; L. K. Muthén & Muthén, 1998-2016). We used the maximum likelihood estimator with robust standard errors (MLR), which is the default estimator for latent transition models in MPlus. The model was estimated using the MPlus default handling of missing data according to which all available data is used to estimate the model (L. K. Muthén & Muthén, 1998-2016). When there was missing data, the covariance coverage of the variables used in our analysis was between .825 and .920 which far exceeded the minimum covariance coverage of .100. In latent profile, latent class, and latent transition analysis it is important to make sure that the best log likelihood value is replicated several times and to avoid improper solutions due to local maxima. We therefore increased the default number of random starts in the initial stage to 400 sets and the number of final stage optimizations to 100. We used unstandardized (raw) scores of the indicator variables for the model estimation to yield unbiased model results. These scores are reported in Table 4. After model estimation and to aid the interpretation of the results, we standardized the profile mean values and standard deviations to T-scores, which have a mean of 50 and a standard deviation of 10. This was done by first subtracting the sample mean from the profile mean and dividing the result by the sample standard deviation separately for each profile indicator in each profile. The sample means and SDs were $M = 0.86$, $SD = .48$, for static storage, $M = 1.66$, $SD =$

.59, for incorrect processing, and *M* = 4.22, *SD* = .65, for correct processing. Following, the resulting values were multiplied by 10 and added to 50. We report these standardized scores in the following to aid interpretation.

Table 4

Descriptive Statistics of All Unstandardized Scores

	M	*SD*	*N*
Static Storage T1	1.19	.609	137
Incorrect Processing T1	2.00	.599	137
Correct Processing T1	3.79	.715	137
Static Storage T2	.798	.591	126
Incorrect Processing T2	1.66	.654	126
Correct Processing T2	4.41	.738	126
Static Storage T3	.736	.628	116
Incorrect Processing T3	1.49	.693	116
Correct Processing T3	4.41	.823	116
Static Storage T4	.584	.579	114
Incorrect Processing T4	1.32	.714	114
Correct Processing T4	4.49	.859	114
IMTS T3	31.0	10.2	116
Grades T4	1.99	.897	111

9.4 Results

In the following three subsections, we first describe how we determined the number of latent profiles, then characterize the latent profiles and describe the transition paths between the latent profiles over time. Finally, we analyze the relationship between the latent profiles and outside criteria, that is, the Implicit Memory Theory Scale and course grades.

9.4.1 Determining the Number of Latent Profiles

We selected the number of latent profiles based on the recommendations given by Nylund et al. (2007). The best fitting model is determined by repeatedly estimating the model with increasing numbers of classes or profiles and comparing their model fits. There is no definite standard for deciding on the number of latent profiles (Nylund et al., 2007). Better-fitting mixture models are characterized by lower comparative fit indices, i.e. lower values in the Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted BIC. Another way to examine model fit is to compare the classification quality of individuals into latent profiles between models that differ in the number of profiles assumed. Classification quality or *entropy* is measured on a scale that ranges between zero and one, with a value of one indicating perfect classification of individuals into latent profiles (Clark & Muthén, 2009).

Table 5

Fit Indices for the Latent Transitions Models with One to Five Latent Profiles

Index	One profile	Two profiles	Three profiles	Four profiles	Five profiles
Log Likelihood	-1513	-1350	-1278	-1245	-1212
Free parameters	24	25	41	63	91
AIC	3074	2749	2638	2617	2607
BIC	3145	2822	2758	2801	2873
Sample-size adjusted BIC	3069	2743	2628	2601	2585
Entropy	-	0.84	0.83	0.86	0.84

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion.

We repeatedly estimated the model, each time with a different number of specified latent profiles. For each model, the best log likelihood value was replicated several times with the initial specification of 400 sets of random starts and 100 iterations, that is, 99 times in the one-class model, 100 times in the two-class and three-class models, 25 times in the four-class model, and 8 times in the five-class model. In the second step, in which we ran the models again with twice the number of random starts and final stage iterations, each of these values was replicated again. Table 5 presents the model fit indices for the models with one to five latent profiles. The models with two to five latent profiles had entropies above .80, which is high according to (Clark & Muthén, 2009). The BIC was lowest for the three-profile model, whereas the AIC and sample size-adjusted BIC were lowest for the five-profile model, and entropy was highest in the four-profile model. Therefore all three models fit the data approximately equally well. We chose the four-profile model because the four-profile model but not the three-profile model included a profile indicating strong misconceptions (high values for static processing only). Even though only small sample proportions showed this profile in our study, it is still interesting, because it corresponds to the lowest level of knowledge of human memory, thus is the starting point of competence development in a domain, and had also been found in previous research (M. Schneider & Hardy, 2013). Furthermore, we chose the four-profile model over the five-profile model because it was more parsimonious and had a lower BIC. All subsequent analyses were conducted with the four-profile model.

Interpretation of the Profile Means. Table 6 lists the T-standardized indicator means for the four latent profiles. In Figure 7 the standardized scores are represented graphically. To aid interpretation, we tested for static storage statements, incorrect processing statements, and correct processing statements in each latent profile at each measurement point whether the latent profile mean significantly deviated from the overall sample mean. For example, we fixed the mean of static storage in profile 1 at the overall sample mean of static storage while fixing the other means at the values of the initial 4 x 4 latent transition model. In total, we computed 3 x 4 = 12 models for the three mean scores across the four latent profiles. Then, we computed log likelihood ratio chi-square difference tests to test for the difference between the fixed mean models and the initial model. We computed the robust test statistic for MLR estimation (Satorra & Bentler, 2010) as reported by Asparouhov and Muthén (2010; formula 2) and adjusted the significance level for the number of comparisons using the Bonferroni method, yielding a significance level of $\alpha < \frac{.05}{12} <$ .004. Table 6 shows the results of these comparisons.

We interpreted the latent profiles based on the standardized mean scores, the results of the computed log likelihood ratio chi-square difference tests, and assigned labels to them (see column 2 in Table 6). We labeled the profile C1 *misconceptions profile* because members of this profile show the highest scores for the static storage misconceptions of all classes, average scores for incorrect processing statements and below-average scores for correct processing statements, even though the first and last deviation were not significant in the chi square difference test. However, it should be noted that only a small number of individuals were assigned to the misconceptions profile, yielding large standard errors and the Bonferroni correction is a rather conservative correction of the significance level. We called the profile C2, which shows significant above-average means on two of the three scales, *fragmented profile* because these learners express inconsistent knowledge: On the one hand, they believed that human memory was a static storage and on the other hand they believed that memory processes and re-constructs information.

Table 6

Assigned Labels for the Latent Knowledge Profiles, Sample Proportions, Standardized Scores of the Scales' Means, and Significance of the Deviation of Each Mean From 50

	Label	Sample Proportion in %				Static storage		Incorrect Processing		Correct Processing	
		T 1	T 2	T 3	T 4	*M*	*p*	*M*	*p*	*M*	*p*
C1	Misconceptions profile	7	5	6	4	92	.053	49	1.00	30	.544
C2	Fragmented profile	56	22	17	17	71	<.001*	71	<.001*	39	.028
C3	Indecisive profile	37	46	43	38	40	.281	48	.927	52	.998
C4	Scientific profile	0	27	33	42	26	<.001*	25	<.001*	70	<.001*
	Total	100	100	100	100						

Note. * significant on the level of $\alpha < 0.004$, corrected for multiple testing.

This profile supports our hypothesis H3a, that some learners have fragmented knowledge rather than integrated knowledge. The third profile, C3, indicated that the participants mostly chose the answer category in the middle of the rating scale with average scores on all three scales. We labeled it *indecisive profile*. Finally, we labeled the profile of C4 *scientific profile* because it displayed significant above-average means of scientific concepts and significant below-average scores of static storage statements and incorrect processing statements, representing the ideal learning outcome.

To further aid interpretation, we tested whether the latent profiles differed significantly in their means on the three indicator variables. We computed a separate model for each pair of the four latent profiles. In each model, we constrained two class profiles to be equal whereas all other model parameters were fixed to the values found with the original (unconstrained) four-profile model. We corrected for multiple testing using the Bonferroni method and set the significance level at $\alpha < \frac{.05}{6} < .008$. Five of the six constrained models had a significantly worse fit than the original four-profile model, as found by likelihood ratio chi-square difference tests, all $ps \leq .001$. Only the comparison between the unconstrained model and the model in which the first and the third profile were constrained equal was not statistically significant, $p = .012$. The reasons for this comparison missing the critical significance level was the low frequency of the misconceptions profile and the very conservative Bonferroni-corrected critical significance level. Descriptively, as shown in Figure 7, there is a large difference between these two profiles. Thus, the results indicate that all four class profiles are mutually different and support our first hypothesis, that there are latent profiles of learners differing systematically in their conceptual knowledge (H1).

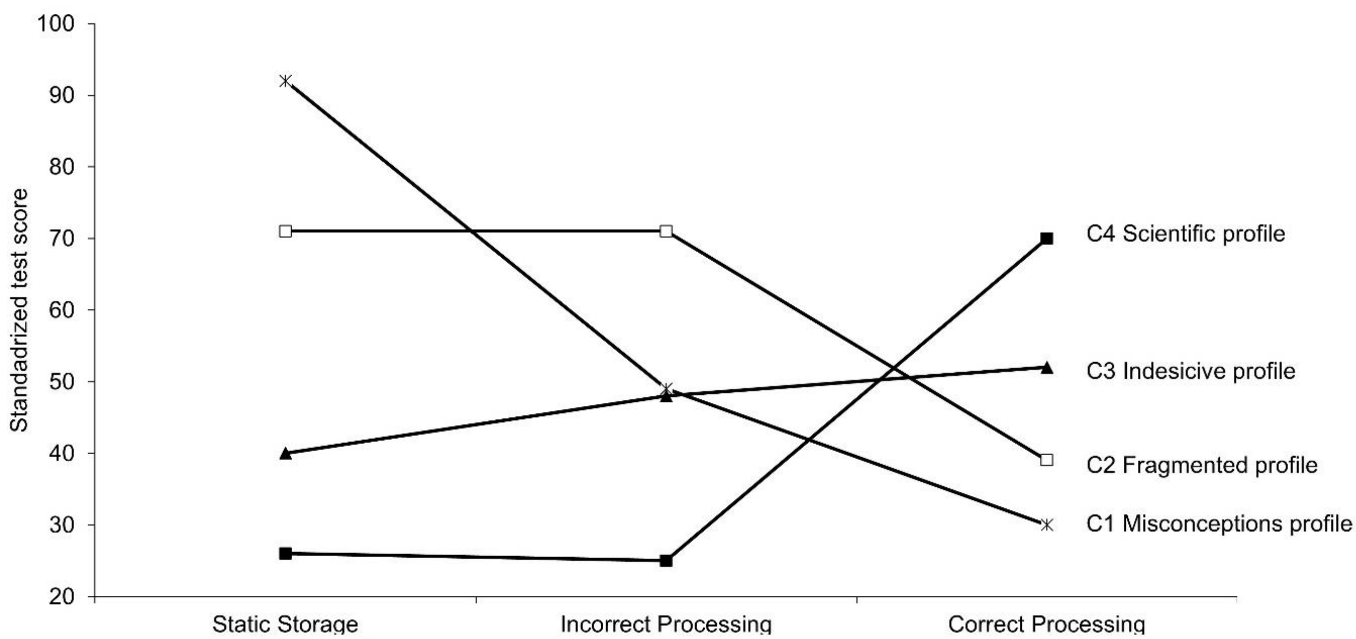

Figure 7. Diagram of the four knowledge profiles

9.4.2 Changing Profile Frequencies on the Sample Level over Time

The proportions of the latent profiles changed substantially from T1 to T4 (see in Table 6), except for the misconceptions profile, which was shown only by a small proportion of the sample (4-7 %) at all four measurement points. In line with hypothesis H3b, the frequency of the fragmented profile decreased from 56 to 17 %, but did not approach 0, thus indicating that some Psychology students still had fragmented knowledge about the memory concept at the end of their fourth semester. The proportion of the indecisive profile increased from T1 to T2 and stayed stable to T3, before finally dropping back to the initial level. The frequency of the scientific profile increased sharply from 0 % at T1 to 37 % at T2 and remained high afterwards. These findings are in line with our second hypothesis (H2b), that there is an overall trend towards a scientific understanding of human memory.

Strengths of the Predictive Relations between Knowledge Profiles. In order to test for possible predictive relationships between the knowledge profiles of T1 to T4, we generated three frequency tables (T1-T2, T2-T3, and T3-T4). We used the profile proportions and the transition probabilities based on the estimated model to compute profile frequencies. Then we build cross tabulations, including the profile frequencies at one point in time (rows) to predict the profile frequencies at the next point in time (columns). For the predictive value of T1 for T2, the chi-square test indicated a highly significant and strong positive relation, $\chi^2 = 174.37$, $df = 6$, $p < .001$, Cramer's $V = .798$. The relations were about equally strong for the other points of time, $\chi^2 = 313.32$, $df = 9$, $p < .001$, Cramer's $V = .876$ (T2xT3), and $\chi^2 = 266.08$, $df = 9$, $p < .001$, Cramer's $V = .808$ (T3xT4). Thus, the knowledge profiles allow predictions about students' future knowledge profiles at a later point in time. Only 23 % of the participants did not change their knowledge profile over the course of the study. For the remaining 77 % of the sample, information about the profile at one point in time helped to predict changes to another profile. In the following section we take a closer look at the participants' individual transition paths.

Latent Transition Paths. MPlus reports overall sample latent transition probabilities based on the estimated model across time (for the latent transition probabilities in the current study, see Appendix B, SM5). These probabilities are used to calculate latent transition paths. In a model with four profiles at four measurement points, there are $4^4 = 256$ possible latent transition paths between the profiles across time. In line with our Hypothesis 1, the participants were on few of these possible paths. There were only six paths taken by at least 5 % of the sample (see Figure 8), and 81 % of the sample was on these six paths. Approximately 17 % of the sample was on another eight paths which were taken by at least one person of the sample. The remaining 3 % were on paths that were estimated to be taken by less than one person of the sample by MPlus. The fact that a high proportion of the sample followed a small number of developmental paths shows that knowledge development was highly systematic in our sample.

Table 7

Pathways of Conceptual Change (i.e. Model-Estimated Latent Transition Paths) in the Sample

Path	Label	Knowledge Profile				Proportion of the sample (%)	
		T1	T2	T3	T4	Alone	Accu-mulated
P1	Increasing correct knowledge	C3	C4	C4	C4	25	25
P2	Decreasing Fragmentation	C2	C3	C3	C3	22	47
P3	Enduring Fragmentation	C2	C2	C2	C2	15	62
P4	Enduring Indecisiveness	C3	C3	C3	C3	8	70
P5	Slowly Evolving Scientific Concepts	C2	C3	C3	C4	6	76
P6	Quickly Evolving Scientific Concepts	C2	C3	C4	C4	5	81
P7-P14	Various Paths found for at least one person	Various profiles	Various profiles	Various profiles	Various profiles	17	97
P15-P256	Various Paths not found for at least one person	Various profiles	Various profiles	Various profiles	Various profiles	3	100

Note. C1 = misconceptions profile; C2 = fragmented profile; C3 = indecisive profile; C4 = scientific profile

We interpreted the transition paths based on between which profiles the transitions occurred and assigned labels to them (see Table 7). The first path P1 was labelled *increasing correct knowledge* because participants on this path displayed an indecisive profile at T1 and transitioned to the scientific profile at T2, where they stayed throughout the course of the study. The second path P2 was called *decreasing fragmentation* because individuals on this path transitioned from the fragmented profile to the indecisive profile which reflects a development towards more integrated pieces of knowledge. Path P3 were called *enduring fragmentation,* as participants on this path stayed with the fragmented profile throughout the whole study. Similarly, participants on P4 stayed with the indecisive profile, thus we labelled the path *enduring indecisiveness*. The paths P5 and P6 both included two transitions, and students on these paths started with the fragmented profile, transitioned to the indecisive profile, and finally moved to the scientific profile. Students on P5, transitioned to the scientific profile between T2 and T3, whereas students on P6 transitioned to the scientific profile between T3 and T4. We labelled the paths *slowly evolving scientific concepts* and *quickly evolving scientific concept* because these transition represent a gradual decrease in the static storage misconception while processing statements are more and more endorsed.

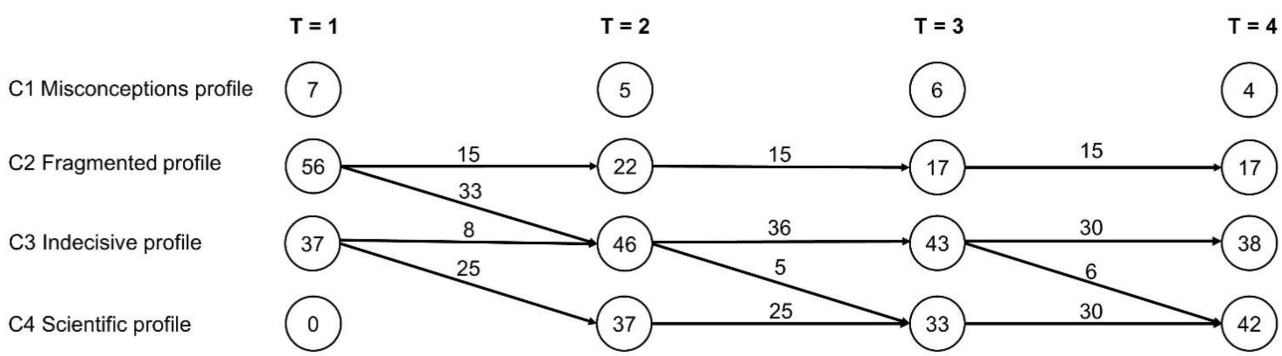

Figure 8. Diagram of the six transition paths taken by at least 5 % of the sample. All numbers are percentages of the sample. The numbers in circles refer to the whole sample (100 % in total at each wave). The numbers next to the arrows refer to the proportion of the sample on the most common six transitions paths (81 % of the sample in total).

As can be seen in Figure 8, on the six most frequent transition paths, the participants either stayed in their respective latent profile over time or transitioned from lower-ranking profiles (e.g., the indecisive profile) to higher-ranking profiles (e.g., the scientific profile). There were no transitions in the opposite direction. Between T1 and T2, the majority (58 %) of the sample transitioned from a lower-ranking to a higher-ranking profile. Conversely, between T2, T3, and

T4 the majority (72 %) of the sample stayed in their respective latent profile. This mirrors the fact that most participants attended a lecture on human memory during the first semester of their program.

9.4.3 Associations without Outside Criteria

To test for relations between the knowledge profiles and outside criteria, we exported the list of participants' most likely profile memberships to SPSS. To test statistically for associations between profile memberships at T3 and T4, IMTS scores, and grades, IMTS and grades were recoded, so that higher scores indicated better achievement or better grades. As the number of students differed strongly between the four knowledge profiles, we used non-parametric tests to conduct the analyses. Parametric tests, such as the analysis of variance (ANOVA) have been shown to be robust against some violations of assumptions (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010). However, multiple problems and unequal group sizes lead to serious constraints to the robustness of parametric tests (Lix & Keselman, 1998). As a non-parametric alternative to one-way ANOVA we used the Kruskal-Wallis H-test. To follow up on the findings, we used the Jonckheere-Terpstra test (Jonckheere, 1954; Terpstra, 1952), which allows to test whether the medians of the groups are ordered in a meaningful way. In SPSS, the Jonckheere-Terpstra test investigates whether the medians ascend or descend in a pre-defined order. We specified the independent variable in the hypothesized developmental rank order of knowledge profiles, that is, from a misconceptions profile, over fragmented and indecisive profiles to a scientific profile. Table 8 shows the recoded IMTS scores and grades by most likely profile memberships.

As hypothesized (H4), the latent profiles at T3 differed significantly in their IMTS scores as indicated by the Kruskal-Wallis H-test, $H(3) = 25.845$, $p < .001$. Jonckheere's test was significant, $J = 3112.500$, $z = 4.546$, $r = .422$, indicating that participants' knowledge profiles predicted their achievement on the IMTS in the assumed developmental rank order from the misconceptions profile to the scientific profile. Thus, persons with a better understanding of memory in general are also more skeptical about the credibility of autobiographical memories. This is in line with a view of human memory as involving the active processing and reconstruction of information.

Persons differing in their knowledge profiles at T4 also differed systematically in their grades on the human memory course, $H(3) = 16.122$, $p = .001$. Jonckheere's test indicated that there was a significant trend in the data, indicating that grades improved from the misconceptions profile, over the fragmented profile and the indecisive profile, to the scientific profile , *J* =2714.500, *z* = 3.547, *r* =.335, which supports our hypothesis H5.

Table 8

Descriptive Statistics for IMTS Scores and Grades by Most Likely Profile Memberships

		IMTS T3			Grade T4		
	Label	*M*	*SD*	*n*	*M*	*SD*	*n*
C1	Misconceptions profile	25.5	10.4	7	2.00	.628	5
C2	Fragmented profile	18.4	10.0	20	1.22	1.06	19
C3	Indecisive profile	30.3	8.43	53	1.97	.860	43
C4	Scientific profile	33.7	8.27	36	2.28	.686	46

Note. IMTS scores and grades are coded so that higher scores reflect higher competence.

9.5 Discussion

9.5.1 Knowledge Profiles

Our aim was to investigate whether conceptual change is still a relevant learning process in higher education students, therefore we conducted the first latent profile transition analysis in this population. In line with our Hypothesis 1, we found that there were individual differences in Psychology students' knowledge about human memory. Our analyses show that these individual differences of the 137 learners can be described in terms of only four knowledge profiles.

The misconceptions profile, in which students agreed mostly with static storage statements, was found in less than 10 % of the sample at all four measurement points, respectively. This does not mean that the students in the sample started the study with perfect prior knowledge, though. The most frequent profile at the beginning of the first semester was the fragmented profile characterized by strong agreement with both static storage statements and incorrect processing statements. These two kinds of statements are mutually incompatible, because either knowledge is static or is processed in memory. We therefore interpreted students' agreement with both kinds of statements as indicating fragmented knowledge, in which the students hold both kinds of beliefs but do not understand their inter-relation. In line with Hypotheses 3a and 3b, the fragmented profile occurred frequently at the beginning of the study, decreased in its frequency over time, but was still found for 17 % of the sample at the end of the study.

Many other students seemed to be aware of their lack of a complete understanding of memory, as the indecisive profile was the second most frequent at the first and the last measurement point and the most frequent on the second and third measurement point. Students with this profile tended to choose the middle category when evaluating statements, thus, neither agreeing nor disagreeing with them. Thus, not all students in our sample held deeply entrenched misconceptions. Instead, some students were aware of their lack of understanding. Students with the scientific profile agreed mostly with correct processing statements. They did not only understand that memory actively constructs and re-constructs information, but also in which situations chunking, interference, or source monitoring are the most relevant memory processes.

The strongly decreasing sample proportions (from 56 % at T1 to 17 % at T4) for the fragmented profile and the strongly increasing sample proportions for the scientific profile (from 0 to 42 %) demonstrated a trend towards more scientifically correct and also more integrated knowledge over the course of the four semesters (Hypotheses 2b and 3b). These findings with respect to the knowledge profiles and their changing frequencies over time are generally well in line with the findings obtained by M. Schneider and Hardy (2013) as well as Edelsbrunner et al. (2015) with elementary school children. These two studies likewise found learners with fragmented knowledge and learners with integrated knowledge in their sample. Also, they found that fragmented knowledge decreased in its frequency in the sample but remained in some learners, and found an overall trend from misconceptions and fragmented knowledge to more integrated and correct knowledge. The main difference between the previous and the present findings is that the most frequent profile before learning was the misconceptions profile in the elementary school children and the fragmented profile in the university students. Future studies will have to investigate whether this is a general difference between the two age groups. Possibly, the longer learning history of the university students led to a great base of badly integrated knowledge in comparison to the school children.

9.5.2 Transition Paths

Students' knowledge changes over the four measurement points could be described in terms of few transition paths between the four knowledge profiles. There were only six paths taken by at least 5 % of the sample, respectively, and a total of 81 % of the sample was on these paths. The remaining part of the sample was on one of eight additional paths. The participants on the six most common paths either stayed in their respective latent profile or moved to a higher ranking profile, as shown in Figure 8. Thus, there was a clear developmental ordering of the profiles from the misconceptions profile, over the fragmented profile and the indecisive profile to the scientific profile. The knowledge profiles differed in how strongly participants agreed with statements about static storage concepts, incorrect processing concepts, or correct processing concepts of human memory. We thus interpret transitions between these knowledge profiles as evidence of conceptual change. The involved profiles demonstrate the strength of conceptual change underlying the transition: Transitions between more similar profiles (e.g., the misconceptions profile and the indecisive profile) arguably require less knowledge restructuring than transitions between dissimilar profiles (e.g., the misconceptions profile and the scientific profile).

The highly systematic pathways imply that the learners' knowledge at each point in time is a good predictor of the learners' knowledge at later points in time. Indeed, we found significant and very strong relations between the knowledge profiles at different points in time. The fact that on the six most common paths more than half of the participants transitioned from one profile to another, that is, restructured their knowledge, shows that conceptual change is a relevant learning mechanism in higher education and in the domain of Psychology. Zero percent of the participants had the scientific profile at the start of the study. The 42 % of the participants having this profile at the end of the study all transitioned there from lower-ranking knowledge profiles over time. This demonstrates that conceptual change does not only happen but is a central learning mechanism in acquiring academic concepts in higher education.

The findings also demonstrate the importance of knowledge fragmentation and integration for understanding and predicting conceptual change in higher education. As predicted, knowledge fragmentation was frequent at the start of the study, stayed constant on one path (the enduring fragmentation path), and decreased in its frequency on other paths (e.g., the decreasing fragmentation path and the slowly evolving scientific concepts path).

The findings demonstrate systematic associations of the profiles and pathways with domain-specific instruction, a finding also observed in the study by M. Schneider and Hardy (2013).

Overall, 58 % of the persons on the most frequent six paths transitioned to a different profile and 33 % integrated previously fragmented knowledge during the first semester, during which most participants also attended a lecture about human memory. In contrast, only 11 % of the participants transitioned to a different profile during the second, third, and fourth semester together, and no participants integrated their knowledge during that time.

9.5.3 Validity of the Model Results and Limitations

At least five findings indicated that the model results reflect systematic and meaningful individual differences rather than random measurement error. First, the transition paths and the predictive relations between the knowledge profiles show a high degree of systematic organization in the results both for individual differences at each point in time and intra-individual differences over time. Second, the basic pattern of results in the current study conceptually replicates the key findings of M. Schneider and Hardy (2013), such as the small numbers of profiles and pathways, the co-existence of fragmented and integrated knowledge in the sample, and general trends towards more integrated and correct knowledge. Third, transitions to higher-ranking profiles were more frequent during domain-specific instruction (i.e. the lecture on human memory during the first semester) than at other times. Fourth, the latent profiles differed in their mean rank-ordered grades, with students in higher-ranking profiles also showing better grades. Though evidence from analyses with most likely profile variables can be somewhat limited when entropy is not perfect, we believe that in our study, in which entropy was only slightly smaller than .90, such bias was only a marginal problem. Therefore, the findings support the criterion validity of the latent profile variables and demonstrate the grade-relevance of academic concepts and conceptual change in higher education. Finally, students in higher-ranking profiles also showed better performance on the implicit memory theory scale, indicating that a better understanding of memory as involving the construction and re-construction of information went along with a greater skepticism with respect to the contents of autobiographical memories.

We expect that the findings of our study are highly generalizable, as the interpretation our knowledge profiles is in line with previous studies conducted outside higher education (Edelsbrunner et al., 2015; McMullen et al., 2015; M. Schneider & Hardy, 2013). Our findings are somewhat limited as the identification of knowledge profiles in latent transition analysis is exploratory. There are no definite standards in choosing the number of profiles. In latent transition analysis, the number of estimated parameter rises exponentially with the number of profiles assumed. With 63 free parameters in our model with four profiles at four measurement points and 137 participants, our design is at the margin of being underpowered. Our sample is restricted to psychology students that represent a highly selected, presumably high-achieving subpopulation. Future studies should investigate whether students from other programs than psychology show similar knowledge profiles and developmental patterns. Another limitation is the use of a newly constructed test in our study. Future studies will have to further examine the validity and reliability of the test and its sensitivity to knowledge changes over time.

9.5.4 Theoretical, Methodological, and Practical Implications

Our findings have theoretical, methodological, and practical implications. On the theoretical level, the results demonstrate that conceptual change still takes place in higher education students and is vital for acquiring an understanding of academic concepts. Conceptual change, its measurement, its prevalence, its dynamics over time, and its stimulation by learning environments are well investigated for school students but almost not examined at all for students in higher education. Our present findings suggest that future studies should systematically investigate parallels and possible differences between conceptual change in K-12 school and in higher education. This will lead to a better understanding of which processes are specific to conceptual change in general and which processes only occur in specific age groups. Further, future studies should include motivational and affective covariates into latent transition models in order to investigate non-cognitive determinants of conceptual change. The small number of profiles and pathways indicated that individual differences matter in conceptual change, but that the number of possible individual differences is quite limited. This suggests that the learners' idiosyncratic knowledge construction processes are constrained by cognitive and environmental variables which guide these processes along a few developmental trajectories.

On the methodological level, our findings indicate that latent profile transition analyses are useful tools for modeling longitudinal changes in multidimensional knowledge structures in general and for modeling conceptual change in specific. Only very few latent profile transition analyses have been published so far (Hickendorff et al., 2017). The convergence of findings from earlier studies with school children and the current study with Psychology students indicates a high degree of replicability, stability, and generalizability of the findings across studies, age groups and content domains. Latent profile transition analyses have also proven to be effective data reduction techniques, because they help to describe individual differences in large samples in terms of just a few profiles and transition paths.

On the practical level, findings from studies like ours can inform higher education teaching. Teachers in higher education might often lack awareness that students enter their programs with prior beliefs and misconceptions, because higher education learning has long been neglected in research on conceptual change. The profiles and pathways identified in the current study can help teachers to better identify students' current knowledge and to predict students' future pathways of learning. Instruction can thus be tailored to fit subpopulations of students differing in their prior knowledge or development. The current study as well as the study by M. Schneider and Hardy (2013) also found evidence of systematic relations between instruction and the knowledge profiles and pathways. Future studies should investigate these interactions in greater detail.

10. Study 3: State and Trait Stress Differentially Predict Health and Achievement in Higher Education Students

10.1 Abstract

Prior research has emphasized that feeling overwhelmed by the challenges encountered in higher education results in high levels of stress among students and consequently, low academic achievement and poor health. However, none of the thousands of studies of stress inside and outside higher education has ever investigated whether impairment in health and achievement differs for temporary elevated stress levels, "state stress" and general elevated stress levels, i.e., "trait stress". In order to close this gap, we conducted two studies using latent state-trait (LST) analyses, which is a powerful research tool to study state and trait effects. We found an excellent fit of the LST model in both studies. In Study 3a (N = 225), there were two stress states, which were predicted by situational factors, while a second order stress trait factor was predicted by stable personality dispositions. Study 3b (N = 77), which was conducted longitudinally over the course of one semester, replicated the findings from Study 3a in a model with five stress states, which were found to vary around a stable stress trait. Students who showed higher levels of state stress at the end of the semester (T5) and higher levels of trait stress reported worse mental and somatic health at T5, whereas academic achievement at T5 was only related to state stress at the same time but not to trait stress. Our results show that the distinction between state and trait stress is valid. Implications for research and practice are discussed.

10.2 Introduction

Being a student in higher education can be stressful due to numerous reasons, including learning demands, examinations, the transition to university, being in a different country, and financial issues (see Robotham, 2008, for a review of stressors). Consequently, high levels of stress have been reported in Europe (Abouserie, 1994; Deasy et al., 2014) and the US (Lust et al., 2015; Reisberg, 2000) for a wide range of higher education programs. For example, in 2015 in Germany, the country of our data collection, interviews with a representative sample of 1,000 higher-education students revealed that 45% of the students reported stress-related exhaustion, 27% felt overwhelmed and unable to cope with the stress at least once during their time in higher education, and 12% reported seeking professional help in that situation (Techniker Krankenkasse, 2015b). As the number of students in higher education programs has dramatically increased to about 40 % of the population of young adults today (OECD, 2014), it is no longer a minority who is at risk of experiencing high levels of stress.

Stress in higher education is a reason for concern, because high amounts of chronic stress have negative effects on mental and physical health. Chronic stress negatively affects the immune system (Segerstrom & Miller, 2004) and is a risk factor for depression, anxiety, cardiovascular problems, obesity, and substance abuse (De Kloet et al., 2005; Deasy et al., 2014; Wang et al., 2008). The overall cost of work-related stress in the European Union has been estimated to be 20 billion Euro per year (EuropeanCommission, 2002, p. 13), and there is no reason to believe that stress would be less harmful in higher education than it is in the workforce. In 2015 21% of a nationally representative sample of 188,394 higher-education students in Germany had been diagnosed with a psychological or behavioral disorder (Techniker Krankenkasse, 2015a, p. 77). Among the most frequently found disorders were stress-related problems, such as depression, anxiety, and somatoform disorders.

Excessive stress can also hamper academic performance, as stress is known to impair cognitive processes, such as learning, memory encoding and recall (Lupien et al., 2007; Vogel & Schwabe, 2016). Consequently, lower academic achievement can be expected for students who experience high levels of stress. The correlation between study-related stress and academic achievement in a sample of 225 Dutch university students was $r = -.27$ (Pluut et al., 2015). A study with a random sample of 176 Australian university students found that higher values of stress-related variables predicted higher psychological distress, lower study satisfaction, and lower academic performance (Cotton et al., 2002). A meta-analysis of eight studies including 1736 students found that the average correlation between students' stress and grades in higher education was $r = -.14$. (Richardson et al., 2012).

10.2.1 Latent State-Trait Analysis as a Method to Study Stress

Latent state trait (LST) analyses (Schmitt & Steyer, 1993; Steyer, 1999; Steyer et al., 2015) have almost universal applications in psychological research, as psychological measurements do never take place in a "situational vacuum and we always measure persons in situations" (Steyer, 1999, p. 392). That means, that interindividual differences in scores on a psychological test, can be explained by interindividual differences in the situation of test taking (state effects), stable interindividual differences (trait effects), and the interaction of these two factors (state × trait interaction effects). For example, two persons could differ in their levels of self-reported stress because they differ in their levels of currently experienced stress, because they differ in their general tendency to experience stress, and the interaction of both. LST analysis is a powerful tool that allows researchers to explicitly model these sources of variance in true scores (Steyer, 1999). This is possible because in LST analysis, the true score at each measurement occasion is decomposed into a latent state variable that represents a psychological state in that specific situation and a latent trait variable that represents the common variation of the states across all measurement occasions (Schmitt & Steyer, 1993).

However, to date there is not a single published study that has used LST models to analyze self-reported stress levels. LST model of stress have been confined to physiological indicators of stress, such as cortisol levels. Previous work in the field of perceived stress has focused on predictors of stress and stress-related outcomes without distinguishing between trait and state sources of variance. Trait stress has been almost completely neglected with studies on driver stress by Hennessy and colleagues (e.g., Hennessy & Wiesenthal, 1997, 1999; Hennessy, Wiesenthal, & Kohn, 2000) as the only exceptions. However, for self-report data of several stress-related moods and emotions, LST models have been used, for example, anxiety (Spielberger & Sydeman, 1994), anger (Forgays, Forgays, & Spielberger, 1997), cheerfulness (Ruch, Köhler, & Van Thriel, 1997), and job satisfaction (Dormann, Fay, Zapf, & Frese, 2006). Over several decades, studies consistently found that state and trait emotions partly differ in their physiological correlates (Satpute, Mumford, Naliboff, & Poldrack, 2012; Tian et al., 2016), and that they are differently related to a wide range of psychological processes, including the startle reflex (Grillon, Ameli, Foot, & Davis, 1993), attention (Pacheco-Unguetti, Acosta, Callejas, & Lupianez, 2010), working memory (Ng & Lee, 2015), decision making (Stöber, 1997), and mental disorders (Kennedy, Schwab, Morris, & Beldia, 2001).

LST analysis with longitudinal cortisol data. In studies of physiological indicators of stress, LST analyses have been used to separate stable basal differences in cortisol activity from individual differences in momentary cortisol activity due to internal and external stressors (e.g., Kertes & van Dulmen, 2012). For this purpose, several measures of cortisol serve as indicators of a latent state factor, which estimates a person’s cortisol level in a specific situation free of measurement error. Baseline differences in cortisol levels on the other hand, are reflected by a second order latent trait factor, which uses the latent state factors as indicators. This logic is in line with cumulative evidence from studies using various analytical strategies. Evidence from these studies suggests that the presence or anticipation of acute stressors is associated with elevated situational cortisol levels, whereas chronic stress leads to dysregulations in basal, diurnal cortisol activity, such as increased or decreased cortisol peaks after waking and a flattened diurnal slope of cortisol (Saxbe, 2008).

Studies using LST models with cortisol data report good to excellent model fits, showing a good differentiation between latent state and trait cortisol (Doane et al., 2015; J. Hellhammer et al., 2007; Kertes & van Dulmen, 2012; Kirschbaum et al., 1990; Shirtcliff, Granger, Booth, & Johnson, 2005; Stalder et al., 2012). In two studies this finding was underpinned by a significantly better fit of the LST model compared to a single trait model (J. Hellhammer et al., 2007; Stalder et al., 2012). Furthermore, the distinction of state and trait cortisol is supported by differential associations between trait and state cortisol levels and other variables (Shirtcliff et al., 2005). Variance proportions in the observed cortisol levels that derive from trait differences (common consistency), state differences (occasion specificity), and unreliability of the measurement have been quantified in a number of studies using standard formulae (Schmitt & Steyer, 1993). Whereas reliability was generally high in studies reporting LST models of cortisol levels, common consistency and occasion specificity varied markedly between studies. One study reported high proportions of common consistency and low proportions of occasion specificity (Stalder et al., 2012); others the opposite (J. Hellhammer et al., 2007; Shirtcliff et al., 2005). And some studies reported almost equal proportions of variance or mixed evidence for separate samples or sampling methods (Kertes & van Dulmen, 2012; Kirschbaum et al., 1990). These inconsistencies suggest that sampling time and methods may be a crucial factor for common trait and occasion specific effects.

Relationship between cortisol levels and perceived stress. An acute increase in cortisol is the consequence of the activation of the hypothalamic–pituitary–adrenal (HPA) axis in stressful situations (Kudielka, Buske-Kirschbaum, Hellhammer, & Kirschbaum, 2004; Kudielka, Schommer, Hellhammer, & Kirschbaum, 2004). In line with this, studies showed that there is a positive relation between elevated basal cortisol levels and self-reported stress (J. C. Pruessner, Hellhammer, & Kirschbaum, 1999; M. Pruessner, Hellhammer, Pruessner, & Lupien, 2003; Schulz, Kirschbaum, Prüßner, & Hellhammer, 1998; Wüst, Federenko, Hellhammer, & Kirschbaum, 2000). However, the interplay between psychological stressors, HPA axis activation and cortisol release is very complex and there are many methodological, psychological and biological reasons to expect an imperfect association between salivary cortisol and perceived stress (D. H. Hellhammer, Wüst, & Kudielka, 2009). Meta-analytical evidence from 208 studies shows that cortisol levels are not affected by all kinds of psychological stressors, such as noise exposure and emotion induction. By contrast, uncontrollable stressors and stressors that include social-evaluative threat elicit the most pronounced changes in cortisol levels (Dickerson & Kemeny, 2004).

The positive, but not perfect association between perceived stress and physiological indicators of stress demonstrates that studying perceived stress is important in its own right. We hypothesized to find a state-trait structure for perceived stress in our study. To our knowledge, none of the several thousand published studies investigating perceived stress in higher education presented factor-analytic evidence on the dissociation of state and trait components of stress.

10.2.2 Potential Predictors of Perceived State and Trait Stress

Potential predictors of state and trait stress come from numerous studies inside and outside higher education and include stable personality dispositions, as well as situation-specific coping strategies, demands and resources. We hypothesize that predictors that show high temporal stability, such as personality dispositions, will be strongly related to trait stress, while predictors that show greater situational variability will be strongly related to state stress.

Two stable personality dispositions that have consistently been associated with stress are neuroticism and extraversion. Evidence from studies inside higher education indicates that lower levels of neuroticism and/ or higher levels of extraversion go along with lower levels of stress and higher levels of favorable outcomes (Bolger & Zuckerman, 1995; Halamandaris & Power, 1999; Saklofske, Austin, Mastoras, Beaton, & Osborne, 2012). The use of coping strategies was associated with lower levels of stress (Saklofske et al., 2012), higher levels of physical health, mental health, and well-being (Darling, McWey, Howard, & Olmstead, 2007; Dyson & Renk, 2006; Eisenbarth, Champeau, & Donatelle, 2013; Pritchard et al., 2007; Sideridis, 2006), and better adjustment to college life (Aspinwall & Taylor, 1992). Despite these positive overall effects of coping, a purely emotion-focused or avoidant coping strategy, wherein students only think about their stress or try to distract themselves, went along with higher stress levels (Aspinwall & Taylor, 1992; Dyson & Renk, 2006; Eisenbarth et al., 2013; Liverant, Hofmann, & Litz, 2004; Pritchard et al., 2007). In contrast, problem-focused coping or a combination of problem-focused and emotion-focused coping went along with lower stress levels (Sideridis, 2006). Demands and resources have been mainly studied in the workforce. Generalizing the definitions given by Bakker and Demerouti (2007, p. 312) for work-related stress, we see demands as “physical, psychological, social, or organizational aspects […] that require sustained physical and/or psychological (cognitive and emotional) effort or skills and are therefore associated with certain physiological and/or psychological costs”. And we see resources as physical, psychological, or social means of an individual that are functional in achieving the individual’s goals, reducing situational demands and the associated physiological and psychological costs, or stimulating personal growth and learning. Demands placed upon higher education students include examinations, financial pressures, the transition to university, and other stressors related to studying (Robotham & Julian, 2006). One important resource against stress is social support, which has consistently been reported to reduce the negative effects of stressors (Bakker, Demerouti, & Schaufeli, 2003; S. Cohen & Wills, 1985; Holahan, 1987) in many different populations over decades (Billings & Moos, 1981; Thoits, 1986; Ystgaard, Tambs, & Dalgard, 1999). Other potential resources for higher education students are factors related to better learning, for example, persistence, intrinsic motivation, interest, and self-efficacy (Garner, Alexander, Gillingham, Kulikowich, & Brown, 1991; Harackiewicz, Barron, Tauer, & Elliot, 2002; Schiefele, 1991; Zajacova, Lynch, & Espenshade, 2005) and factors that reduce daily hassles, such as financial worries and time conflicts (Oropeza, Fitzgibbon, & Barón, 1991; Richardson et al., 2012; R. Roberts, Golding, Towell, & Weinreb, 1999).

10.2.3 Potential Effects of Perceived State and Trait Stress on Physical and Mental Health

Evidence for differential effects of short-term versus long-term stressor exposure comes from research on chronic physiological stress responses, and studies on perceived acute and chronic stress. Models of allostatic overload emphasize that prolonged stressor exposure, resulting in chronic stress, jeopardizes physical and mental health (e.g., Juster et al., 2011; Juster et al., 2010; McEwen, 2008). These negative effects are possibly mediated by changes in diurnal cortisol activity, which are elicited by chronic stress as found in a recent meta-analysis (Miller, Chen, & Zhou, 2007).

Acute and chronic psychological stress are associated with a range of negaitve health outcomes, such as worse immunity (Segerstrom & Miller, 2004), general mental health (Bovier, Chamot, & Perneger, 2004), depressive symptomalogy (McGonagle & Kessler, 1990), migraine attacks (Köhler, 1990), and phantom pain (Arena, Sherman, Bruno, & Smith, 1990). Similar to allostatic load models, exposure to longer lasting stressors, such as unemployment, or stressful life situations following a time-limited stressor, such as losing a loved-one, seems to have more negative consequeneces than relatively time-limited stressful situation (e.g., Hammen, Kim, Eberhart, & Brennan, 2009; Van Der Ploeg & Kleber, 2003). Considering for example changes in the immune system, acute, time-limited stressors, such as giving a speech in a laboratory study, were associated with immediate adaptive changes in immunity, while longer lasting stressors, such as preparing for examniations in higher education, were associated with immune suppression, and more chronic stressors or a sequence of challenges after the occurrence of a specific stressful event were associated with further maladaptive changes in immunity (Segerstrom & Miller, 2004).

In line with the evidence of health effects of actute and chronic stressor exposure, we expect to find that both, perceived state and trait stress in our study are negatively related to physical and mental health. Furthermore, we hypothesize that trait stress will be a stronger predictor of health than state stress because of the detrimental health effects that have been reported for chronic stress. To date there is no study that explicitely investigated the effects of state and trait stress on mental and physical health. Furthermore, most studies investigated the effects of perceived stress on mental health and the evidence on effects of perceived stress on somatic symptoms is scarce.

10.3 The Current Studies

In sum, there is ample evidence of perceived stress, its predictors and effects on health and psychological functioning in and outside higher education. However, to our knowledge no study has investigated stress in higher education students using latent state-trait analysis. Most studies on stress were conducted in the workforce or with clinical populations. Evidence on stress of students in higher education is almost exclusively correlational and rarely included more than one measurement point. Furthermore, it is unclear whether negative effects of stress on health outcomes and academic achievement differ for state and trait stress. In order to close these gaps in the research literature, we conducted two studies. In Study 3a, we measured students' levels of perceived stress at two points in time and used structural equation modeling to test the relations among personal resources, coping-strategies and the Big Five personality dispositions and three latent stress factors, i.e., two state and one trait factor. The data were assessed longitudinally at the beginning (T1) and at the end (T2) of the academic year, about nine months apart, and analyzed by means of an LST model of stress. In Study 3b, we assessed stress levels repeatedly over the course of one semester at ten measurement points, which were approximately two to four weeks apart. We averaged the data points to five stress states, computed an LST model and investigated the relationships between latent state and latent trait stress, mental health, somatic symptoms, and academic achievement at the end of the semester. In both studies, the participants came from a wide range of bachelor’s and master’s programs at a medium-sized university in a medium-sized German city.

Based on the previous studies, we had six hypotheses. We hypothesized a good fit of the LST model in both studies (Hypothesis 1a). We further hypothesized a good fit of the LST model with strong measurement invariance on the second order factors, indicating that differences in stress across time reflect state-variability processes rather than trait changes (Hypothesis 1b). Also, we expected to find that stress is a state-like attribute, which means that the proportion of variance in the observed variables that is attributed to trait effects (common consistency) is lower than the proportion of variance that is attributed to state effects (occasion specificity) (Hypothesis 2). We further hypothesized that stable personality dispositions predict trait stress (Hypothesis 3). More specifically, we hypothesized that neuroticism positively predicts stress (Hypothesis 3a) and that extraversion negatively predicts stress (Hypothesis 3b). Furthermore, we expected that person × situation factors, i.e., perceived personal resources and coping strategies in Study 3a and perceived situational demands in Study 3b predict the state residual factors. Specifically, we hypothesized that in Study 3a problem-focused coping is negatively associated with stress (Hypothesis 4a), while emotion-focused coping is positively associated with stress (Hypothesis 4b), and personal resources and stress correlate negatively (Hypothesis 4c). In Study 3b, we expected that state stress is positively associated with situational demands (Hypothesis 4d). Furthermore, in Study 3b we expected that self-reported symptoms at the end of the semester are more strongly associated with trait stress than the stress state residuals (Hypothesis 5). Contrarily, we expected that academic achievement at the end of the semester is associated with the stress state residuals at the end of the semester but not with trait stress (Hypothesis 6).

10.4 Study 3a

10.4.1 Method

Participants. All 14,909 students at a mid-sized university in a mid-sized German town were informed about the study via email and could volunteer to participate. Three-hundred-thirty-six students volunteered at the first measurement point. Most students (68%) were enrolled in study programs relating to arts, language and cultural sciences. The most common programs in our samples were Psychology (34%) and Educational Science (14 %). Twenty-three per cent were enrolled in one of the law and economic science programs, and about 9% were studying mathematics or science. Seventy-seven per cent of the students were female and about 5% were foreign students. Ninety-three per cent of the students were from lower semesters (1 to 9). Of the 336 individuals who participated at T1, 225 also participated at T2; so the dropout rate was 33%. Students who dropped out between T1 and T2 did not differ significantly from students who did not drop out on any of the demographic characteristics and psychological scales with the exception of conscientiousness ($t(334) = 2.347$, $p = .019$, $d = .272$), and neuroticism ($t(334) = 2.327$, $p = .021$, $d = .270$). Higher conscientiousness and neuroticism were associated with participation at both measurement points. However, we see no reasons to expect that this would bias the findings with respect to the relations between protective factors and stress.

Procedure. Data were collected via an online survey a few weeks after the beginning of the winter semester, in November 2013 (T1), and again near the end of the summer semester, in July 2014 (T2), approximately at the middle of the summer semester in Germany. After receiving information about the study and giving informed consent to their participation, students completed the online questionnaire at home. Participation in the study was compensated with 24 Euro and we hold all participants' written consent on file. The descriptives and internal consistencies of all scales are reported in Table 9. We analyzed the data from all 225 participants, as there was no missing data on either of the variables and we did not have to exclude any of the participants. The online survey included questions regarding demographic characteristics and the measures described below.

Measures.

Stress. To assess students' self-experienced stress levels, we used the German version (Engling, 2010) of the four-item version (PSS-4, as reported in Warttig, Forshaw, South, & White, 2013) of the *Perceived Stress Scale* (PSS; S. Cohen, Kamarck, & Mermelstein, 1983).

The participants rated the frequency of their stress experiences during the last three months on a 5-point Likert scale ranging from 1 = *never* to 5 = *very often*.

The different versions of the PSS are ubiquitously used in research on subjectively experienced stress levels. The scale was developed and validated with large samples of college students and is well suited for, but not restricted to, research in higher education. A recent review found that the psychometric properties of the different versions of the PSS were acceptable with mostly good internal consistencies (Cronbach's α > .70) and moderate to strong correlations with criterion variables, such as depression and anxiety (E.-H. Lee, 2012). Even though, the psychometric properties of the PSS-10 and PSS-14 were found to be superior, the PSS-4 suited our need for a quick assessment of stress in Study 3a.

Well-Being. Even though our study did not primarily focus on well-being, we presented the participants with the items of the *WHO-Five Well-being Index* (Bech, 2004). This allowed us to correlate stress and well-being scores as a plausibility check in our analyses, because a strong negative correlation between these two constructs has been firmly established in previous empirical studies (e.g. Carmody & Baer, 2008).

Personality traits. Personality predispositions were assessed using the German short version (BFI-K, Rammstedt & John, 2005) of the *Big Five Inventory* (John, Donahue, & Kentle, 1991; John, Naumann, & Soto, 2008; Rammstedt, 1997). The BFI-K is a 21 item short scale used to assess five central dimensions of personality (i.e., neuroticism, extraversion, openness, agreeableness, and conscientiousness). Participants gave their answers on a 5-point Likert scale.

Personal Resources. We constructed test items in order to assess a range of student resources, including social support from peers and family, competence-related, interest-related and context-related resources. Most items were statements about having sufficient or insufficient amounts of a resource. For example, one statement was *I'm satisfied with my current financial situation* (context-related). The participants rated how well each item described their situation on a five-point Likert scale from 1 = *doesn't apply* to 5 = *applies*. Only one item did not fit this format, that is, students' reported hours of extracurricular work. These were reverse coded by subtracting the number of hours worked per week from the maximum number of hours per week in our survey, which was 40. All items were transformed to lie on the same scale by z standardizing them individually. Higher values indicated more available resources. The resulting scores were then averaged and combined for T1 and T2 separately. The items, their mean values, and standard deviations can be found in Appendix C (SM 9).

Coping Strategies. We also assessed the extent to which students employed emotion-focused and problem-focused coping strategies. We did this by using the German 36-item short version (Ferring & Filipp, 1989) of the Ways of Coping Checklist (Folkman & Lazarus, 1980). Participants were instructed to think of the most stressful study situation they had experienced during the last two months and how they had dealt with it. The participants read descriptions of 18 emotion-focused and 18 problem-focused coping strategies and indicated for each whether they had used it or not

10.4.2 Results

Descriptives. Table 9 shows the means and standard deviations of the items of the stress scale and of the mean scores for the protective factors and personality. All variable distributions, except resources at T1, significantly deviated from normality according to the Kolmogorov-Smirnov test. However, the skewness and kurtosis indicated only slight deviations from normality. None of the variables showed ceiling or floor effects. We computed the internal consistencies of the PSS scores, which were good, $\alpha = .789$ at T1 and $\alpha = .839$ at T2. The stress sum scores at T1 and T2 correlated negatively with the WHO-Five Well-being Index in our sample, $r = -.579$, $p < .001$, for T1 and $r = -.609$, $p < .001$, for T2.

Table 9

Study 3a: Descriptive Statistics of the Observed Variables, Cronbach's Alphas for the Scales, Cohen's ds for Changes from T1 to T2, and Correlations r between T1 and T2

	T1			T2			$d_{T1\text{-}T2}$	$r_{T1\text{-}T2}$
	M	*SD*	α	*M*	*SD*	α		
PSS Item 1	2.960	.950	-	2.890	1.018	-	0.090 *ns*	.403**
PSS Item 2	2.610	.814	-	2.600	.819	-	0.033 *ns*	.446**
PSS Item 3	2.690	.903	-	2.640	.875	-	0.070 *ns*	.455**
PSS Item 4	2.800	1.070	-	2.690	1.090	-	0.114 *ns*	.508**

Resources	1.969	.377	.794	1.970	.398	.821	-.025 *ns*	.718**
Emotion-focused coping	.493	.198	.725	.479	.202	.735	0.089 *ns*	.512**
Problem-focused coping	.551	.179	.670	.564	.187	.703	-0.035 *ns*	.512**
Extraversion	3.563	.836	.787	3.616	.824	.811	-.096*	.837**
Conscientiousness	3.569	.674	.693	3.65	.680	.697	-.031 *ns*	.695**
Neuroticism	3.283	.919	.778	3.303	.942	.808	.067 *ns*	.801**
Openness	3.898	.707	.728	3.890	.747	.758	.006 *ns*	.750**
Agreeableness	3.100	.822	.680	3.047	.846	.711	.035 *ns*	.768**

ns not significant; * $p < .05$; ** $p <$; PSS refers to the Perceived Stress Scale.

T-tests were performed to investigate mean differences in stress, coping, personality, and resources between T1 and T2 (see Table 9). Despite the high statistical power that comes with a large sample size, only the mean difference of extraversion between T1 (M = 3.563, SD = .833) and T2 (M = 3.616, SD = .824) was statistically significant but the effect was small, t = -2.532, p = .012, d = -.096. The other mean differences were insignificant, suggesting that neither group average personality scale scores, nor stress levels, nor group average levels of the protective factors changed over time.

Reliability. Table 9 also reports the Cronbach's alphas of all scales. Regarding personality, resources and coping, all alphas were $.700 \geq \alpha \leq .942$, which shows acceptable to high internal consistency (Nunnally, 1967), except for problem-focused coping and conscientiousness at T1.

Latent State Trait Model.

Model Specification. We prepared the data for structural equation modeling by averaging the scores of the personality variables at T1 and T2 and forming item parcels of the PSS for reasons of model parsimony. The first parcel of the PSS included items 1 and 4, and the second parcel included item 2 and 3. Thus, the first parcel contained items reflecting "being overwhelmed" while the second parcel contained the reverse coded positive items which reflect a tendency to feel in control of one's life.

While the two first-order factors represent situation-specific variance components, i.e., state stress at T1 and T2, the second-order factor models variance that is common to both measurement, i.e., trait stress. Based on our hypotheses, we specified regression paths from the person × situation protective factors (resources and coping strategies) at T1 and T2 to state stress at T1 and T2, and regression paths from the stable person factors (personality dispositions) to trait stress. We allowed the correlations among the predictor variables to be freely estimated, specifically, the correlations between person × situation factors at T1 and T2 and the personality dispositions, and the pairwise correlations between the person × situation factors at T1 and T2. We also included the pairwise correlations between the item parcels of stress at T1 and T2 into the model to allow for indicator specific effects. The structure of the LST model is displayed in Figure 9. Note that we omitted the correlation paths to enhance the clarity of the illustration. We used the robust maximum likelihood estimator MLR to estimate the model (L. K. Muthén & Muthén, 1998-2016) in MPlus Version 7.31 (B. O. Muthén & Muthén, 1998-2012).

Measurement Invariance. We first tested for measurement invariance of the latent state and trait factors following the recommendations by Widaman, Ferrer, and Conger (2010) and Geiser et al. (2015). This was done by estimating a series of five models with increasingly stricter parameter constraints. First, measurement invariance was tested on the level of the second-order factor by comparing the models M1, M2, and M3. M1 was the baseline model in which all parameters were freely estimated (configural invariance). In M2, only the factor loadings on the latent trait factor were constrained equal (metric invariance); in M3, the intercepts of the first-order factors on the latent trait factor were also constrained equal (scalar invariance). In M4 additionally the factor loadings of the item parcels on the first-order factors were constrained equal; and in M5 the constraints of strong measurement invariance were applied on the first-order factors. M3 was the model, which displayed the lowest values on the Akaike and Bayesian Information Criteria, which indicates the best relative fit according to these indices. However, Satorra-Bentler corrected chi-square difference tests were computed for M1 to M5 and indicated that none of the constraints significantly worsened the model fits (see last column of Table 10).

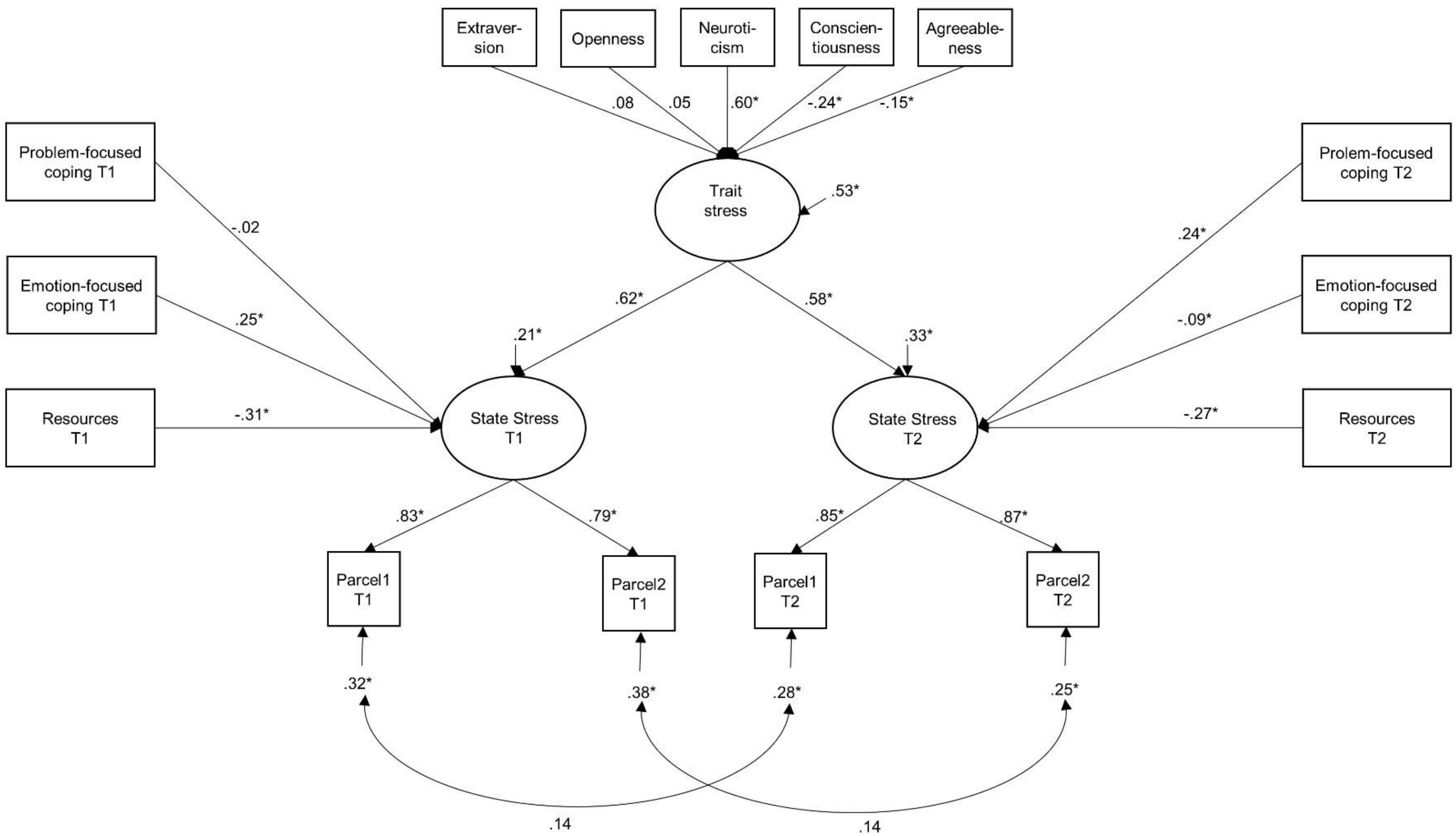

Figure 9. Path diagram of model M4 in Study 3a. For reasons of clarity we omitted the correlation paths among the predictors.

Table 10

Study 3a: Model Fits and Results of Chi-Square Difference Tests for the Models M1 to M5

Model	Assumptions	χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC	χ^2 difference test			
										Compared against	$\Delta\chi^2$	Δdf	*p*
M1	no equality constraints	52.660	31	.009	.969	.046	.027	5350.790	5747.769	-	-	-	-
M2	second-order factor: metric invariance	52.354	32	.013	.971	.044	.027	5348.792	5741.955	M1	0.827	1	.363
M3	second-order factor: scalar invariance	54.032	33	.012	.970	.044	.027	5348.587	5737.932	M2	1.348	1	.246
M4	first-order factors: metric invariance	57.288	34	.008	.970	.043	.027	5347.175	5345.807	M3	0.711	1	.399
M5	first-order factors: scalar invariance	57.985	35	.009	.971	.042	.027	5732.704	5727.518	M4	0.641	1	.423

Thus, for reasons of parsimony, the high level of measurement invariance and excellent model fit, we chose M5 for our further analyses (see Table 10 and Figure 9). In M5 the state residual factors were substantially related to trait stress as indicated by their factor loadings (λ = .615, p < .001, for T1; λ = .558, p < .001, for T2). These findings lend support to Hypothesis 1a and demonstrate that the distinction of a state component and a trait component of subjectively experienced stress is valid. Overall, the finding of measurement invariance for the indicators of the state stress factors indicates that stress was assessed in a comparably ways at the two measurement points. Furthermore, the finding of time-invariant intercepts and loadings on the second-order factor supports the assumption that potential changes in student stress over time reflect state variability processes instead of trait changes of perceived stress (Geiser et al., 2015) (Hypothesis 1b).

Common consistency, occasion specificity and reliability. We further computed common consistency, occasion specificity and reliability according to the formulae given by Steyer (1999). This was done in order to quantify the proportions of variance in the observed variables due to trait effects, situation effects and unreliability. The reliability coefficient reflects the proportion of variance that is free of measurement error, which was .620 in our sample (see Table 11). Contrary to our expectations (Hypothesis 2), the largest proportion of variance was not explained by occasion specificity, which was only .153. Thus, individual differences in students' stress levels were mainly explained by trait differences, which is shown by a common consistency of .467.

Predictors of Trait Stress. Neuroticism exhibited the expected positive association with trait stress (Hypothesis 3a), β = .597, p < .001, whereas the expected negative relation between extraversion and stress (Hypothesis 3b) was not found. Instead trait stress showed significant negative relations with conscientiousness, β = -.240, p = .002, and agreeableness, β = -.153, p = .043. Altogether, the amount of explained variance in trait stress was 47%.

Predictors of State Stress. Against our prediction (Hypothesis 4a), problem-focused coping did not go along with less stress (T1: β = -.017, p = .710; T2: β = -.089, p = .122). However, Hypothesis 4b and Hypothesis 4c were supported, as emotion-focused coping positively predicted stress (T1: β = .246, p < .001; T2: β = .239, p < .001) and resources were negatively related to stress (T1: β = -.310, p < .001; T2: β = -.268, p < .001) as expected.

The proportions of explained variance in state stress in M5 were very high: 79% at T1 and 67% at T2. However, in order to determine the explained variance by the predictors alone, without the variance explained by trait stress, we computed an additional model M5.1. In M5.1

the state and trait factor loadings and intercepts were fixed to the values obtained from M5 and only the personality variables but not the situation-specific predictors were included. In this model 68% of the variance of stress at T1 and 57% of the variance of stress at T2 were explained. This means, that the situation-specific predictors explain a unique amount of variance, which is the difference in R^2 between M5 and M9, that is, 79% - 68% = 11% at T1 and 68% - 57% = 11% at T2.

Table 11

Coefficients of reliability, common consistency, method specificity, and occasion specificity for student stress in both studies

	Study 3a	Study 3b
Reliability (Y_{ik})	.614	.639
Common Consistency (Y_{ik})	.471	.130
Occasion Specificity (Y_{ik})	.142	.471
Method Specificity	-	.038

Note. Y_{ik} refers to the aggregated indicators over all measurement occasions.

10.4.3 Discussion

The main focus of Study 3a was to provide empirical evidence for the distinction between stable and situationally varying components of stress. The good model fit and the results of the invariance testing of the measurement model support the assumption that a stable and situationally variable level of stress can be distinguished. More specifically, time invariant trait and state factor loadings and intercepts indicate that changes in students' stress levels over time reflect mere situational fluctuations but no changes of an underlying disposition to experience stress. As expected, personal resources and emotion-focused coping were related to state stress at the start and the end of the academic year, which was indicated by the significant regression coefficients of these predictors. Unexpectedly, problem-focused coping was unrelated to stress. However, our findings are in line with previous studies in the field of higher education, in which problem-focused coping was unrelated to negative affect (Eisenbarth et al., 2013) and adjustment to university life (Halamandaris & Power, 1999). Studies outside the field of higher education found that problem-focused coping is only helpful when people have some degree of control over the situation, and that in uncontrollable situations at least some emotion-focused strategies (e.g., accepting) are essential (Compas, Malcarne, & Fondacaro, 1988). These findings underpin the claim that the distinction between problem- and emotion-focused coping is merely theoretical and that, in practice, the dynamic interplay of different strategies is necessary to cope effectively with stress (Folkman & Moskowitz, 2004; Lazarus, 2000). Neuroticism had a negative impact on stress as expected and was the strongest predictor of students' stress levels. Unexpectedly, extraversion had no effect, while higher scores of conscientiousness and agreeableness were associated with lower stress levels. Students who are more conscientious may be better-organized and show better performance and thus, experience less stress. Students who score higher on agreeableness and conscientiousness may be more cooperative and make more effort, which translates into better academic performance (Poropat, 2009). These students might in turn also experience less stress.

Our overall findings of the close relation between state and trait stress, person and person × situation protective factors complement earlier studies on stress in higher education, which mainly focused on situational stress levels and potential stressors. Our findings support a dual approach against stress in higher education. First, stress could be reduced by removing stressors where possible and by strengthening protective factors where removing stressors is not possible. Second, stress could be reduced by changing individual tendencies to experience stress, possibly by changing the way students' perceive stressors and stress. Our findings regarding the importance of protective factors in higher education are well in line with earlier evidence collected outside higher education in research on work stress (Holahan, 1987; Holahan & Moos, 1990; Holahan, Moos, Holahan, & Cronkite, 1999) and resilience from the perspective of Positive Psychology (cf. Snyder et al., 2002).

Against our hypothesis, individual differences in stress levels in Study 3a were mostly explained by trait differences and not occasion specific factors, such as the situation and the interaction between the person and the situation, which was reflected by the high proportion of common consistency and the relatively low proportion of occasion specificity. However, as the participants in Study 3a were instructed to indicate their stress experiences during the last three months, the state stress residuals may reflect a more global level of stress than expected.

We observed self-selection effects regarding study participation and dropout. Female students, German residents, students from lower semesters, and students from certain programs were overrepresented in our sample compared to the population. Dropout in the study was not related to any of the demographic characteristics, stress, coping, and student resources. Regarding personality, only students who scored higher on neuroticism and conscientiousness were less likely to drop out of the study between T1 and T2. Despite the selective dropout, there were no significant mean differences in any of the variables between T1 and T2, so we are confident that dropout does not affect the interpretability of our results.

A number of findings demonstrate the trustworthiness of our findings in Study 3a. We used standardized questionnaires for stress, coping and personality. Our analyses had the high statistical power that comes with a large sample. Stress was modeled as a latent factor in order to reduce measurement error. The two item parcels of the PSS-4 items had good convergent validities, as indicated by their factor loadings greater than .60. There were substantial individual differences and no signs of ceiling or floor effects for all variables. Stress was negatively associated with well-being, which provided some evidence for the validity of the dependent variable.

10.5 Study 3b

In Study 3b, we used similar methods as in Study 3a but we assessed student stress and its covariates at ten measurement points over the course of one semester. At T1 participants were instructed to indicate their stress during the last 12 months, and at T2 to T10 we instructed the participants to indicate their stress during the last week. By the end of the semester, we additionally assessed mental health and somatic symptoms. After the end of the semester, we asked the participants to provide us with a printout of their current record of achievement. Study 3b allowed us to investigate Hypotheses 1 to 3 in more detail and investigate Hypothesis 5 and 6 concerning the differential effects of trait and state stress on health and academic achievement.

10.5.1 Method

We used the same measures as in Study 3a with the following exceptions: Rather than using the four-item version of the PSS (S. Cohen, Kamarck, & Mermelstein, 1983), we used the ten-item version (German PSS-10; Engling, 2010) in order to assess students' stress levels. The PSS-10 has the best psychometric properties of all three versions of the scale as reported by E.-H. Lee (2012). Also, rather than using the short version of the German Big Five Inventory (Rammstedt & John, 2005), we used the longer, original scale (Rammstedt, 1997).

We additionally assessed self-reported psychological and somatic symptoms using the German version (Franke & Derogatis, 2002) of the Symptom Checklist-90-revised (Derogatis & Unger, 2010) at T10 and grades and credit points at the beginning of the following semester. Furthermore, we asked students to indicate which demands they experienced out of a list at T2 to T10. The list of demands was modified from a German handbook of a stress-reduction intervention (Hammer, 2006) for the assessment of stressors in a student population. The items are listed in Appendix C (SM10).

Participants. As in Study 3a, we called for students to volunteer in our study via the university mailing list. They provided us with their email addresses and we sent the link to the online survey first to the 100 respondents. Eighty-five students responded at T1. Most participants were enrolled in one of the arts, language and cultural sciences programs. Most participants were Psychology students (31%) or Educational Science students (10 %), fewer were studying German Studies, English Studies, Classical Philology, Business Studies, Economics, Mathematics, Media Studies, Law, Political Science, History, Art History, and others. Forty-two per cent were studying in a bachelor's program, 35 % in a master's program, 23 % did not indicate their aspired degree or had missing data. Approximately 81 % of the participants indicated that they were female. Ninety-seven per cent of the students indicated that their nationality was German. Of the initial $N = 85$ participants, only two participants completely dropped out of the study after T1, 14 completed the survey at five to nine time points and 69 participated at all ten time points. At each of the time points T2 - T10, there were between 73 and 77 respondents to the survey.

Procedure. As in Study 3a, we collected data via an online survey. The ten measurement points were distributed over the course of the fall/winter semester 2016/2017. T1 data were collected between October 17 and October 23, 2016, which was the week before the beginning of classes. T2 to T7 data were collected between the beginning of November and the end of January and the measurement points were two weeks apart. T8 was at the beginning of February, and T9 and T10 took place at the end of February and March, respectively, when there were no regular classes. Students were reminded of the data collection via e-mail and the online survey was approximately open for four to seven days at each measurement occasion. In order to complete the survey, students had to log in to the system, using their email-addresses. These data were not saved, instead students provided an anonymized code, which allowed us to assemble the data from different time points. At each measurement occasion we informed all persons which we had recruited for T1 (N = 100). After T10 we asked the students to hand in an anonymized printout of their record of achievement and rewarded their participation with 50 € for the participation at all ten measurement occasions and 40 € for their participation, when they had missed one to five measurement occasions.

10.5.2 Results

Descriptives. The descriptive statistics of the variables used in the analyses of Study 3b are reported in Table 12.

Stress. To compare our data to the PSS-10 norms (S. Cohen, Kamarck, & Mermelstein, 1994) we recoded the items to lie on the original scale. The mean stress levels in the normative sample for 18 to 29 year olds was M = 14.2 [95% CI: 13.7, 14.7]. The mean stress levels in our sample from T1 to T10 ranged between M = 14.652 and M = 19.958 and thus, were all except for one time point (T6) above the upper limit of the confidence interval reported by Cohen and colleagues, indicating that participants in our sample felt more stressed than the norm sample. We found that the stress means across the ten time points were in good agreement with phases of tension and relaxation during the semester: the means were lower during the semester break than in the time when the students attended lectures and classes. The lowest mean stress level was observed at T6, M = 2.446, which was right after the Christmas holidays, which was a two week break during the winter semester. The highest mean was observed at T9, M = 2.996, which was during the exam period and the means during the lecture time were between M = 2.694 and M = 2.811. We tested whether the mean differences were significant in a one-way repeated-measures ANOVA. As the assumption of sphericity had been violated, χ^2 (44) = 103.450, p < .001, we used the Greenhouse-Geisser corrected estimate of the F-value, which was significant, F(6.659, 399.563) = 4.670, p < .001, $\eta^2 p$ = .072. This indicates that the means of the stress levels significantly differed over the course of the semester. Rather than conducting the analyses with the ten stress states, for model simplicity, we averaged two time points each and computed item parcels as in Study 3a before the main analyses in Mplus. Parcel 1 of each time point comprised six items, which indicated the experience of being overwhelmed by demands and (including items 1, 2, 3, 6, 9, 10). Parcel 2 comprised the four remaining items (items 4, 5, 7, 8), which were reverse coded and indicated having a sense of control of what was happening in one's life. We report the following results using the item parcels.

Table 12

Study 3b: Descriptive Statistics

	N	*M*	*SD*
PSS Parcel 1 T1	77	2.579	.602
PSS Parcel 2 T1	77	2.814	.660
PSS Parcel 1 T2	72	2.802	.667
PSS Parcel 2 T2	72	2.805	.827
PSS Parcel 1 T3	72	2.668	.680
PSS Parcel 2 T3	72	2.510	.734
PSS Parcel 1 T4	72	2.782	.700
PSS Parcel 2 T4	72	2.680	.744
PSS Parcel 1 T5	72	2.756	.621
PSS Parcel 2 T5	72	2.733	.770
Demands T1	67	55.105	5.480

Demands T2	72	54.271	5.229
Demands T3	72	51.986	5.921
Demands T4	72	52.944	5.704
Demands T5	72	52.646	6.160
Extraversion	77	3.312	.603
Conscientiousness	77	3.921	.530
Neuroticism	77	2.919	.720
Openness	77	3.547	.568
Agreeableness	77	3.734	.540
Grades	58	1.84	.826
Global Burden PSDI	72	1.56	.463

Normality tests. All stress item parcels were normally distributed according to the results of the Kolmogorov-Smirnov (K-S) test (all *p*s > .200) except for the second parcel at T2. The same was true for the Big Five personality variables and the five item parcels of demands, except for openness, and demands at T1, for which the K-S test was significant. However, for the two variables for which the K-S test was significant, the deviations from normality were rather small as indicated by the values of skewness and kurtosis (*z*-scores < 1.96).

Latent State-Trait Model. The latent state-trait model was computed in MPlus Version 7.31 (B. O. Muthén & Muthén, 1998-2012). To account for the significant, yet small deviations from normality, we used the robust maximum likelihood estimator MLR (L. K. Muthén & Muthén, 1998-2016).

Model Specification. We specified the initial latent state-trait model using the item parcels as indicators of the latent state factors with the overwhelmed-by-demands-parcel as marker indicator. To account for covariances in the residuals of the reverse coded item parcels, we formed a method factor. The model is depicted in Figure 10.

Measurement Invariance. As in Study 3a, we tested for measurement invariance as suggested by Geiser et al. (2015). Thus, we computed a series of models with increasingly stricter parameter constraints. The first model M1 was a model of configural measurement invariance. In the second model M2 the loadings of the first order factors on the second order factor were constrained equal across time (weak measurement invariance). In M3 additionally to the loadings, the intercepts of the second to fifth latent state factors on the trait factor were constrained equal across time, while the intercept of the first state factor was fixed to zero as well as the intercepts of the first indicator on each state factor. This enabled us to estimate the latent mean structure, including the mean of the latent trait variable and the means of the second to fifth stress states. In M4 additionally to the constraints of strong measurement invariance on the trait factor, the constraints of weak measurement invariance were implemented on the state residual factors and in M5 the constraints of strong invariance on the state residual factors were added to the model.

We compared the model fits of the four models and computed chi-square difference tests using the Satorra-Bentler correction (see Table 13). The chi-square test was insignificant models M1 to M4 and the values of CFI and SRMR indicated good fit, while the values of RMSEA indicated good fit in M1 and M2 (below .05), and acceptable fit in M3 and M4 (below .08). Chi-square difference tests indicated that there was no significant worsening in model fit between M1 and M2, M2 and M3, and M3 and M4. However, in M5 in which the constraints of strong measurement invariance on the state residual factors (M5) were implemented, all fit indices indicated a worsening of model fir and the chi-square difference test between M4 and M5 was highly significant, χ^2_{diff} (4) = 40.285, $p < .001$. Thus, M4 was the model, which was the best fitting and most parsimonious model, χ^2 (35) = 45.1, $p = .118$; *CFI* = .980; *RMSEA* = .061; *SRMR* = .080; *AIC* = 1017.631; *BIC* = 1087.945, and all further analyses were conducted using M4. The good fit of M4, which is a model that included strong measurement invariance on the latent trait factor and weak invariance on the state residual factors, further supports our Hypotheses 1a and 1b that students' stress levels show a state-trait structure and changes in stress levels reflect state variability processes instead of a trait change process

Consistency, Occasion Specificity, Method Specificity and Reliability. As in Study 3a we computed the common consistency, occasion specificity coefficients (Hypothesis 2). We found that in Study 3b reliability was comparable to Study 3a. Contrary to Study 3a and in line with our hypothesis, the largest proportion of variance in Study 3b was explained by occasion specificity, which was 47 % (see Table 11). Trait differences on the other hand explained only a small amount of variance in the observed variables, which is indicated by a common consistency of 13 %. The method factor accounted for only about 4% of the variance in the observed variables.

Predictors of Trait Stress. We computed a model M6, which included the invariance constraints of M4 and the Big Five dimensions of personality as predictors of trait stress (M6). M6 showed a slightly worse but still acceptable model fit, χ^2 (80) = 118.951, p = .003; *CFI* = .940; *RMSEA* = .080; *SRMR* = .073; *AIC* = 963.309; *BIC* = 1045.343. Neuroticism significantly predicted trait stress with a strong effect, β = .86, $p < .001$, which lends support to our Hypothesis 3a. As neuroticism was the only significant predictor, our Hypothesis 3b that extraversion negatively predicts stress was not supported. Personality altogether explained 68.5 % of the variance in the latent stress trait.

Predictors of State Stress. Another model, M7, in which we included the aggregated situational demands at the five time points showed an overall acceptable fit to the data, χ^2 (70) = 98.043, p = .015; *CFI* = .960; *RMSEA* = .072; *SRMR* = .091; *AIC* = 2843.908; *BIC* = 2996.256. In this model, we regressed each state factor on situational demands at the same time point, which were included as observed variables in the model. The situational demands at T1 to T5 were strong predictors of the state residual factors with significant, positive regression weights, which were of comparable size across time, $.694 \geq \beta \leq .750$ (Hypothesis 4d). Situational demands and the trait factor together explained a large amount of variance in the state residual factors, 73% at T1, 76% at T2, 75% at T3, 69% at T4, and 68% at T5. For each state residual factor the amount of explained variance was around 3-16% higher than in M4, indicating that situational demands explained a unique amount of variance of student stress.

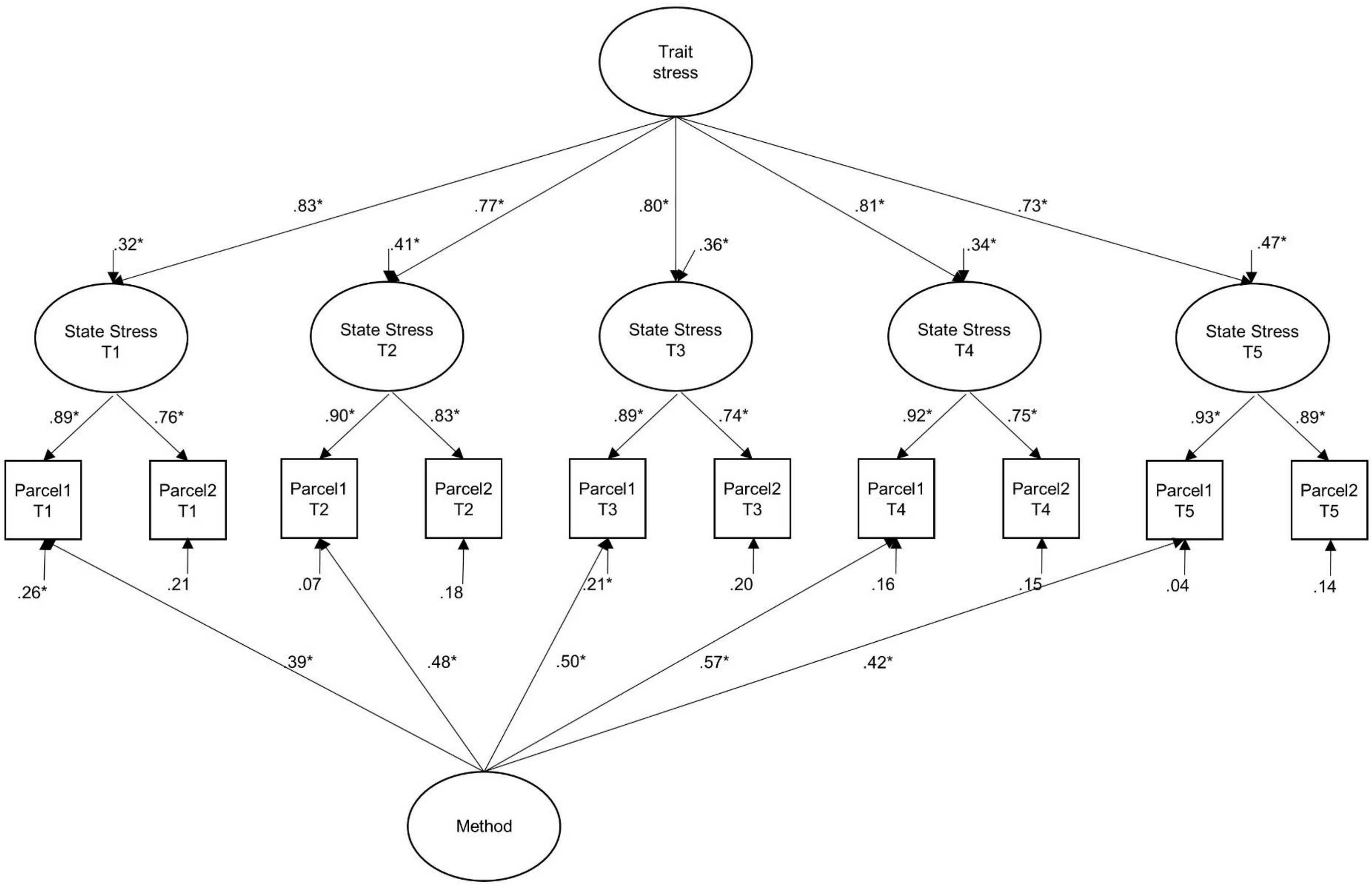

Figure 10. Path diagram of model M5 in Study 3b.

Table 13

Study 3b: Model Fits and Results of Chi-Square Difference Tests for the Models

Model	Assumptions	χ^2	*df*	*p*	CFI	RMSEA	SRMR	AIC	BIC	χ^2 difference test			
										Compared against	$\Delta\chi^2$	Δdf	*p*
M1	no equality constraints	27.910	25	.312	.994	.039	.044	1019.964	1113.716	-	-	-	-
M2	second-order factor: metric invariance	30.492	28	.340	.995	.034	.071	1016.482	1103.203	M1	2.567	3	.463
M3	second-order factor: scalar invariance	45.175	31	.072	.976	.071	.079	1023.341	1103.030	M2	4.373	3	.224
M4	first-order factors: metric invariance	45.095	35	.118	.980	.061	.080	1017.631	1087.945	M3	5.830	4	.212
M5	first-order factors: scalar invariance	82.128	39	<.001	.916	.120	.111	1046.260	1107.199	M4	40.285	4	<.001
M6	M4 including Big Five	118.951	80	.003	.940	.080	.073	963.309	1045.343	-	-	-	-
M7	M4 including demands	98.043	70	.015	.960	.072	.091	2843.908	2996.256	-	-	-	-
M8	M4 including health outcomes	64.066	43	.020	.963	.080	.083	1086.525	1166.215	-	-	-	-
M9	M4 including achievement	65.453	43	.015	.961	.078	.088	1224.625	1307.675	-	-	-	-

Health. We added the correlations between self-reported symptoms and trait stress, and symptoms and the stress state residuals at the end of the semester (T5) to our model M4 and ran the analysis (M8). The correlation between the overall number of self-reported symptoms and the state residual at T5 was r = .401 and the correlation between symptoms and trait stress was r = .501. The number of self-reported somatic symptoms showed also significant correlations with latent trait stress (r = .386) and the state residuals at T5 (r = .467). We tested whether the magnitude of the correlation between symptoms and state and trait stress differed for overall number of reported symptoms and number somatic symptoms. Satorra-Bentler corrected chi-square difference tests showed that there was no significant difference for the overall number of reported symptoms, χ^2_{diff} (1) = .106, p = .745, and the number of reported somatic symptoms, χ^2_{diff} (1) = .241, p = .624. Thus, our Hypothesis 5 that the relationship between self-reported symptoms and trait stress is stronger than the relationship between symptoms and occasion specific stress at the end of the semester was not supported.

We further analyzed the correlations between self-reported somatic symptoms; the stress residuals at T5 and trait stress by computing a separate model for each of the 12 symptoms in Mplus (see Table 14). We found that nine out of the 12 symptoms, including headaches, pains in lower back, nausea or upset stomach, trouble getting one’s breath, hot or cold spells, a lump in one’s throat and feeling weak in parts of one’s body, showed significant correlations with either the state stress residuals at T5, trait stress, or both ($.223 \leq r\text{s} \leq .461$). The strongest association was found between the state residuals at T5 and headaches.

Achievement. In order to analyze the associations between stress and achievement we computed M4 in which the correlations between trait stress and the state residuals at T5 and GPA of the current semester were added to the model (M9). In line with our Hypothesis 6, we found that GPA was associated with occasion specific stress at the end of the semester (r = .388, p = .003) but not with trait stress (r = .037, p =.765). These results indicate that a general tendency to experience stress does not per se translate into lower achievement but that stress stemming from the situation and person × situation interaction during the period of exams is associated with lower achievement.

Table 14

Correlations between self-reported symptoms latent trait stress and latent state stress state residuals at T5

	Trait Stress	State Residuals T5
PSDI	.496**	.409**
Overall number of somatic symptoms	.386**	.467**
Headaches	.256*	.461**
Faintness or dizziness	.096	.109
Pains in heart or chest	.243**	.226*
Pains in lower back	.207	.300*
Nausea or upset stomach	.246*	.397**
Soreness of your muscles	.052	.124
Trouble getting your breath	.253*	.050
Hot or cold spells	.262*	.237
Numbness or tingling in parts of your body	.148	.115
A lump in your throat	.307**	.291
Feeling weak in parts of your body	.250**	.389**
Heavy feelings in your arm or legs	.223*	.212

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

10.5.3 Discussion

Study 3b served to provide further evidence for our Hypothesis 1-4 about state variability processes, predictors of state and trait stress and to investigate Hypothesis 5 and 6 of health and achievement as correlates of trait stress and state residuals at the end of the semester. As in Study 3a the latent-state trait model showed an excellent fit to the data and supported our hypothesis of state variability processes rather than trait changes across time. Furthermore, in Study 3b we found support for the hypothesis that self-reported stress levels reflect a state-like rather than trait-like attribute because of the high proportion of occasion specificity and relatively low proportion of common consistency in the variance of stress levels. As expected, neuroticism predicted trait stress and situational demands predicted state stress at the same point in time. Unexpectedly, but as in Study 3a extraversion was not related to student stress. We found that self-reported health and GPA of the current semester were correlated with stress. As hypothesized only the state stress residuals were associated with achievement but not trait stress, which shows that situational stress factors and the interaction between personal and situational stress factors compromise students' achievement but not trait stress. Unexpectedly, more symptoms and burden were not only associated with more trait stress but also more occasion specific stress at the end of the semester.

Over and above the use of standardized questionnaires if possible, modeling stress as a latent factor and good convergent validities of the stress item parcels, Study 3b had several other strengths compared to Study 3a. Instead of using the short versions of the PSS-4 and the BFI-K we used the longer versions of these questionnaires which show excellent psychometric properties (Lang, Lüdtke, & Asendorpf, 2001; E.-H. Lee, 2012; Rammstedt, 1997). The model included five measurement points, which is in line with recent recommendations for LST analyses (Geiser et al., 2015). We tested for measurement invariance on the first and second order factors as recommended and found that there was a high level of measurement invariance, which is in line with a theoretical model in which the stress states fluctuate around a stable disposition to experience distress.

10.6 General Discussion

We investigated state and trait stress, their predictors, and their effects on health and achievement in higher education students using an LST modeling approach. Going beyond previous studies, we examined whether stress is a state-like attribute and whether changes in stress levels over time reflect effects of situational and persons x situation interactions effects rather than trait differences. Furthermore, we investigated the relationship between state and trait stress and two important groups of correlates, achievement and health with a specific emphasis on somatic symptoms. In the following, we would like to discuss the findings of Studies 3a and 3b and their insights on state and trait stress, the relations between stress, academic achievement, and health, the generalizability of our findings, and practical implications.

10.6.1 State and Trait Stress

One of our main findings was that the distinction of state and trait stress is valid, which was indicated by the excellent fit of the LST model of perceived stress. We had hypothesized this good fit based on the usefulness of LST models in research on cortisol levels and stress-related emotions. The loadings of the first-order factors (“state stress”) on the second-order factor for (“trait stress”) were, on the one hand, significantly greater than zero and, on the other hand, substantially less than one. In line with findings from LST models of cortisol levels (e.g., Giesbrecht, Bryce, Letourneau, Granger, & Team, 2015; Kertes & van Dulmen, 2012; Kirschbaum et al., 1990; Saxbe, 2008; Stroud, Chen, Doane, & Granger, 2016), anxiety (Spielberger & Sydeman, 1994), and anger (Forgays et al., 1997), this indicates that the state and trait variance components of stress are related but not identical. Furthermore, our results supported the distinction of state and trait stress, as they were differently related to person × situation protective factors and personality dispositions. Person × situation factors, such as perceived situational demands and resources, the use of coping strategies, and university achievement were related only to the state stress residuals at the respective point in time and not to the state stress residuals at other points in time or trait stress. As expected, stable personality traits, and in particular, neuroticism, were related only to trait stress and not to the state stress residuals. This is in line with the view that students with low neuroticism are generally more happy with their lives and tend to focus on positive stimuli (Diener, Oishi, & Lucas, 2003; Larsen & Ketelaar, 1991). The findings of stable and situationally varying factors that are closely associated with experienced stress can explain why, in large samples drawn from relatively homogeneous student populations, some students report high levels of stress while others report only mild or negligible levels of stress (e.g. Reisberg, 2000; Techniker Krankenkasse, 2015b, pp. 7-15).

Another central finding of our studies is the time-invariance of the assessment of experienced stress in our studies. The time-invariant loadings and intercepts of the state stress residuals on the trait factors is consistent with a theoretical model in which changing levels of students' stress over time are mere situational fluctuations and no changes in the general tendency to experience distress (Geiser et al., 2015). Despite we think that changes in the general tendency to experience distress are possible, it would have been implausible to find such changes in our studies considering the relative short periods of time in which the studies were conducted.

However, from the results of our studies it is not completely clear whether students' stress levels are more affected by situationally varying factors or trait differences. Study 3a suggests that trait effects are stronger than occasion specific effects, whereas we found the opposite in Study 3b. As Study 3a included only two measurement points, which were several months apart and participants were instructed to report their stress during the last three months versus one week in Study 3b, we believe that both, the instruction and the frequency of the assessment can affect the estimates of common consistency in our samples. Thus, we see the results in line with our assumption that stress is a state-like attribute, when participants consider relatively short periods of time when reporting their stress levels. In this case, participants' stress levels depend mostly on how they appraise their current situation including situation-specific demands, their available resources and coping efforts. However, further work is needed to investigate if and how frequency and timing of different measurement occasions in longitudinal studies can affect the measurement of state-like and trait-like attributes.

10.6.2 Stress, Health, and Academic Achievement

The results of our studies confirm the presumptions from previous research that even short-term stress goes along with worse health and lower academic achievement in higher education students. As expected, trait stress was also associated with a higher number of self-reported symptoms, including among others depression, anxiety and somatic symptoms. Moreover, a substantial amount of somatic symptoms was related to the experience of distress, both trait stress and state stress. Unexpectedly, the association between symptoms and trait stress was not stronger than the association with state stress at the same point in time.

This might be due to the fact, that we considered a relatively short period of time in Study 3b. Trait stress experienced over the course of one or several academic years may be more strongly related to health than concurrent state stress. Another explanation might be that state stress could also result from the experience of specific symptoms, for example, headaches or breathing problems. Our analyses of the relationship between stress and specific somatic symptoms indicated that stress can go along with a broad range of somatic symptoms, and that the association between stress and headaches is particularly strong. Future research should clarify whether there is a causal relationship between state stress, trait stress and self-reported symptoms and whether trait stress makes students more vulnerable for specific conditions.

Consistent with our expectations, GPA in the semester of our data collection was only associated with state stress at the end of the semester and not with trait stress. There are good reasons for these relations. First, neuroticism, which was closely related to trait stress in our studies, has been found to have only a minor or no association with academic performance in meta-analyses of personality and performance in secondary and post-secondary education (O’Connor & Paunonen, 2007; Poropat, 2009). Even though, neuroticism goes along with test anxiety and thus, may impair performance (Chamorro-Premuzic & Furhnam, 2006), moderator effects in a meta-analysis by Poropat (2009) showed that the association between neuroticism and performance decreases in higher educational levels. Possible reasons for this moderator effect are a less biased assessment of performance in higher education and students’ longer learning histories, in which they have acquired skills and strategies to compensate for their neuroticism.

10.6.3 Generalizability of Findings

We expect that future studies will be able to show a high generalizability of our results. The sample size of the first study was large (N = 225). Even though our sampling procedure was not free of self-selection effects, our samples were not restricted to a specific study phase or program, as in most previous studies. Our samples comprised students from a wide range of semesters (1 to 20) and more than 20 bachelor’s and master’s programs. Our findings concerning the relationship between stress, person × situation factors and stable personality traits were highly consistent over Study 3a and Study 3b, suggesting a good replicability of our findings. Measurement invariance was found in both studies for the indicators of state stress. The generalizability of our findings is supported by the fact that we assessed stress, coping strategies, personality traits, mental health and somatic symptoms with established and widely used standardized tests. In particular, the PSS is ubiquitously used in psychological stress research and had been found to work well in many populations differing in nationality, health status, or personal backgrounds (S. Cohen, 1988; S. Cohen & Janicki-Deverts, 2012; Warttig et al., 2013), especially in student populations (E.-H. Lee, 2012). The PSS is associated with many psychological and even physiological variables related to stress (Ebrecht et al., 2004; Hewitt, Flett, & Mosher, 1992; Leung, Lam, & Chan, 2010; J. C. Pruessner et al., 1999), raising the hypothesis that the LST structure found in our study with the PSS can be found with alternative stress measures as well.

However, the generalizability of our findings is limited in that we used convenience samples from a single university in a single country, included almost no science programs, and had a small sample size in Study 3b. Testing the generalizability of the current findings in different institutional contexts, schools and universities for example, and different cultures, remains as an important task for future research.

10.6.4 Practical Implications

The finding of state and trait stress and their predictors from our studies has practical implications for higher education institutions and mental health services. Higher education institutions can reduce state stress by reducing demands and strengthening protective factors. There already is a number of available effective stress reduction interventions for affecting protective factors and students' stress levels in higher education. A meta-analysis of 24 studies with 1,431 higher-education students found that cognitive, behavioral, and mindfulness-based stress reduction trainings reduced stress levels with a Cohen's d of -0.77 [95% CI: -0.88, -.58] (Regehr, Glancy, & Pitts, 2013). Three of the studies also found a negative effect of the trainings on salivary cortisol, d = -0.52 [95% CI: -0.83, -0.20].

Our study results show that it is worthwhile to consider measures that reduce the general tendency to experience stress, as trait stress predicted students' mental and physical health at the end of the semester. Trait stress-reduction interventions could focus on neuroticism. There is evidence that pharmaco-psychotherapy reduces neuroticism in patients diagnosed with depression (De Fruyt, Van Leeuwen, Bagby, Rolland, & Rouillon, 2006; Tang et al., 2009). More research is needed whether these changes go along with lower trait stress levels, enhanced well-being and better health.

11. General Discussion

Despite the fact that there is a large body of research on determinants of academic achievement in higher education, there was a lack of meta-analytical and longitudinal investigations of the relationships between prior knowledge, stress, and academic achievement. Thus, in the course of this dissertation, three studies were conducted to close these gaps: a meta-analysis of prior knowledge and learning, a latent transition analysis of multivariate knowledge trajectories in first year undergraduates, and a longitudinal study of state and trait stress over the course of one semester and their relationships to academic achievement in a sample of university students. In the following chapter, I present key insights from these three studies, discuss methodological and practical implications, give recommendations for future research, and provide a final conclusion.

11.1 Key Insights

From the three studies of this dissertation, six key insights emerged which I present in the following. As Studies 1 and 2 investigated the relationship between prior knowledge and achievement, the first four key insights relate to knowledge, whereas the last two key insights are derived from Study 3 and relate to stress.

The first two key insights stem from Study 1 and concern the average effect of prior knowledge on learning. First, domain-specific prior knowledge predicts learning outcomes on all educational levels, including higher education. Second, the association with prior knowledge does not differ between the learning outcomes indicating domain-specific knowledge and achievement in an academic domain.

Key insights 3 and 4 pertain to Study 2 and provide more detailed information about how the nature of prior knowledge affects learning and achievement inside higher education. Key insight 3 states that learners in the same domain show distinct differences in the content and organization of their structures of (prior) knowledge, and learning trajectories that indicate conceptual change processes. Key insight 4 states that, multivariate assessments of knowledge are useful to explain interindividual differences in academic achievement in higher education.

The last two key insights are derived from Study 3 which investigated stress and achievement in higher education by means of a latent state-trait model. Key insight 5 states, that students' stress experiences inside higher education can be explained by state and trait stress which are conceptually different. Finally, key insight 6 states that state stress and trait stress show differential relationships with academic achievement in higher education.

11.1.1 Key Insight 1

Study 1 addresses a timely issue, as the study of prior knowledge on later learning outcomes has been a vital aspect of educational research for many years (Dochy et al., 2002), yet, prior meta-analytical investigations had been confined to the predictive validity of grades, specific knowledge domains or educational levels (e.g., Duncan et al., 2007; Kuncel et al., 2001; Schuler et al., 1990). Thus, the average effect of prior knowledge for learning outcomes was unknown. Study 1 closes this gap by providing an average estimate of the effect based on 4327 effect sizes from 240 articles over a wide range of educational levels, showing that in general, prior knowledge has a strong, positive effect on learning outcomes, $r^{+} = .521$, representing a Cohen's *d* of 1.23. Thus, on average the effect of prior knowledge on learning is three times as large as the benchmark effect size of educational interventions that lead to real-word changes (Hattie, 2009). Whereas the effect of prior knowledge on learning outcomes is generally strong, the effect is significantly stronger in primary education than in higher education, as shown by the moderator analyses of Study 1. This leads to two conclusions for the role of prior knowledge in higher education. First, differences in prior knowledge between learners who enter higher education programs are highly predictive of their learning outcomes inside higher education. These findings are in line with previous meta-analyses (Kuncel & Hezlett, 2007; Richardson et al., 2012; M. Schneider & Preckel, 2017) and support the use of assessments of prior knowledge and achievement for selecting students for higher education programs. Second, besides domain-specific prior knowledge, there are other important factors that affect learning outcomes in higher education as the effect of prior knowledge is smaller than in primary education. This could be due to the fact that the contribution of school instruction to the highly specialized knowledge conveyed in higher education can only be limited. However, there are also statistical and methodological reasons that could explain the difference between primary and higher education, such as restrictions of variance in knowledge in higher education, or a confounding effect between socioeconomic status and knowledge in primary education.

11.1.2 Key Insight 2

Domain-specific prior knowledge predicts academic achievement, as the correlation between prior knowledge and assessments of achievement was high and did not significantly differ from the correlation between prior knowledge and subsequent assessments of domain-specific knowledge in the moderator analysis of Study 1. This is remarkable as achievement tests are designed to assess not only content knowledge, but various facets of academic achievement, such as reasoning skills, general scientific competencies, problem-solving skills, critical thinking, and analytical writing skills (Educational Testing Service, 2017; Kuncel & Hezlett, 2007; OECD, 2017). The results from Study 1 support the idea that (prior) knowledge is the basis for academic reasoning, critical thinking, problem-solving and other skills which are supposed to underlie students' performance in achievement tests. This is in line with research on expertise in which studies demonstrated that experts in various disciplines, such as chess (e.g., Chase & Simon, 1973a, 1973b), music (e.g., Kalakoski, 2007), and mathematics (e.g., Schoenfeld & Herrmann, 1982), displayed superior reasoning and problem solving because they have rich and well-organized content knowledge (Committee on Developments in the Science of Learning, 2004). While results from standardized achievement tests usually comprise multiple facets of learning outcomes in various domains (Steinmayr et al., 2014), domain-specific knowledge is a much narrower and more dynamic construct as knowledge can be changed by small interventions within weeks, days, or even hours. In conclusion, investigating the knowledge structures that underlie academic achievement is a very promising way to better understand learning and instructional processes that lead to students' academic achievement in and outside higher education.

11.1.3 Key Insight 3

Previous research has demonstrated that higher education students still hold many misconceptions, showing that there often is a need for conceptual change (e.g., Hughes et al., 2013; Shtulman & Calabi, 2013). However, few studies have investigated conceptual change processes quantitatively and these studies have been confined to learning situations outside higher education only (Edelsbrunner et al., 2015; Kainulainen et al., 2017; McMullen et al., 2015; M. Schneider & Hardy, 2013). Study 2 closed this gap as it provided insights into processes of conceptual change in higher education. The results of Study 2 show that even inside higher education, students show profound differences in their profiles of conceptual knowledge. Over time, a development from an initial understanding that is flawed by misconceptions and fragmented knowledge, towards correct, scientific conceptions and better integrated structures of knowledge can be expected for most students. Compared to the findings from studies outside higher education, there was a clearer developmental trend towards correct knowledge, which is in line with findings from other studies that showed that once domain-specific knowledge is acquired in higher education, it is relatively stable across time (Bahrick & Hall, 1991; Conway, Cohen, & Stanhope, 1992; Custers, 2010; Semb, Ellis, & Araujo, 1993). Second, there were very few students with pure misconceptions profiles. This can be explained by the fact that higher education students have long learning histories during which they have acquired scientific knowledge. Thus, at the beginning of instruction, inside higher education one may expect rather fragmented structures of correct and incorrect knowledge than pure misconceptions profiles.

11.1.4 Key Insight 4

Multivariate assessments of conceptual, domain-specific knowledge are useful to study differences in academic achievement in higher education as Study 2 provided compelling evidence that students who are more knowledgeable show higher achievements. Specifically, the results of Study 2 indicate that, students who hold fewer misconception, more correct knowledge and better integrated structures of knowledge achieve better grades. The correlation between grades and the knowledge profiles found in Study 2, which was a medium strong effect according to J. Cohen (1992), can be considered a lower-bound estimate of the relation between conceptual knowledge and academic achievement. This is because grades have some limitations when studying academic achievement as they can be multidimensional, comprising not only performance quality but also other factors, such as effort and aptitude (Stiggins, 2001). These confounding influences consequently limit the contribution of domain-specific knowledge to explain interindividual differences in students' grades which is in line with findings from previous studies (Griggs & Jackson, 1988; Plant et al., 2005; Xu et al., 2013). In conclusion, multivariate assessments of domain-specific knowledge are strongly related to students' academic achievement in higher education.

11.1.5 Key Insight 5

Several stress theories have emphasized that stress depends on stable long-term and dynamic situational factors (Bakker & Demerouti, 2007; Folkman, 2011; Hobfoll, 1989; Lazarus & Folkman, 1984), and research has shown that acute and chronic stress can have very different effects (Segerstrom & Miller, 2004). Yet, Study 3 is one of the first studies that used latent state-trait analysis to account for long-term and situational components of stress statistically. Study 3 indicates that both state and trait stress explain students' stress experiences in higher education and that state and trait stress differ conceptually and statistically. Furthermore, Study 3 indicates that there is considerable conceptual overlap in studies using self-report data and physiological indicator of stress (e.g., Doane et al., 2015; J. Hellhammer et al., 2007; Kertes & van Dulmen, 2012; Kirschbaum et al., 1990; Shirtcliff et al., 2005; Stalder et al., 2012). Evidence from these studies suggests that the presence or anticipation of acute stressors is associated with elevated situational cortisol levels, whereas chronic stress leads to dysregulations in basal, diurnal cortisol activity (Saxbe, 2008). Finding that levels of state and trait stress are conceptually different implies that state and trait stress may have differential effects on academic achievement. Differential effects may be found in the strength of the relation to achievement, and the pathways through which state and trait stress affect academic achievement (see LIKE framework, Figure 1).

11.1.6 Key Insight 6

State and trait stress differentially affect academic achievement as Study 3 found a substantial negative correlation between state stress and GPA, but no correlation between GPA and trait stress. Finding a negative association between state stress and academic achievement is in line with previous studies conducted inside and outside higher education (Cotton et al., 2002; Evans et al., 2013; Evans & Schamberg, 2009; Pluut et al., 2015; Pungello et al., 1996; Richardson et al., 2012; Schwartz et al., 2005). A possible explanation for the negative effects of state stress on academic achievement is that stress hinders learning, deeper conceptual understanding, and transfer if stress is experienced around the time of learning or right before memory recall, as shown in laboratory studies (Dandolo & Schwabe, 2016; Merz et al., 2016; Vogel & Schwabe, 2016). The fact that trait stress was unrelated to semester GPA in Study 3 concurs with findings from meta-analyses on the effects of a stress-related trait on learning and achievement, i.e., neuroticism. In these meta-analyses, neuroticism showed only small or no effects on academic performance in secondary and higher education (O'Connor & Paunonen, 2007; Poropat, 2009; Richardson et al., 2012). Even though individuals who score high on neuroticism may be more anxious, which can impair cognitive performance (Chamorro-Premuzic & Furhnam, 2006), moderator effects in a meta-analysis by Poropat (2009) showed that the association between neuroticism and performance is weaker in higher education than in lower levels of education. Possibly, the association is weaker because grades in higher education are less biased by relationship factors, or individuals with higher intelligence have better strategies to compensate for the potential hindering effects of neuroticism. Thus, learners in higher education may compensate for the potentially hindering effects of trait stress by resources such as prior knowledge, aptitudes, motivation, and self-regulation strategies.

11.2 Methodological Implications

In the current dissertation I studied how prior knowledge and stress are related to academic achievement in higher education. I presented three studies which use different methodological approaches to investigate the relations between prior knowledge and academic achievement, and stress and academic achievement. In the following, I would like to highlight the benefits, but also limitations and potential drawbacks of using these specific methods to study prior knowledge, stress and academic achievement in higher education.

11.2.1 Meta-analysis as a Research Tool to Study Prior Knowledge Effects on Academic Achievement

Meta-analysis is a useful method for educational research, and particularly, research on academic achievement, as it is a wide research field that has produced large numbers of empirical studies, both inside and outside higher education (Glass, 1976; Winne & Nesbit, 2010). Therefore, synthesizing findings from single studies is a promising strategy to integrate and explain inconsistent findings, provide more precise estimates of effect sizes, find gaps in the research literature, and thereby, fuel further development in theory and practice. Study 1 synthesized findings from 240 studies that investigated the impact of prior knowledge on learning outcomes, including more than 4000 effect sizes over different domains of knowledge, educational levels, types of knowledge, and further aspects. As a result, Study 1 provided a precise estimate of the effect of prior knowledge on later learning, i.e., the stability of interindividual differences in knowledge, and an estimate of the effect of prior knowledge on learning gains based on the small number of studies that provided the respective data.

A central determinant of meta-analysis is the literature search as it provides the data for the statistical analyses. Defining search strings, inclusion and exclusion criteria, and choosing the data bases in which the search will be conducted are crucial decisions in the initial stages of conducting a meta-analysis. When studying prior knowledge, one comes across a plethora of studies that is devoted to studying the effects of prior knowledge on learning outcomes, either explicitly or implicitly (Dochy et al., 2002). The quick increase of published empirical evidence is illustrated by the fact that the initial literature search in Study 1 yielded 4572 hits, and the updated search which was performed two years later, provided 890 additional hits. Even though so many articles were scanned for inclusion, it is still possible that the meta-analysis did not include all relevant studies as studying prior knowledge effects is ubiquitous in various fields of research (e.g., conceptual change, expertise, mathematics learning, and more). Possibly, searching other data bases (e.g., ERIC) and including unpublished research would also yield additional information of the association between prior knowledge and knowledge gains, and some moderator levels on which only limited information was available, e.g., knowledge domains, such as health sciences, medicine, history, and sports.

Integrating findings from single studies using different designs, measures, and samples in a meta-analysis is both a chance but also a potential drawback. When calculating the combined effect size in the meta-analysis, differences between the single studies that lead to differences in the effect size estimates are eliminated. This is a chance if the differences consist in irrelevant details of the study design or flaws in the measurement instruments that lead to measurement error, but it can be a drawback when the differences stem from investigating different effects in the population. As this is a potential drawback that stems from the very nature of meta-analysis – the synthesis of findings from primary studies that differ in various aspects – it can only be diminished. For example, researchers can include methodological moderators and estimate effect sizes for the different levels of these moderators. For example, in the meta-analysis conducted in Study 1, the effect of domain-specific prior knowledge on later learning outcomes was collapsed over a great variety of knowledge domains, educational interventions, levels, and age groups. Moderator analyses revealed that the correlation between prior knowledge and posttest knowledge did not differ in different domains of knowledge, study designs, learning outcomes, and countries. On the contrary, relatively strong differences were found for various educational levels, instruction with low versus high cognitive demands, and whether prior knowledge and posttest knowledge were from the same versus different domains of knowledge. These findings are crucial for making predictions for real-world situations as there is no cognitive or educational equivalent for prior knowledge averaged over several knowledge domains, educational levels, and instructional methods. Thus, carefully considering the meaning and validity of the effect size that is obtained from meta-analyses is indispensable when drawing conclusions from meta-analyses.

11.2.2 LTA as a Research Tool to Study Multivariate Trajectories of Knowledge

When conducting meta-analysis, one is typically interested in average effects, collapsing over several scales, interventions, and age groups. On the other hand, latent variable mixture modeling can be used to conduct fine-grained analysis of intra- and interindividual differences in learning. While traditional statistical approaches do not justice to the complex and often heterogeneous patterns of learning, latent variable mixture modeling has been shown to be a useful tool to study multivariate trajectories of knowledge (Hickendorff et al., 2017). Study 2 demonstrates that latent profile transition analyses can yield insights into profiles of conceptual knowledge and their development over the course of two academic years, providing evidence for conceptual change as a learning mechanism in higher education and the importance of a thorough understanding of academic concepts for achievement in higher education.

The profiles and trajectories of conceptual knowledge in Study 2 illustrate that the choice of knowledge tests and indicator variables is crucial when conducting latent class, latent profile models and their longitudinal extensions to study conceptual change processes. First of all, the nature of the indicator variables constrains the interpretation of knowledge profiles and learning pathways, i.e., whether it is a study of concepts of human memory or concepts of physical force, procedural knowledge or declarative knowledge, misconceptions, everyday concepts, prescientific conceptions, and so on. When studying conceptual change processes, it is necessary to choose indicator variables that represent different qualities and not only quantities of conceptual knowledge. For example in Study 2, transitions between profiles with a high mean of misconceptions and profiles with lower means of misconceptions, and higher means of correct concepts, provide compelling evidence for conceptual change mechanisms as these changes cannot be explained by learning mechanisms of the enrichment type only but rather reflect a reorganization of conceptual knowledge due to conceptual differentiation, coalescence, or changes in the ontological status of concepts, i.e., mechanisms of conceptual change (M. Schneider et al., 2012).

Second, psychometric properties of the knowledge tests play an important role for the results of latent class and latent transition analyses, in particular as knowledge tests are often constructed ad hoc and have unknown psychometric properties. Inappropriate item difficulties can lead to poor differentiation of knowledge profiles and poor classification of individuals into profiles, as they restrict item variance by floor and ceiling effects. Even though, the differentiation of knowledge profiles and the classification of individuals into latent classes was excellent, using a newly constructed test could have been a drawback for other reasons. As there is no information on retest-reliability, using the same test at all measurement occasions could have been a drawback, as participants may have remembered their answers and may have been reluctant to reconsider them which could lead to overestimating the stability of knowledge profiles. In the transition analyses this would result in fewer transitions between different knowledge profiles over time. When assessing misconceptions, a potential pitfall is that students may show higher levels of affirmations to misconceptions because of the employed measurements and not their faulty beliefs as the language and the structure of the test can lead to affirmation bias (Hughes et al., 2013). In the knowledge test used in Study 2, affirmation bias was not a potential pitfall, on the contrary, students may have been too reluctant to affirm the misconceptions items because indicating no group difference at all may have seemed implausible. This could have led to an underestimation of students with a pure misconceptions profile in Study 2.

Finally, a related methodological issue lies in the measurement model of latent class, latent profile, and latent transition models as it is assumed that the correlations between the indicator variables are fully explained by the latent class variable (Hickendorff et al., 2017; Nylund, 2007). This can be problematic as participants may have specific expectations or response tendencies when answering the items in a knowledge test. For example, students might expect that there must be at least one correct answer when presented with a list of incorrect statements and a dichotomous true/false answer format. In this case, the latent class variable would not account for all the relations between the indicator variables and a basic assumption of latent class analysis would be violated.

In sum, latent class and latent profile analysis, and their longitudinal extensions are useful tools to study knowledge and its trajectories, both inside and outside higher education. However, an informed application and careful consideration of the used knowledge measures is necessary to ensure meaningful, valid and generalizable model solutions.

11.2.3 LST Analysis as a Research Tool to Study Temporal Dynamics of Stress

Latent state-trait analyses has been used by prior research to study experienced emotions and physiological indicators of stress, and has been shown to be useful to differentiate and quantify state and trait effects, and study the temporal dynamics of states and traits. Study 3 was the first study that used a latent state-trait model to study stress in higher education students and the relationship between state and trait stress, academic achievement, and health. Using latent state-trait analyses to study stress has several methodological implications that are discussed in the following, including the choice of measurement instruments and the interpretation of the latent factors, in particular, the latent trait factor. Choosing appropriate measurement instruments is crucial because they determine the latent state and trait factors. By its very nature, latent state-trait analysis requires measurement instruments that are suited to assess situational and temporal fluctuations in psychological constructs. Therefore, questionnaires that are designed to estimate pure traits can be problematic, even though it can be interesting to investigate whether these instruments actually meet these requirements, e.g., social desirability (Schmitt & Steyer, 1993). In Study 3, it was expected that students' acute stress experience would impact their answers in the Perceived Stress Scale, although the PSS was originally conceptualized as a "global level of perceived stress" and measures the degree of stress over various life situations within the last month (S. Cohen et al., 1983, p. 387). Using the PSS to assess experiences of acute stress is supported by the large proportion of occasion specificity in Study 3b which indicated that most variance in students' stress levels was attributable to situational effects, not trait effects.

Second, whereas the interpretation of the latent state factors is relatively straightforward, as they represent a psychological state at a specific measurement occasion, the interpretation of the latent trait factor is more difficult. The latent trait factor represents individual differences in the psychological states that are common to all measurement occasions. Mathematically, the trait score of a person is a function of the person and situation variable, i.e., the trait is implicitly determined by the situational states as trait scores are estimated from persons in situations (Steyer et al., 2015). This can be problematic as the calculation of the trait relies on the assumption that all the situations in which the psychological states are assessed share the same underlying trait. Specifically, in the study of stress, where the distinction between state and trait stress is relatively new, the question remains what a common trait factor of stress represents. It could be possible that state stress in the first half of the semester is subject to general neuroticism, while the underlying trait of state stress in the second half of the semester is test anxiety. Thus, in order to understand latent traits in more depth, it is advisable to include covariates in the model, such as the Big Five personality factors in Study 3. Furthermore, it is advisable to consider and test other possible models, such as single trait and multiple trait models.

11.3 Practical Implications

The findings from the three studies of the current dissertation have several practical implications regarding student academic achievement in higher education. These can be broadly grouped into two categories: 1) monitoring the development of knowledge in order to foster academic achievement, and 2) decreasing stress as a risk factor for low achievement. In the following two subchapters, I present the practical implications relating to these two broad topics.

11.3.1 Monitoring the Development of Knowledge in order to Foster Academic Achievement

Processes of knowledge acquisition and development are closely related to academic achievement, as grades and standardized achievement are designed to reflect the degree to which students have accomplished the learning objectives of higher education programs (Steinmayr et al., 2014). As shown in Study 1 and 2, both univariate scores of correct prior knowledge and multivariate assessments of prior knowledge, including misconceptions and scientific concepts, and the learning trajectories between profiles of knowledge over time are important predictors of knowledge and academic achievement in higher education. This leads to the conclusion that, academic achievement among higher education students can be fostered by taking students' prior knowledge into account, and monitor the processes of knowledge development.

Monitoring the development of knowledge has several implications for teaching practices in higher education, including formative assessments, and constructivist approaches to teaching. Formal or informal formative assessments of knowledge contribute to student learning and achievement as they provide feedback about the current knowledge state to students and teachers (Nicol & Macfarlane-Dick, 2006). Therefore, formative assessment practices address several determinants of knowledge development illustrated in the LIKE Framework (Figure 1), e.g., learner's motivation, self-regulatory strategies, dialogue between students and teachers, and instructional strategies (Nicol & Macfarlane-Dick, 2006; Orlich, Harder, Callahan, Trevisan, & Brown, 2010; Yorke, 2003). Formative assessments inform students about learning goals and help them to regulate their learning activities in order to achieve these goals. Teachers on the other hand can regulate their teaching activities according to the current state of students' knowledge. Several formative assessment strategies are available that can be easily implemented in higher education, e.g., questioning, peer assessments, and providing feedback for tests, reports, and projects before assigning a final grade (Orlich et al., 2010). Formal assessment practices, such as student-peer and student self-assessments were among the strongest correlates of academic achievement in higher education in the review of meta-analyses by M. Schneider and Preckel (2017), suggesting that assessment practices are as important as presentation techniques in higher education. Furthermore, asking open ended questions can act as an informal formative assessment strategy as it can be used by teachers to uncover gaps in students' knowledge and misconceptions. Asking open-ended questions has been found to be strongly associated with student achievement in higher education (M. Schneider & Preckel, 2017).

Constructivist approaches to teaching can be used to foster students' transition from their preconceptions and everyday understanding to an academic understanding that is in line with state-of-the art research, i.e. conceptual change (Lawson, 2000). Conceptual change researchers widely agree that an important condition for conceptual change is cognitive conflict (M. Schneider et al., 2012). Cognitive conflict occurs when learners are confronted with information that does not fit their prior beliefs. Based on these findings, researchers have attempted to develop principles and teaching strategies to foster conceptual change in various disciplines inside higher education, for example in psychology (Hughes et al., 2013), biology (Alters & Nelson, 2002), and chemistry (N. T. Taylor, 2001). These strategies include refuting misconceptions with empirical evidence (Hughes et al., 2013), concept mapping and group discussions (N. T. Taylor, 2001), as well as prediction and simulation of data (Alters & Nelson, 2002). However, in particular in higher education, teaching for conceptual change is still in the early stages and more work is needed to develop and evaluate specific instructional methods in various disciplines.

Domain-specific knowledge is an informative indicator of academic achievement as it is closely tied to the learning objectives of sessions, courses, and programs. Moreover, it is more dynamic in nature than other indicators of academic achievement, such as grades and standardized achievement tests as they also comprise relatively stable factors, such as aptitude and general academic skills (Kuncel & Hezlett, 2007; Stiggins, 2001). Thus, uncovering students’ prior conceptions and monitoring students’ learning processes leading to knowledge acquisition processes are fruitful strategies to foster academic achievement. This can be achieved by using formative assessments and teaching approaches designed for conceptual change.

11.3.2 Decreasing Stress as a Risk Factor for Low Achievement in Higher Education

Excessive stress is a risk factor for psychological disorders, physical illness, and does also directly affect academic achievement (Cotton et al., 2002; Evans et al., 2013; Evans & Schamberg, 2009; Pluut et al., 2015; Pungello et al., 1996; Richardson et al., 2012; Schwartz et al., 2005). Study 3 has shown that state stress is associated with achievement loss in higher education, whereas state and trait stress go along with increased levels of psychological and psychosomatic symptoms. Therefore, decreasing stress promotes health, well-being, and study success. The results of Study 3 suggest that decreasing students’ stress levels can be achieved by either decreasing experienced stress in a specific situation, i.e., state stress, or decreasing trait stress as individuals who score high on trait stress are more likely to experience state stress, too. While decreasing high levels of trait stress should be a very effective strategy to decrease stress, theory-based interventions designed to change personality traits are just emerging as there is an ongoing debate in psychology about the changeability of traits (N. W. Hudson & Fraley, 2015; Magidson, Roberts, Collado-Rodriguez, & Lejuez, 2014; B. W. Roberts & DelVecchio, 2000).

Stress theories point to several possibilities for decreasing stress among higher education students. For example, one pathway might be to eliminate stressors that are associated with studying. However, this is not always possible, as students might experience stress in situations that are inherent to learning and achievement(Whitman, 1985), such as studying for examinations. These situations might cause stress because students are afraid of low achievements. Therefore, ensuring good teaching practices, for example, preparation and organization of the course, and clarity of course objectives and requirements, friendliness and respect for students (M. Schneider & Preckel, 2017), might not only ensure high achievements but also lower levels of stress among students. A second pathway to decreasing stress is changing the way students perceive potential stressors and stress. Perceiving potential stressors as manageable challenges instead of threats decreases experienced stress and bodily stress reactions (Seery, Weisbuch, Hetenyi, & Blascovich, 2010). Besides, stress mindsets play a crucial role for behavior in stressful situations (Crum et al., 2013). Students who believe that stressful situations enhance their performance, psychological and physical functioning, report fewer psychological symptoms, higher levels of life satisfaction, and are more likely to actively utilize their stress experiences in demanding situations to reach their goals (Crum et al., 2013).

Another possibility to decreasing stress is to enhance adaptive coping with stress as suggested by the results of Study 3a. Two theoretically important functions of coping are problem-solving and emotion regulation (Folkman & Lazarus, 1980; Folkman & Moskowitz, 2004). Accordingly, problem-focused and emotion-focused coping strategies can be distinguished. In correlational studies in the field of higher education, the use of coping strategies was associated with lower levels of stress (Saklofske et al., 2012) and higher levels of physical health, mental health, and well-being (Darling et al., 2007; Dyson & Renk, 2006; Eisenbarth et al., 2013; Pritchard et al., 2007). In a longitudinal study, coping strategies predicted better adjustment to college life, including less academic and general stress, three months later (Aspinwall & Taylor, 1992). However, studies showed that not all coping strategies are equally effective to reduce stress. A purely emotion-focused or avoidant coping strategy, wherein students only think about their stress or try to distract themselves, actually goes along with higher stress levels (Aspinwall & Taylor, 1992; Dyson & Renk, 2006; Eisenbarth et al., 2013; Liverant et al., 2004; Pritchard et al., 2007). In contrast, problem-focused coping or a combination of problem-focused and emotion-focused coping go along with lower stress levels (Sideridis, 2006). This is partly in line with findings from Study 3a, in which emotion-focused coping was negatively associated with state stress, while problem-focused coping was unrelated to stress. More recently, studies have shown that whether a coping strategy is adaptive or not depends mainly on the individual's appraisal of the situation, for example the controllability, which in turn impacts on the choice of coping strategies (Cheng, 2001; Cheng & Cheung, 2005). Based on these findings, Cheng and colleagues argue that stress management workshops should focus on conveying metacognitive skills that help students to differentiate between different stressful situations and deploy coping strategies that fit situational demands, instead of focusing only on broadening students' repertoire of coping strategies (Cheng & Cheung, 2005).

Several interventions that aim at reducing stress in higher education students have been evaluated meta-analytically (Regehr et al., 2013; Winzer, Lindberg, Guldbrandsson, & Sidorchuk, 2018; Yusufov, Nicoloro-SantaBarbara, Grey, Moyer, & Lobel, 2018) . These meta-analyses show that students who receive cognitive, behavioral, and mindfulness-based interventions experience lower levels of anxiety, depression, and cortisol than students in the control conditions (Regehr et al., 2013). The significant, but small effects from these types of interventions are sustained up to twelve months post intervention for negative mental health outcomes, and up to six months post intervention for positive mental health outcomes and academic achievement (Winzer et al., 2018). Moderator analyses suggest that the effectiveness of stress reduction interventions may differ for undergraduate and graduate students, and for reductions in stress and anxiety (Yusufov et al., 2018). Interventions that are designed according to cognitive-behavioral principles, interventions that address coping skills, and social support effectively reduce stress, whereas greater reduction in anxiety can be expected by relaxation training, mindfulness-based stress reduction, and psychoeducation. Moreover, the results suggest that graduate students profit to a greater extent than undergraduates from relaxation trainings and psychoeducational interventions.

In sum, higher education institutions should eliminate stressors where possible, for example, by ensuring good teaching practices and respectful interactions inside and outside classes. Results from experimental and correlative studies suggests that stress can be reduced by interventions that target students’ appraisals of stress and potentially stressful demands, and enhance students’ ability to adapt their coping strategies. Meta-analyses show that interventions designed according to cognitive-behavioral and, mindfulness principles, relaxation trainings, psychoeducation, and interventions that target students’ coping strategies and social support effectively reduce levels of stress and stress-related outcomes in higher education students with mostly sustainable, small effects.

11.4 Recommendations for Future Research

From the current dissertation, I would like to give three recommendations for future research: 1) investigate knowledge as a proxy for academic achievement, 2) focus on learning processes instead of on outcomes alone, and 3) use meta-analyses to synthesize findings. In the following three subsections, I explain these three recommendations in more detail.

11.4.1 Investigate Knowledge as a Proxy for Academic Achievement

Academic achievement in higher education refers to the attainment of the specific learning goals in higher education programs. The attainment of these goals can be operationalized in multiple ways, using standardized achievement tests, grades, degrees, and knowledge measures. Knowledge measures have rarely been used to study academic achievement because there is a lack of standardized instruments, so researchers often have to develop their own tests. Therefore, grades and standardized achievement tests have been used more frequently to study academic achievement as they can be attained more easily. However, as this dissertation shows, it is informative to use assessments of domain-specific knowledge to study academic achievement. First, achievement can be studied on a more detailed level as domain-specific knowledge is a much narrower and dynamic construct than academic achievement. For example, standardized achievement tests mostly represent a broader range of academic abilities that have been developed over several years of formal education (Educational Testing Service, 2017 51) and are strongly associated with general cognitive abilities (Kuncel et al., 2004). Grading practices have been criticized outside higher education for often represent multiple dimensions of learning, not only achievement of the learning objectives, but also effort and aptitude (Stiggins, 2001).

Investigating knowledge as a proxy for academic achievement holds several challenges as standardized instruments are scarce, and the assessment of knowledge costs more time and effort than the assessment of other indicators, for example grades and GPA. However, investigating domain-specific knowledge and its development in higher education provides the opportunity to gain a better understanding of the learning processes involved in academic achievement. This can help to further develop instructional strategies to support these learning processes.

11.4.2 Focus on Learning Processes Instead of on Outcomes Alone

Prior meta-analyses on academic achievement and their syntheses have identified correlates of student achievement in various levels of education, including higher education (e.g., Hattie, 2009; M. Schneider & Preckel, 2017). These studies have highlighted the impact of prior achievement, knowledge, intelligence, and other abilities for later achievement. Prior knowledge is a strong predictor of later knowledge (Dochy et al., 2002) and performance, both inside and outside higher education. Study 1 of the current dissertation shows that the effect of prior knowledge on learning is likely to be causal. Yet, there is still a lack of longitudinal studies on the development of knowledge, and/or studies that report the association between prior knowledge and knowledge gains, in particular inside higher education. Studying the bivariate relation between prior knowledge and posttest knowledge allows only statements about the strength of the association between knowledge at pretest and knowledge at posttest, i.e., the stability of interindividual differences in knowledge. When interested in making statements about learning, information about learners' gains in knowledge over time are needed. Unfortunately, this information was not provided by the majority of the primary studies included in the meta-analysis of Study 1, showing that there is a need to provide more information about learning gains in future studies. Investigating how knowledge develops *within* learners can help to understand the difficulties that students may encounter in the acquisition of academic knowledge. Some of these difficulties may be due to learners' preconceptions as prior research has shown that even inside higher education, many students hold pervasive misconceptions about academic topics (e.g., Clement, 1982; Hughes et al., 2013; Shtulman & Calabi, 2013; Xu et al., 2013). Multivariate assessments of knowledge, including students' preconceptions, as well as academic correct conceptions in a domain, can help to uncover the processes underlying the dynamics of knowledge development, e.g., knowledge fragmentation and knowledge integration. Study 2 of the current dissertation showed that latent profile transition analysis is a useful statistical tool to model these knowledge trajectories.

Studying academic achievement by investigating the development of academic knowledge affords the opportunity to uncover the intra-learner processes and learner-environment interaction processes that underlie academic achievement. LIKE (Figure 1) provides a framework to guide future research to study knowledge acquisition, and thus, academic achievement by means of a mediated moderation. Based on the LIKE framework and the results of Study 1 and 2 of the current dissertation, future research should put greater emphasis on studying intra-learner processes and their determinants. In quasi-experimental and experimental studies on prior knowledge this can be implemented by including information about knowledge gains, and/ or using longitudinal and multivariate assessments of knowledge.

11.4.3 Use Meta-Analyses to Synthesize Findings

Academic achievement is an important outcome of educational research, which has led to an enormous amount of literature (Winne & Nesbit, 2010). In particular in higher education research, it is one of the most frequently studied outcomes (Astin, 1993, as cited by Strayhorn, 2013). This is also illustrated by a growing amount of meta-analyses in the field. A recent review identified 124 meta-analyses of academic achievement in higher education of which 38 were eligible for the analysis, including 3330 effect sizes obtained from studies with almost 2 million students (M. Schneider & Preckel, 2017). Similarly, the literature search of Study 1 reported in the current dissertation, yielded an enormous data base of 4572 initial hits, of which 240 studies were eligible for the meta-analysis, providing 4327 effect sizes for the relationship between prior knowledge and posttest knowledge and achievement over all educational levels, and 358 effect sizes from 57 studies inside higher education. These numbers illustrate the breadth of the research field on academic achievement, which is both a chance and a potential drawback. Having a lot of evidence available on academic achievement can be a drawback as there likely is a large variation of inconsistent findings in various educational contexts, for various outcomes, and in various populations, making it hard to draw conclusions for educational practice (Glass, 1976).

Thus, meta-analyses are needed in order to synthesize the evidence gathered on the many questions that pertain to academic achievement, such as the effectivity of specific instructional methods, characteristics of learners and teachers, and more. Particularly, in the light of a process perspective on academic achievement (see Chapter 11.4.2), meta-analytic structural equation model (MASEM) seems to be a promising strategy for future research. MASEM can be used to test hypotheses about learning processes, their determinants, and outcomes on a meta-analytical level (Cheung, 2015). MASEM allows to address research questions that cannot be addressed by structural equation models using primary data. For example, MASEM can help to explain different relationships between prior knowledge, learning processes, teaching strategies and achievement in different knowledge domains or educational levels. Furthermore, MASEM can be used to reexamine the factor structure of relevant determinants of academic achievement meta-analytically, for example, experienced state and trait stress which are differentially related to academic achievement as shown in Study 3. Strong evidence is provided if structural models that fit the data well in single studies, such as latent state-trait models, still show good model fits in a meta-analytical model. Whereas this is an example of a quasi-experimental or correlative application, MASEM can also be used to synthesize findings from experimental studies. When combining MASEM with data from randomized-controlled trials, particularly compelling evidence for how instruction interacts with intra-learner processes in the development of academic achievement can be provided.

Meta-analyses have higher statistical power and allow for more moderator analyses than primary studies, therefore conducting meta-analyses is a promising strategy to synthesize findings from primary studies in educational research and address research questions that will lead to a better understanding of the learning and teaching processes that lead to academic achievement. Yet, meta-analysis is not perfect either and relies on primary studies of high methodological quality in order to address specific research questions. Therefore, the most fruitful strategy for future research seems to be two-fold: First, conduct primary studies of high validity for the understanding of learning processes and academic achievement, i.e., use multivariate indicators of knowledge, report knowledge gains, use experimental designs, and consider factors that mediate and moderate these learning processes. Second, address the questions and problems that cannot be resolved by primary studies by carefully apply meta-analytical techniques, for example MASEM when investigating learning processes that lead to academic achievement.

12. Conclusion

Academic achievement has been a vital aspect of educational research as educating high achieving individuals has immediate and long-term benefits for both, the individual and the society (OECD, 2000). Cultivating academic achievement among learners is not only a task undertaken by teachers, but a complex interaction of multiple factors that pertain to teaching, the learning environment, learning resources, and the cognitive and non-cognitive stable and dynamic characteristics of learners(Hattie, 2009). Prior knowledge and stress are two of these dynamic learner characteristics which have been shown to have powerful impacts on academic achievement (Committee on Developments in the Science of Learning, 2004; Dandolo & Schwabe, 2016; Vogel, Kluen, Fernández, & Schwabe, 2018). They both affect the learning processes that lead to academic achievement directly as they affect cognitive processes involved in learning and indirectly as for example, stress can alter the ways students interact with their learning environments. However, in particular inside higher education, there was a gap in our knowledge concerning the specific contributions of prior knowledge to learning outcomes across different domains, the nature of students’ prior knowledge and the learning trajectories that lead to academic achievement. Furthermore, the contributions of state and trait stress to academic achievement were unknown. Specifically, there was a lack of meta-analyses on the relation between prior knowledge and learning outcomes over a wide range of knowledge domains and educational levels, and longitudinal investigations of discontinuous learning trajectories, experienced stress and academic achievement in higher education.

In order to close these gaps, three studies were conducted as part of the current dissertation which have led to theoretical and practical conclusions about the contributions of prior knowledge and stress to academic achievement. On a theoretical level, the results of the current dissertation support the idea that the acquisition of domain-specific knowledge is a cumulative and long-term process. Studying these learning processes is a promising strategy to better understand how academic achievement evolves among learners in and outside higher education as knowledge is a much narrower and more dynamic construct than achievement. Moreover, studying whether and under which circumstances students' restructure their incompatible prior knowledge, i.e., processes of conceptual change, is crucial as learners often hold misconceptions, synthesize scientific facts with incorrect pieces of information, and show fragmented structures of knowledge that can hamper a thorough understanding and academic achievement in a domain.

On a practical level, the current dissertation has two principal implications. First, domain-specific knowledge is an important prerequisite for academic achievement in and outside higher education. Implementing formative assessments of students' conceptual knowledge, and teaching strategies that foster conceptual change can help monitoring the development of academic knowledge that is needed to support the learning processes that lead to academic achievement. Second, stress is a risk factor for low academic achievement and poor mental health in higher education students. Teaching practices that foster academic achievement may also prevent state stress among higher education students, such as communicating clear learning goals and course requirements. Stress that students experience due to the demands placed upon them either as state or trait stress can be particularly harmful for students' mental health. Interventions that target student stress and anxiety, promote well-being and academic performance could be offered by higher education institutions as a means of helping students with high levels of stress who are unable to cope on their own.

Despite numerous investigations in the field have provided many insights into the contributions of cognitive and non-cognitive characteristics of learners, teaching strategies, and learning environment to academic achievement, we are still at the early stages of understanding the causal network of prior knowledge, stress, and academic achievement in higher education. The current dissertation is a starting point from which the contributions of knowledge and stress to achievement in higher education can be explored in more depth. For this purpose, the LIKE framework which is presented as part of the current dissertation can be a helpful heuristic to test hypotheses about how prior knowledge and stress affect learning processes that lead to academic achievement. Promising strategies for testing these hypotheses include conducting longitudinal studies, using randomized controlled trials and meta-analytical techniques, such as MASEM, to synthesize findings from primary studies. Promoting our understanding of academic achievement, can help to ensure a high quality of tertiary education and foster numerous benefits for individuals and societies.

References

Abouserie, R. (1994). Sources and levels of stress in relation to locus of control and self esteem in university students. *Educational Psychology, 14*(3), 323-330.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 78*(5), 218-223.

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders, 36*(4), 312-333. doi:10.1097/TLD.0000000000000102

Alters, B. J., & Nelson, C. E. (2002). Perspective: Teaching evolution in higher education. *Evolution, 56*(10), 1891-1901.

Amadieu, F., Salmerón, L., Cegarra, J., Paubel, P.-V., Lemarié, J., & Chevalier, A. (2015). Learning from concept mapping and hypertext: An eye tracking study. *Educational Technology & Society, 18*(4), 100-112.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American psychologist, 57*(12), 1060-1073. doi:10.1037/0003-066X.57.12.1060

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglas, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060. doi:10.1037/0033-295X.111.4.1036

Arena, J. G., Sherman, R. A., Bruno, G. M., & Smith, J. D. (1990). The relationship between situational stress and phantom limb pain: cross-lagged correlational data from six month pain logs. *Journal of Psychosomatic Research, 34*(1), 71-77.

Arkes, H. R., & Freedman, M. R. (1994). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition, 12*(1), 84-89. doi:10.3758/BF03197000

Asparouhov, T., & Muthén, B. O. (2010). *Computing the strictly positive Satorra-Bentler chi-square test in Mplus* Vol. 12. *Mplus Web Notes*

Aspinwall, L. G., & Taylor, S. E. (1992). Modeling cognitive adaptation: A longitudinal investigation of the impact of individual differences and coping on college adjustment and performance. *Journal of personality and social psychology, 63*(6), 989-1003.

Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests.

Educational Researcher, 38(9), 665-676.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York and Toronto: Holt, Rinehart, and Winston.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.

Badham, S. P., Hay, M., Foxon, N., Kaur, K., & Maylor, E. A. (2016). When does prior knowledge disproportionately benefit older adults' memory? *Aging, Neuropsychology, and Cognition, 23*(3), 338-365. doi:10.1080/13825585.2015.1099607

Bahrick, H. P., & Hall, L. K. (1991). Lifetime maintenance of high school mathematics content. *Journal of experimental psychology: general, 120*(1), 20.

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (in press). Risky business: Correlation and causation in longitudinal studies of skill development. *American psychologist*.

Bakker, A., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of managerial psychology, 22*(3), 309-328.

Bakker, A., Demerouti, E., & Schaufeli, W. (2003). Dual processes at work in a call centre: An application of the job demands–resources model. *European Journal of Work and Organizational Psychology, 12*(4), 393-417.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological bulletin, 128*(4), 612-637. doi:10.1037//0033-2909.128.4.612

Bates, A. K., & Bourke, P. Q. (2016). *Findings from the 2016 Administration of the Your First College Year (YFCY) Survey*. Los Angeles, CA.

Bates, A. W., & Poole, G. (2003). *Effective Teaching with Technology in Higher Education: Foundations for Success*: ERIC.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers’ mathematical knowledge, cognitive activation in the classroom, and student progress. *American educational research journal, 47*(1), 133-180. doi:10.3102/0002831209345157

Baumert, J., Nagy, G., & Lehmann, R. (2012). Cumulative advantages and the emergence of social and ethnic inequality: Matthew effects in reading and mathematics development within elementary schools? *Child Development, 83*, 1347-1367. doi:10.1111/j.1467-8624.2012.01779.x

Bech, P. (2004). Measuring the dimension of psychological general well-being by the WHO-5. *Quality of Life Newsletter*, 15-16.

Beeth, M., Duit, R., Prenzel, M., Ostermeier, C., Tytler, R., & Wickman, P.-O. (2003). Quality development projects in science education. In D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos, & M. Kallery (Eds.), *Science education research in the knowledge-based society* (pp. 447-457). Dordrecht: Springer Netherlands.

Bewick, B. M., Gill, J., Mulhearn, B., Barkham, M., & Hill, A. J. (2008). Using electronic surveying to assess psychological distress within the UK student population: a multi-site pilot investigation. *E-Journal of Applied Psychology, 4*(2).

Biggs, J. B. (1993). From theory to practice: A cognitive systems approach. *Higher education research and development, 12*(1), 73-85.

Bilalić, M., McLeod, P., & Gobet, F. (2008a). Inflexibility of experts - reality or myth? Quantifying the Einstellung effect in chess masters. *Cognitive psychology, 56*, 73-102.

Bilalić, M., McLeod, P., & Gobet, F. (2008b). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition, 108*(3), 652-661.

Bilalić, M., McLeod, P., & Gobet, F. (2010). The mechanism of the Einstellung (set) effect: A pervasive source of cognitive bias. *Current Directions in Psychological Science, 19*(2), 111-115. doi:0.1177/0963721410363571

Billings, A. G., & Moos, R. H. (1981). The role of coping responses and social resources in attenuating the stress of life events. *Journal of Behavioral Medicine, 4*(2), 139-157. doi:10.1007/bf00844267

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of research in science teaching, 27*(5), 415-427.

Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 59-68). New York: Worth.

Bjork, R. A. (1992). Interference and memory. In L. R. Squire (Ed.), *Encyclopedia of learning and memory* (pp. 283-288). New York: Macmillan.

Bligh, D. (2002). *What's the use of lectures*. Exeter, Egnland: Intellect.

Bolger, N., & Zuckerman, A. (1995). A framework for studying personality in the stress process. *Journal of personality and social psychology, 69*(5), 890.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods, 1*(2), 97-111.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothenstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons Ltd.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences, 113*(47), 13354-13359.

Bovier, P. A., Chamot, E., & Perneger, T. V. (2004). Perceived stress, internal resources, and social support as determinants of mental health among young adults. *Quality of Life Research, 13*(1), 161-170.

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159.

Breakspear, S. (2012). The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers*(71), 0_1.

Bretz Jr., R. D. (1989). College Grade Point Average as a Predictor of Adult Success: A Meta-Analytic Review and Some Additional Evidence *Public Personnel Management, 18*(1), 11-22. doi: 10.1177/009102608901800102

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician, 48*(3), 209-213.

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

Brooks, L. R. (1967). The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology, 19*(4), 289-299. doi:10.1080/14640746708400105

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. *Science Education, 68*(4), 493-503.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1992). The origin and evolution of everyday concepts. In R. N. Giere (Ed.), *Cognitive models of science* (Vol. 15). Minneapolis: University of Minnesota Press.

Carey, S., & Spelke, E. S. (1994). Domain-specific knowledge and conceptual change. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169-200). Cambridge, UK: Cambridge University Press.

Carmody, J., & Baer, R. A. (2008). Relationships between mindfulness practice and levels of mindfulness, medical and psychological symptoms and well-being in a mindfulness-based stress reduction program. *Journal of Behavioral Medicine, 31*(1), 23-33.

Case, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale: Erlbaum.

Castel, A. D., McCabe, D. P., Roedinger, H. L., & Heitman, J. L. (2007). The dark side of expertise. *Psychological Science, 18*(1), 3-5. doi:10.1111/j.1467-9280.2007.01838.x

Chamorro-Premuzic, T., & Furhnam, A. (2006). Intellectual competence and the intelligent personality: A third way in differential psychology. *Review of General Psychology, 10*(3), 251.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess *Visual information processing* (pp. 215-281): Elsevier.

Chase, W. G., & Simon, H. A. (1973b). Perception in chess. *Cognitive psychology, 4*(1), 55-81. doi:10.1016/0010-0285(73)90004-2

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: a multimethod approach. *Journal of personality and social psychology, 80*(5), 814.

Cheng, C., & Cheung, M. W. (2005). Cognitive processes underlying coping flexibility: Differentiation and integration. *Journal of personality, 73*(4), 859-886.

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods, 19*(2), 211.

Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*: John Wiley & Sons.

Chi, M. T. H., & Ceci, S. J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 91-142). New York, NY: Academic Press.

Clark, S. L., & Muthén, B. O. (2009). Relating latent class analysis results to variables not included in the analysis.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of physics, 50*(1), 66-71.

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*, 81-89.

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

Cohen, P. A. (1984). College grades and adult achievement: A research synthesis. *Research in higher education, 20*(3), 281-293. doi:10.1007/BF00983503

Cohen, S. (1988). Perceived stress in a probability sample of the United States. In S. S. S.

Oskamp (Ed.), *The social psychology of health* (pp. 31-67). Thousand Oaks, CA, US: Sage Publications, Inc.

Cohen, S., & Janicki-Deverts, D. (2012). Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. *Journal of applied social psychology, 42*(6), 1320-1334.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of health and social behavior*, 385-396.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*.

Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological bulletin, 98*(2), 310-357.

Collins, L. M., & Lanza, S. T. (2010). Latent class analysis with covariates. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*, 149-177.

Committee on Developments in the Science of Learning. (2004). *How people learn: Brain, mind, experience, and school: Expanded edition* (J. D. Bransford, A. L. Brown, & R. R. Cocking Eds.). Washington, DC: National Academies Press.

Compas, B. E., Malcarne, V. L., & Fondacaro, K. M. (1988). Coping with stressful events in older children and young adolescents. *Journal of Consulting and Clinical Psychology, 56*(3), 405-411. doi:10.1037/0022-006X.56.3.405

Conway, M. A., Cohen, G., & Stanhope, N. (1992). Very long-term memory for knowledge acquired at school and university. *Applied cognitive psychology, 6*(6), 467-482.

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*: Russell Sage Foundation.

Cotton, S. J., Dollard, M. F., & De Jonge, J. (2002). Stress and student job design: Satisfaction, well-being, and performance in university students. *International Journal of Stress Management, 9*(3), 147-162.

Crum, A. J., Salovey, P., & Achor, S. (2013). Rethinking stress: the role of mindsets in determining the stress response. *Journal of personality and social psychology, 104*(4), 716.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis* New York, NY: Routledge.

Custers, E. J. (2010). Long-term retention of basic science knowledge: a review study. *Advances in Health Sciences Education, 15*(1), 109-128.

Dagher, Z. R., & BouJaoude, S. (1997). Scientific views and religious beliefs of college students: The case of biological evolution. *Journal of research in science teaching, 34*(5), 429-445.

Dahlström, Ö. (2012). Learning during a collaborative final exam. *Educational Research and Evaluation, 18*(4), 321-332.

Dandolo, L. C., & Schwabe, L. (2016). Stress-induced cortisol hampers memory generalization. *Learning & Memory, 23*(12), 679-683.

Darling, C. A., McWey, L. M., Howard, S. N., & Olmstead, S. B. (2007). College student stress: The influence of interpersonal relationships on sense of coherence. *Stress and Health, 23*(4), 215-229.

De Fruyt, F., Van Leeuwen, K., Bagby, R. M., Rolland, J.-P., & Rouillon, F. (2006). Assessing and interpreting personality change and continuity in patients treated for major depression. *Psychological assessment, 18*(1), 71-80.

De Kloet, E. R., Joëls, M., & Holsboer, F. (2005). Stress and the brain: from adaptation to disease. *Nature Reviews Neuroscience, 6*(6), 463-475.

Deasy, C., Coughlan, B., Pironom, J., Jourdan, D., & Mannix-McNamara, P. (2014). Psychological Distress and Coping amongst Higher Education Students: A Mixed Method Enquiry. *PLoS ONE, 9*(12), e115193. doi:10.1371/journal.pone.0115193

Derogatis, L. R., & Unger, R. (2010). *Symptom checklist‐90‐revised*: Wiley Online Library.

Detterman, D. K., & Sternberg, R. J. (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Westport, CT: Ablex.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin, 130*(3), 355.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. *Annual review of psychology, 54*(1), 403-425.

diSessa, A. A. (2008). *A bird's-eye view of the" pieces" vs" coherence" controversy (from the" pieces" side of the fence").* Paper presented at the International handbook of research on conceptual change.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*(6), 843-900.

doi:10.1016/j.cogsci.2004.05.003

Doane, L. D., Chen, F. R., Sladek, M. R., Van Lenten, S. A., & Granger, D. A. (2015). Latent trait cortisol (LTC) levels: Reliability, validity, and stability. *Psychoneuroendocrinology, 55*, 21-35.

Dochy, F. (1990). Instructional Implications of Recent Research and Empirically-Based Theories on the Effect of Prior Knowledge on Learning *Learning Environments* (pp. 339-355): Springer.

Dochy, F., & Alexander, P. A. (1995). Mapping prior knowledge: A framework for discussion among researchers. *European Journal of Psychology of Education, 10*(3), 225-242. doi:10.1007/BF03172918

Dochy, F., De Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning: How far have we progressed since Bloom? Implications for educational practice and teaching. *Active learning in higher education, 3*(3), 265-284.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of educational research, 69*(2), 145-186.

Dormann, C., Fay, D., Zapf, D., & Frese, M. (2006). A state-trait analysis of job satisfaction: on the effect of core self-evaluations. *Applied psychology, 55*(1), 27-51.

Duff, D., Tomblin, J. B., & Catts, H. (2015). The influence of reading on vocabulary growth: A case for a Matthew effect. *Journal of Speech, Language, and Hearing Research, 58*, 853-864. doi:10.1044/2015_JSLHR-L-13-0310

Duit, R., Treagust, D. F., & Widodo, A. (2008). Teaching science for conceptual change: Theory and practice. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 629-646). New York: Routledge.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89-98. doi:10.2307/2669529

Dyson, R., & Renk, K. (2006). Freshmen adaptation to university life: Depressive symptoms, stress, and coping. *Journal of clinical psychology, 62*(10), 1231-1244.

Ebrecht, M., Hextall, J., Kirtley, L.-G., Taylor, A., Dyson, M., & Weinman, J. (2004). Perceived stress and cortisol levels predict speed of wound healing in healthy male adults. *Psychoneuroendocrinology, 29*(6), 798-809.

Edelsbrunner, P. A., Schalk, L., Schumacher, R., & Stern, E. (2015). *Pathways of Conceptual Change: Investigating the Influence of Experimentation Skills on Conceptual Knowledge Development in Early Science Education.* Paper presented at the CogSci.

Edelsbrunner, P. A., Schalk, L., Schumacher, R., & Stern, E. (2018). Variable control and conceptual change: A large-scale quantitative study in elementary school. *Learning and Individual Differences*.

Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society, 51*(3), 599-635.

Educational Testing Service. (2017). *GRE Information Bulletin.* Retrieved from https://www.ets.org/gre/revised_general/about/?WT.ac=grehome_greabout_a_150213

Edwards, K. J., Hershberger, P. J., Russell, R. K., & Markert, R. J. (2001). Stress, negative social exchange, and health symptoms in university students. *Journal of American College Health*, *50*(2), 75-79.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634. doi:10.1136/bmj.315.7109.629

Eisenbarth, C. A., Champeau, D. A., & Donatelle, R. J. (2013). Relationship of appraised stress, coping strategies, and negative affect among college students. *International Journal of Psychology and Behavioral Sciences, 3*(5), 131-138.

Engling, J. (2010). *German PSS (4 item)*. [Word document]. Retrieved from PSS Translations Website: http://www.psy.cmu.edu/~scohen/scales.html

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist, 49*(8), 725-747.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181-1182.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406. doi:10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual review of psychology, 47*, 273-305. doi:10.1146/annurev.psych.47.1.273

European Commission. (2008). *The European Qualifications Framework for Lifelong Learning (EQF)*. Luxembourg: Office for Official Publications of the European

Communities.

EuropeanCommission. (2002). *Guidance on work-related stress: Spice of life or kiss of death*. Luxembourg: European Communities.

Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological bulletin, 139*(6), 1342.

Evans, G. W., & Schamberg, M. A. (2009). Childhood poverty, chronic stress, and adult working memory. *Proceedings of the National Academy of Sciences, 106*(16), 6545-6549.

Ferguson, C. J., & Brannick, M. T. (2011). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17*(1), 120-128. doi:10.1037/a0024445

Fischer, F. T., Schult, J., & Hell, B. (2013). Sex-specific differential prediction of college admission tests: A meta-analysis. *Journal of Educational psychology, 105*(2), 478.

Fisher, M., & Keil, F. C. (2016). The curse of expertise: When more knowledge leads to miscalibrated explanatory insight. *Cognitive Science, 40*(5), 1251-1269.

Fisher, Z., & Tipton, E. (2014). Robumeta: Robust variance meta-regression. Retrieved from http://cran.rproject.org/web/packages/robumeta/index.html

Flaig, M., Simonsmeier, B. A., Mayer, A.-K., Rosman, T., Gorges, J., & Schneider, M. (2018). Conceptual change and knowledge integration as learning processes in higher education: A latent transition analysis. *Learning and Individual Differences, 62*, 49-61. doi:10.1016/j.lindif.2017.12.008

Folkman, S. (2011). *The Oxford handbook of stress, health, and coping*. New York: Oxford University Press.

Folkman, S., & Moskowitz, J. T. (2004). Coping: Pitfalls and promise. *Annual review of psychology, 55*, 745-774.

Forgays, D. G., Forgays, D. K., & Spielberger, C. D. (1997). Factor structure of the State-Trait Anger Expression Inventory. *Journal of Personality Assessment, 69*(3), 497-507. doi:10.1207/s15327752jpa6903_5

Franke, G. H., & Derogatis, L. R. (2002). *Symptom-Checkliste von LR Derogatis: SCL-90-R; deutsche Version*: Beltz Test.

Fry, H., Ketteridge, S., & Marshall, S. (2008). *A handbook for teaching and learning in higher education: Enhancing academic practice*. New York, NY: Routledge.

Furnée, C. A., Groot, W., & van Den Brink, H. M. (2008). The health effects of education: a meta-analysis. *European journal of public health, 18*(4), 417-421.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational psychology, 108*(1), 82-97. doi:10.1037/edu0000053

Gadzella, B. M. (1994). Student-life stress inventory: Identification of and reactions to stressors. *Psychological reports, 74*(2), 395-402.

Garner, R., Alexander, P. A., Gillingham, M. G., Kulikowich, J. M., & Brown, R. (1991). Interest and learning from text. *American educational research journal, 28*(3), 643-659.

Garry, M., Loftus, E. F., & Brown, S. W. (1994). Memory: A river runs through it. *Consciousness and Cognition, 3*(3), 438-451.

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. *Journal of Educational psychology, 78*(3), 225.

Geiser, C., Keller, B. T., Lockhart, G., Eid, M., Cole, D. A., & Koch, T. (2015). Distinguishing state variability from trait change in longitudinal data: The role of measurement (non) invariance in latent state-trait analyses. *Behavior research methods, 47*(1), 172-203.

Gentner, D., Brem, S., Ferguson, R. W., Markman, A. B., Levidow, B. B., Wolff, P., & Forbus, K. D. (1997). Analogical reasoning and conceptual change: A case study of Johannes Kepler. *Journal of the Learning Sciences, 6*(1), 3-40.

Giesbrecht, G. F., Bryce, C. I., Letourneau, N., Granger, D. A., & Team, A. s. (2015). Latent trait cortisol (LTC) during pregnancy: Composition, continuity, change, and concomitants. *Psychoneuroendocrinology, 62*, 149-158.

Gilboa, S., Shirom, A., Fried, Y., & Cooper, C. (2008). A meta-analysis of work demand stressors and job performance: examining main and moderating effects. *Personnel Psychology, 61*(2), 227-271.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Gobet, F. (2005). Chunking models of expertise: Implications for instruction. *Applied cognitive psychology, 19*, 183-204.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in cognitive sciences, 5*, 236-243.

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*(1), 20-33.

doi:10.3758/BF03211367

Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. In I. B. Weiner (Ed.), *Handbook of Psychology* (Vol. Eight:22, pp. 597–621).

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological bulletin, 142*(7), 729-757. doi:10.1037/bul0000043

Gómez, E. J., Benarroch, A., & Marín, N. (2006). Evaluation of the degree of coherence found in students' conceptions concerning the particulate nature of matter. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 43*(6), 577-598.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms and the theory theory. *Psychological bulletin, 138*(6), 1085-1108. doi:10.1037/a0028044

Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience, 59*(9), 792-799.

Grek, S. (2009). Governing by numbers: The PISA ‘effect’in Europe. *Journal of education policy, 24*(1), 23-37.

Griggs, R. A., & Jackson, S. L. (1988). A reexamination of the relationship of high school psychology and natural science courses to performance in a college introductory psychology class. *Teaching of Psychology, 15*(3), 142-144.

Grillon, C., Ameli, R., Foot, M., & Davis, M. (1993). Fear-potentiated startle: relationship to the level of state/trait anxiety in healthy subjects. *Biological psychiatry, 33*(8-9), 566-574.

Groot, W., & Van Den Brink, H. M. (2007). The health effects of education. *Economics of Education Review, 26*(2), 186-200.

Hailikari, T., Katajavuori, N., & Lindblom-Ylanne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education, 72*(5), 113.

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: A structural model. *Educational Psychology, 28*(1), 59-71.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of physics, 66*(1), 64-74. doi:10.1119/1.18809

Halamandaris, K., & Power, K. (1999). Individual differences, social support and coping with the examination stress: A study of the psychosocial and academic adjustment of first year home students. *Personality and Individual Differences, 26*(4), 665-685.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive psychology, 44*(4), 339-387.

Hammen, C., Kim, E. Y., Eberhart, N. K., & Brennan, P. A. (2009). Chronic and acute stress and the prediction of major depression in women. *Depression and anxiety, 26*(8), 718-723.

Hammer, M. (2006). *SBT, Stressbewältigungstraining für psychisch kranke Menschen: ein Handbuch zur Moderation von Gruppen*: Psychiatrie-Verlag.

Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational psychology, 94*(3), 562-575.

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational psychology, 98*(2), 307-326.

Harrison, A. G., Grayson, D. J., & Treagust, D. F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 36*(1), 55-87.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50*(1-3), 171-188.

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

Hecht, H., & Proffitt, D. R. (1995). The price of expertise: Effects of experience on the water-level task. *Psychological Science, 6*(2), 90-95. doi:10.1111/j.1467-9280.1995.tb00312.x

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods, 1*(1), 39-65.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology, 34*(2), 163-171.

Hellhammer, J., Fries, E., Schweisthal, O., Schlotz, W., Stone, A., & Hagemann, D. (2007). Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state-and trait components. *Psychoneuroendocrinology, 32*(1), 80-86.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*(4), 389–395. doi:10.3758/BF03202431

Hennessy, D. A., & Wiesenthal, D. L. (1997). The relationship between traffic congestion, driver stress and direct versus indirect coping behaviours. *Ergonomics, 40*(3), 348-361.

Hennessy, D. A., & Wiesenthal, D. L. (1999). Traffic congestion, driver stress, and driver aggression. *Aggressive behavior, 25*(6), 409-423.

Hennessy, D. A., Wiesenthal, D. L., & Kohn, P. M. (2000). The influence of traffic congestion, daily hassles, and trait stress susceptibility on state driver stress: An interactive perspective1. *Journal of Applied Biobehavioral Research, 5*(2), 162-179.

Hewitt, P. L., Flett, G. L., & Mosher, S. W. (1992). The Perceived Stress Scale: Factor structure and relation to depression symptoms in a psychiatric sample. *Journal of Psychopathology and Behavioral Assessment, 14*(3), 247-257.

Hewson, P. W., Tabachnick, B. R., Zeichner, K. M., & Lemberger, J. (1999). Educating prospective teachers of biology: Findings, limitations, and recommendations. *Science Education, 83*(3), 373-384. doi:10.1002/(SICI)1098-237X(199905)83:3<373::AID-SCE6>3.0.CO;2-3

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2017). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences*.

Hiebert, J., Stigler, J. W., & Manaster, A. B. (1999). Mathematical features of lessons in the TIMSS video study. *ZDM – International Journal on Mathematics Education, 31*(6), 196-201. doi:10.1007/BF02652695

Hieggelke, C., Maloney, D., Van Heuvelen, A., & O'Kuma, T. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of physics, 69*(7), S12-S23.

Hirschfeld, L. A., & Gelman, S. A. (Eds.). (1994). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.

Hobfoll, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American psychologist, 44*(3-13), 513-524.

Holahan, C. J. (1987). Risk, resistance, and psychological distress: a longitudinal analysis with adults and children. *Journal of Abnormal Psychology, 96*(1), 3.

Holahan, C. J., & Moos, R. H. (1990). Life stressors, resistance factors, and improved psychological functioning: an extension of the stress resistance paradigm. *Journal of personality and social psychology, 58*(5), 909-917.

Holahan, C. J., Moos, R. H., Holahan, C. K., & Cronkite, R. C. (1999). Resource loss, resource gain, and depressive symptoms: a 10-year model. *Journal of personality and social psychology, 77*(3), 620-629.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635-650. doi:10.1086/266350

Hudson, H., & Rottmann, R. M. (1981). Correlation between performance in physics and prior mathematics knowledge. *Journal of research in science teaching, 18*(4), 291-294.

Hudson, N. W., & Fraley, R. C. (2015). Volitional personality trait change: Can people choose to change their personality traits? *Journal of personality and social psychology, 109*(3), 490.

Hughes, S., Lyddy, F., & Lambe, S. (2013). Misconceptions about psychological science: A review. *Psychology Learning & Teaching, 12*(1), 20-31.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362. doi:10.1016/0001-8791(86)90013-8

Hunter, L. W., & Thatcher, S. M. (2007). Feeling the heat: Effects of stress, commitment, and job experience on job performance. *Academy of Management Journal, 50*(4), 953-968.

Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive science quarterly, 2*(1), 5-62.

Izsak, A. (2005). " You Have to Count the Squares": Applying Knowledge in Pieces to Learning Rectangular Area. *The Journal of the Learning Sciences, 14*(3), 361-403.

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M. S., & Krugers, H. J. (2006). Learning under stress: how does it work? *Trends in cognitive sciences, 10*(4), 152-158.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The big five inventory—versions 4a and 54: Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research, 3*(2), 114-158.

Johnson, A. (2003). Procedural memory and skill acquisition. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 499-523). Hoboken, NJ: Wiley.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin, 114*(1), 3-28. doi:10.1037/0033-2909.114.1.3

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika, 41*(1/2), 133-145. doi:10.2307/2333011

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of personality and social psychology, 57*(3), 469.

Jussim, L., & Eccles, J. S. (1992). Teacher expectations: II. Construction and reflection of student achievement. *Journal of personality and social psychology, 63*(6), 947.

Juster, R.-P., Bizik, G., Picard, M., Arsenault-Lapierre, G., Sindi, S., Trepanier, L., . . . Lord, C. (2011). A transdisciplinary perspective of chronic stress in relation to psychopathology throughout life span development. *Development and psychopathology, 23*(3), 725-776.

Juster, R.-P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews, 35*(1), 2-16.

Kainulainen, M., McMullen, J., & Lehtinen, E. (2017). Early developmental trajectories toward concepts of rational numbers. *Cognition and Instruction, 35*(1), 4-19. doi:10.1080/07370008.2016.1251287

Kalakoski, V. (2007). Effect of skill level on recall of visually presented patterns of musical notes. *Scandinavian Journal of Psychology, 48*(2), 87-96.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31. doi:10.1207/S15326985EP3801_4

Keil, F. C. (1996). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kennedy, B. L., Schwab, J. J., Morris, R. L., & Beldia, G. (2001). Assessment of State and Trait Anxiety in Subjects with Anxiety and Depressive Disorders. *Psychiatric Quarterly, 72*(3), 263-276. doi:10.1023/a:1010305200087

Kerfoot, B. P., Shaffer, K., McMahon, G. T., Baker, H., Kirdar, J., Kanter, S., . . . Armstrong, E. G. (2011). Online "spaced education progress-testing" of students to confront two upcoming challenges to medical schools. *Academic Medicine, 86*(3), 300-306.

Kertes, D. A., & van Dulmen, M. (2012). Latent state trait modeling of children’s cortisol at two points of the diurnal cycle. *Psychoneuroendocrinology, 37*(2), 249-255. doi:10.1016/j.psyneuen.2011.06.009

Kintsch, W. (1994). Text comprehension, memory, and learning. *American psychologist, 49*(4), 294-303.

Kirschbaum, C., Steyer, R., Eid, M., Patalla, U., Schwenkmezger, P., & Hellhammer, D. (1990). Cortisol and behavior: 2. Application of a latent state-trait model to salivary cortisol. *Psychoneuroendocrinology, 15*(4), 297-307.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86. doi:10.1207/s15326985ep4102_1

Köhler, T. (1990). Daily stress as a trigger of migraine attacks: Results of thirteen single-subject studies. *Journal of Consulting and Clinical Psychology, 58*(6), 870.

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research synthesis methods, 2*(1), 61-76.

Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158-170. doi:10.1016/j.learninstruc.2008.03.003

Kudielka, B. M., Buske-Kirschbaum, A., Hellhammer, D., & Kirschbaum, C. (2004). HPA axis responses to laboratory psychosocial stress in healthy elderly adults, younger adults, and children: impact of age and gender. *Psychoneuroendocrinology, 29*(1), 83-98.

Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., & Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology, 29*(8), 983-992.

Kuhle, B. X., Barber, J. M., & Bristol, A. S. (2009). Predicting students' performance in introductory psychology from their psychology misconceptions. *Journal of Instructional Psychology, 36*(2), 119.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of educational research, 75*(1), 63-82.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2007). A Meta-Analysis of the Predictive Validity of the Graduate Management Admission Test (GMAT) and Undergraduate Grade Point Average (UGPA) for Graduate Student Academic Performance. *Academy of Management Learning & Education, 6*(1), 51-68.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success.

Science.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological bulletin, 127*(1), 162.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of personality and social psychology, 86*(1), 148.

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1), 101-107.

Kuncel, N. R., Wee, S., Serafin, L., & Hezlett, S. A. (2010). The validity of the Graduate Record Examination for master's and doctoral programs: A meta-analytic investigation. *Educational and Psychological Measurement, 70*(2), 340-352.

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: a meta-analytic review *Review of educational research, 70*(4), 443-484.

Laird, J. E. (2012). *The SOAR cognitive architecture*. Cambridge, MA: MIT Press.

Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica, 47*(3), 111-121.

Lanza, S. T., Tan, X., & Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural equation modeling: a multidisciplinary journal, 20*(1), 1-26. doi:10.1080/10705511.2013.742377

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of personality and social psychology, 61*(1), 132-140.

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *Journal of research in science teaching, 20*(2), 117-129. doi:10.1002/tea.3660200204

Lawson, A. E. (2000). How do humans acquire knowledge? And what does that imply about the nature of knowledge? *Science & Education, 9*(6), 577-598.

Lazarus, R. S. (2000). Toward better research on stress and coping.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal and coping*.

Lee, E.-H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian nursing research, 6*(4), 121-127.

Lee, Y., & Nelson, D. W. (2005). Viewing or visualising--Which concept map strategy works best on problem-solving performance? *British Journal of Educational Technology, 36*(2), 193-203. doi:10.1111/j.1467-8535.2005.00452.x

Lejeune, C., Desmottes, L., Catale, C., & Meulemans, T. (2015). Age difference in dual-task interference effects on procedural learning in children. *Journal of Experimental Child Psychology, 129*(January 2015), 165-172. doi:10.1016/j.jecp.2014.07.007

LePine, J. A., Podsakoff, N. P., & LePine, M. A. (2005). A meta-analytic test of the challenge stressor–hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal, 48*(5), 764-775.

Leung, D. Y., Lam, T.-h., & Chan, S. S. (2010). Three versions of Perceived Stress Scale: validation in a sample of Chinese cardiac patients who smoke. *BMC public health, 10*(1), 1.

Lewandowsky, S., & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory & Cognition, 28*(2), 295-305. doi:10.3758/BF03213807

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243-264). New York: Cambridge University Press.

Linn, R. L., & Hastings, C. N. (1984). A meta analysis of the validity of predictors of performance in law school. *Journal of Educational Measurement, 21*(3), 245-259.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Liverant, G. I., Hofmann, S. G., & Litz, B. T. (2004). Coping and anxiety in college students after the September 11th terrorist attacks. *Anxiety, Stress & Coping, 17*(2), 127-139.

Lix, L. M., & Keselman, H. J. (1998). To trim or not to trim: Tests of location equality under heteroscedasticityand nonnormality. *Educational and Psychological Measurement, 58*(3), 409-429.

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive psychology, 7*(4), 560-572. doi:10.1016/0010-0285(75)90023-7

Lonigan, C. J., Burgees, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596-613. doi:10.1037/0012-1649.36.5.596

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung*. Eugene, OR: University of Oregon Press.

Luchins, A. S., & Luchins, E. H. (1987). Einstellung effects. *Science, 238*(4827), 598.

Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and cognition, 65*(3), 209-237.

Lust, K., Ehlinger, E., & Golden, D. (2015). *College student health survey report: Health and health-related behaviors Minnesota postsecondary students*. Retrieved from http://www.bhs.umn.edu/surveys/index.htm.

Lynn, S. J., & McConkey, K. M. (Eds.). (1998). *Truth in memory*. New York: Guilford.

Ma, J., Pender, M., & Welch, M. (2016). Education Pays 2016: The Benefits of Higher Education for Individuals and Society. Trends in Higher Education Series. *College Board*.

Machery, E. (2010). Précis of doing without concepts. *Behavioral and Brain Sciences, 33*(2-3), 195-206. doi:10.1111/j.1468-0017.2010.01403.x

Magidson, J. F., Roberts, B. W., Collado-Rodriguez, A., & Lejuez, C. (2014). Theory-driven intervention for changing personality: Expectancy value theory, behavioral activation, and conscientiousness. *Developmental Psychology*, *50*(5), 1442.

Maris, E. (1998). Covariance adjustment versus gain scores - revisited. *Psychological Methods, 3*, 309-327. doi:10.1037/1082-989X.3.3.309

May, K., & Hittner, J. B. (2010). Reliability and validity of gain scores considered graphically. *Perceptual and Motor Skills, 111*, 399-406. doi:10.2466/03.PMS.111.5.399-406

Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.

McCutcheon, L. E., Hanson, E., Apperson, J. M., & Wynn, V. (1992). Relationships among critical thinking skills, academic achievement, and misconceptions about psychology. *Psychological reports, 71*(2), 635-639.

McEwen, B. S. (2008). Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European journal of pharmacology, 583*(2-3), 174-185.

McGonagle, K. A., & Kessler, R. C. (1990). Chronic stress, acute stress, and depressive symptoms. *American journal of community psychology, 18*(5), 681-706.

McMullen, J., Laakkonen, E., Hannula-Sormunen, M., & Lehtinen, E. (2015). Modeling the developmental trajectories of rational number concept (s). *Learning and Instruction, 37*(1), 14-20. doi:10.1016/j.learninstruc.2013.12.004

McMullen, J., Van Hoof, J., Degrande, T., Verschaffel, L., & Van Dooren, W. (2018). Profiles of rational number knowledge in Finnish and Flemish students–A multigroup latent class analysis. *Learning and Individual Differences*.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254-278. doi:10.1037/0033-295X.100.2.254

Melcher, J. M., & Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language, 35*(2), 231-245. doi:10.1006/jmla.1996.0013

Merz, C. J., Dietsch, F., & Schneider, M. (2016). The impact of psychosocial stress on conceptual knowledge retrieval. *Neurobiology of learning and memory, 134*, 392-399.

Misra, R., & McKean, M. (2000). College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction. *American Journal of Health Studies, 16*(1), 41.

Möhring, A., Schroeders, U., & Wilhelm, O. (2018). Knowledge is power for medical assistants: Crystallized and fluid Intelligence as predictors of vocational knowledge. *Frontiers in psychology, 9*, 28. doi:10.3389/fpsyg.2018.00028

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364-386. doi:10.1177/1094428106291059

Müller, B. (1999). Use specifity of cognitive skills: Evidence for production rules? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 191-207.

Mullis, I. V., Martin, M., Foy, P., & Arora, A. (2012). TIMMS International Results in Mathematics. *Chestnut Hill, MA: TIMMS & PIRLS International Study Center, Boston College. Retrieved from the web at http://timss. bc. edu/timss2011/downloads/T11_IR_ Mathematics_FullBook. pdf*.

Mullis, I. V., Martin, M. O., Foy, P., & Drucker, K. T. (2012). *PIRLS 2011 International Results in Reading*: ERIC.

Muthén, B. O. (2001). Latent variable mixture modeling. *New developments and techniques in structural equation modeling*, 1-33.

Muthén, B. O., & Asparouhov, T. (2011). LTA in Mplus: Transition probabilities influenced by covariates. *Mplus Web Notes, 13*, 1-30.

Muthén, B. O., & Muthén, L. K. (1998-2012). *Mplus software*.

Muthén, L. K., & Muthén, B. O. (1998-2016). *Mplus: Statistical analysis with latent variables: User's guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural equation modeling, 9*(4), 599-620.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience, 57*(3), 263-272.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Sternberg, R. J. (1996). Intelligence: Knowns and unknowns. *American psychologist, 51*(2), 77.

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education, 8*(3), 545-563.

Ng, E., & Lee, K. (2015). Effects of trait test anxiety and state anxiety on children's working memory task performance. *Learning and Individual Differences, 40*, 141-148. doi:http://dx.doi.org/10.1016/j.lindif.2015.04.007

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Niedźwieńska, A., Neckar, J., & Baran, B. (2007). Development and validation of the implicit memory theory scale. *European Journal of Psychological Assessment, 23*(3), 185-192. doi:10.1027/1015-5759.23.3.185

Nonis, S. A., Hudson, G. I., Logan, L. B., & Ford, C. W. (1998). Influence of perceived control over time on college students' stress and stress-related outcomes. *Research in higher education, 39*(5), 587-605.

Nugent, P. M. S. (2013). *Achievement* PsychologyDictionary.org, Retrieved from https://psychologydictionary.org/knowledge/

Nunnally, J. C. (1967). *Psychometric theory* (1st ed.). New York: McGraw-Hill.

Nylund, K. L. (2007). *Latent transition analysis: Modeling extensions and an application to peer victimization.* University of California, Los Angeles Doctoral dissertation.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling, 14*(4), 535-569.

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971-990.

OECD. (2000). *Literacy in the Information Age. Final report of the international adult literacy survey*. Paris: OECD Publishing.

OECD. (2012). *Education at a glance 2012: OECD indicators*. Paris: OECD Publishing.

OECD. (2014). *Education at a glance 2014: OECD indicators*. Paris: OECD Publishing.

OECD. (2016). *PISA 2015 assessment and analytical framework: Science, reading,* *mathematic and financial literacy*. Paris: OECD Publishing.

OECD. (2017). What is PISA? In OECD (Ed.), *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. Paris: OECD Publishing.

Orlich, D. C., Harder, R. J., Callahan, R. C., Trevisan, M. S., & Brown, A. H. (2010). *Teaching Strategies: A Guide to Effective Instruction* (Ninth ed.). Boston, MA: Wadsworth.

Ormrod, J. E. (2011). *Human learning* (Fourth ed.). Upper Saddle River, New Jersey: Pearson Education.

Oropeza, B. A. C., Fitzgibbon, M., & Barón, A. (1991). Managing mental health crises of foreign college students. *Journal of Counseling and Development, 69*(3), 280-284.

Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., & Lupianez, J. (2010). Attention and anxiety: different attentional functioning under state and trait anxiety. *Psychological Science, 21*(2), 298-304. doi:10.1177/0956797609359624

Pfost, M., Dörfler, T., & Artelt, C. (2011). Reading competence development of poor readers in a German elementary school sample: an empirical examination of the Matthew effect model. *Journal of Research in Reading, 35*, 411-426. doi:10.1111/j.1467-9817.2010.01478.x

Piaget, J. (1971). *Biology and knowledge*. Chicago: University of Chicago Press.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of educational research, 63*(2), 167-199.

Plant, E. A., Ericsson, K. A., Hill, L., & Asberg, K. (2005). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology, 30*(1), 96-116.

Pluut, H., Curşeu, P. L., & Ilies, R. (2015). Social and study related stressors and resources among university entrants: Effects on well-being and academic performance. *Learning and Individual Differences, 37*, 262-268.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin, 135*(2), 322.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227. doi:10.1002/sce.3730660207

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education, 13*(1), 21-43. doi:10.1007/s10763-014-9520-6

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate behavioral research, 42*(1), 185-227.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*(1), 91-109. doi:10.1207/s15326985ep2701_7

Pritchard, M. E., Wilson, G. S., & Yamnitz, B. (2007). What predicts adjustment among college students? A longitudinal panel study. *Journal of American College Health, 56*(1), 15-22.

Pruessner, J. C., Hellhammer, D. H., & Kirschbaum, C. (1999). Burnout, perceived stress, and cortisol responses to awakening. *Psychosomatic medicine*, *61*(2), 197-204.

Pruessner, M., Hellhammer, D. H., Pruessner, J. C., & Lupien, S. J. (2003). Self-reported depressive symptoms and stress levels in healthy young men: associations with the cortisol response to awakening. *Psychosomatic medicine, 65*(1), 92-99.

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational psychology*, *103*(3), 683.

Pungello, E. P., Kupersmidt, J. B., Burchinal, M. R., & Patterson, C. J. (1996). Environmental risk factors and children's achievement from middle childhood to early adolescence. *Developmental Psychology, 32*(4), 755.

Pustejovsky, J. (2017). ClubSandwich: Cluster robust (sandwich) variance estimators with small-sample corrections. R package version 0.2.3.

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. .   Retrieved from http://www.R-project.org/

Rammstedt, B. (1997). *Die deutsche version des Big Five Inventory (BFI): Übersetzung und validierung eines fragebogens zur erfassung des fünf-faktoren-modells der persönlichkeit.* (Doctoral Dissertation).

Rammstedt, B., & John, O. P. (2005). Kurzversion des big five inventory (BFI-K). *Diagnostica, 51*(4), 195-206.

Ramsden, P. (2003). *Learning to teach in higher education*. New York, NY: Routledge.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 295-315). New York: Russell Sage Foundation.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology, 80*(6), 721-730. doi:10.1037/0021-9010.80.6.721

Regehr, C., Glancy, D., & Pitts, A. (2013). Interventions to reduce stress in university students: a review and meta-analysis. *Journal of affective disorders, 148*(1), 1-11. doi:10.1016/j.jad.2012.11.026

Reisberg, L. (2000). Student Stress Is Rising, Especially among Women. *Chronicle of Higher Education, 46*(21), 49-52.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist, 31*, 115-121. doi:10.1207/s15326985ep3102_3

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin, 138*(2), 353.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667-706.

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology, 53*(4), 713.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587-597. doi:10.1007/s10648-015-9302-x

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: American Psychological Association.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological bulletin, 126*(1), 3.

Roberts, R., Golding, J., Towell, T., & Weinreb, I. (1999). The effects of economic circumstances on British students' mental and physical health. *Journal of American College Health, 48*(3), 103-109.

Robotham, D. (2008). Stress among higher education students: Towards a research agenda.

Higher Education, 56(6), 735-746.

Robotham, D., & Julian, C. (2006). Stress and the higher education student: a critical review of the literature. *Journal of further and higher education, 30*(02), 107-117.

Ross, S. E., Niebling, B. C., & Heckert, T. M. (1999). Sources of stress among college students. *Social psychology, 61*(5), 841-846.

Roth, P. L., BeVier, C. A., III, F. S. S., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance: American Psychological Association.

Roth, P. L., & Clarke, R. L. (1998). Meta-analyzing the relation between grades and salary. *Journal of Vocational Behavior, 53*(3), 386-400.

Ruch, W., Köhler, G., & Van Thriel, C. (1997). To be in good or bad humour: Construction of the state form of the State-Trait-Cheerfulness-Inventory—STCI. *Personality and Individual Differences, 22*(4), 477-491.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science, 331*(6022), 1269-1270. doi:10.1126/science.1198976

Saklofske, D. H., Austin, E. J., Mastoras, S. M., Beaton, L., & Osborne, S. E. (2012). Relationships of personality, affect, emotional intelligence and coping with student stress and academic success: Different patterns of association for stress and success. *Learning and Individual Differences, 22*(2), 251-257.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science*, *26*, 515–520. doi:10.1177/0963721417712760

Samson, G. E., Graue, M. E., Weinstein, T., & Wahlberg, H. J. (1984). Academic and occupational performance: A quantitative synthesis. *American educational research journal, 21*(2), 311-321. doi:10.3102/00028312021002311

Satorra, A., & Bentler, P. M. (2010). Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika, 75*(2), 243-248. doi:10.1007/s11336-009-9135-y

Satpute, A. B., Mumford, J. A., Naliboff, B. D., & Poldrack, R. A. (2012). Human anterior and posterior hippocampus respond distinctly to state and trait anxiety. *Emotion, 12*(1), 58-68.

Saxbe, D. E. (2008). A field (researcher's) guide to cortisol: tracking HPA axis functioning in everyday life. *Health Psychology Review, 2*(2), 163-190.

Schiefele, U. (1991). Interest, Learning, and Motivation. *Educational Psychologist, 26*(3-4), 299-323. doi:10.1080/00461520.1991.9653136

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology, 6*(4), 147-151. doi:10.1027/1614-2241/a000016

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Schmitt, M. J., & Steyer, R. (1993). A latent state-trait model (not only) for social desirability. *Personality and Individual Differences, 14*(4), 519-529.

Schneider, M. (2012). Knowledge integration. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1684-1686). New York: Springer.

Schneider, M., & Grabner, R. H. (2012). Fact learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1264-1266). New York: Springer.

Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology, 49*(9), 1639-1649. doi:10.1037%2Fa0030976

Schneider, M., & Preckel, F. (2017). Variables Associated with Achievement in Higher Education: A Systematic Review of Meta-analyses. *Psychological bulletin*. doi:Advance online publication. http://dx.doi.org/10.1037/bul0000098

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology*, *46*, 1525-1538.

Schneider, M., Vamvakoussi, X., & Van Dooren, W. (2012). Conceptual change. *Encyclopedia of the sciences of learning*, 735-738.

Schneider, W. (1993). Domain-specific knowledge and memory performance in children. *Educational Psychology Review*, *5*(3), 257-273.

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

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(5), 484-494. doi:10.1037/0278-7393.8.5.484

Schrader, F.-W., & Helmke, A. (2015). School achievement: Motivational determinants and processes. In J. Wright (Ed.), *International encyclopedia of the social and behavioral sciences* (2nd ed., Vol. 21, pp. 48-54). Oxford: Elsevier.

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.2016.08.003

Schuler, H., Funke, U., & Baron-Boldt, J. (1990). Predictive Validity of School Grades-A Meta-analysis. *Applied psychology, 39*(1), 89-103.

Schulz, P., Kirschbaum, C., Prüßner, J., & Hellhammer, D. (1998). Increased free cortisol secretion after awakening in chronically stressed individuals due to work overload. *Stress and Health, 14*(2), 91-97.

Schwartz, D., Gorman, A. H., Nakamoto, J., & Toblin, R. L. (2005). Victimization in the Peer Group and Children's Academic Functioning. *Journal of Educational psychology, 97*(3), 425.

Schwippert, K. (2007). *Progress in reading literacy: The impact of PIRLS 2001 in 13 countries*: Waxmann.

Seery, M. D., Weisbuch, M., Hetenyi, M. A., & Blascovich, J. (2010). Cardiovascular measures independently predict performance in a university course. *Psychophysiology, 47*(3), 535-539.

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin, 130*(4), 601-630.

Semb, G. B., Ellis, J. A., & Araujo, J. (1993). Long-term memory for knowledge learned in school. *Journal of Educational psychology, 85*(2), 305.

Shirtcliff, E. A., Granger, D. A., Booth, A., & Johnson, D. (2005). Low salivary cortisol levels and externalizing behavior problems in youth. *Development and psychopathology, 17*(1), 167-184.

Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naive theories of evolution. *Merrill-Palmer Quarterly, 59*(2), 141-167.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*(1), 118-137. doi:10.1111/tops.12174

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124*(2), 209-215. doi:10.1016/j.cognition.2012.04.005

Sideridis, G. D. (2006). Coping is not an ‘either’‘or’: The interaction of coping strategies in regulating affect, arousal and performance. *Stress and Health, 22*(5), 315-327.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in cognitive sciences, 17*(1), 13-19. doi:10.1016/j.tics.2012.11.004

Siegler, R. S., & Svetina, M. (2002). A microgenetic/cross-sectional study of matrix completion: Comparing short-term and long-term change. *Child Development, 73*(3), 793-809.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith III., J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences, 3*(2), 115-163.

Snyder, C., Lopez, S. J., Aspinwall, L., Fredrickson, B. L., Haidt, J., Keltner, D., . . . Wrzesniewski, A. (2002). The future of positive psychology: A declaration of independence.

Spielberger, C. D., & Sydeman, S. J. (1994). State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory. In M. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 292-321). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Stalder, T., Steudte, S., Miller, R., Skoluda, N., Dettenborn, L., & Kirschbaum, C. (2012). Intraindividual stability of hair cortisol concentrations. *Psychoneuroendocrinology, 37*(5), 602-610.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy *Reading Research Quarterly, 21*(4), 360-407. doi:10.1177/0022057409189001-204

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School, 3*, 268-275.

Steinberg, L. (2014). *Adolescence* (10th Ed.). New York: McGraw-Hill.

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. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality, 13*(5), 389.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits-revised. *Annual review of clinical psychology, 11*, 71-98.

Stiggins, R. J. (2001). *Student-involved classroom assessment*. Upper Saddle River, New Jersey: Prentice Hall.

Stöber, J. (1997). Trait anxiety and pessimistic appraisal of risk and chance. *Personality and Individual Differences, 22*(4), 465-476.

Strayhorn, T. L. (2013). Academic Achievement - A Higher Education Perspective. In J. Hattie & E. M. Anderman (Eds.), *International Guide to Student Achievment* (pp. 16-18). New York: Routledge.

Stroud, C. B., Chen, F. R., Doane, L. D., & Granger, D. A. (2016). Individual differences in early adolescents' latent trait cortisol (LTC): Relation to recent acute and chronic stress. *Psychoneuroendocrinology, 70*, 38-46.

Sundberg, M. D., & Dini, M. L. (1993). Science majors vs nonmajors: is there a difference? *Journal of College Science Teaching, 23*, 299–304.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Essex: Pearson Education Limited.

Tanaka, D., Kiyokawa, S., Yamada, A., Dienes, Z., & Shigemasu, K. (2008). Role of selective attention in artificial grammar learning. *Psychonomic Bulletin & Review, 15*(6), 1154-1159. doi:10.3758/PBR.15.6.1154

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., Amsterdam, J., Shelton, R., & Schalet, B. (2009). Personality change during depression treatment: a placebo-controlled trial. *Archives of general psychiatry, 66*(12), 1322-1330.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations and a software tutorial in Stata and SPSS. *Research synthesis methods, 5*(1), 13-30. doi:10.1002/jrsm.1091

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*(1), 85-112. doi:10.1007/s40865-016-0026-5

Tata, J. (1999). Grade distributions, grading procedures, and students' evaluations of instructors: A justice perspective. *The Journal of Psychology, 133*(3), 263-271.

Taylor, A. K., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength, and sources of misconceptions. *The Psychological Record, 54*(1), 15.

Taylor, N. T. (2001). Using constructivism to inform tertiary chemistry pedagogy. *Chemistry Education Research and Practice, 2*(3), 215-226.

Techniker Krankenkasse. (2015a). *Gesundheitsreport der Techniker Krankenkasse 2015 - Gesundheit von Studierenden*. Hamburg: Techniker Krankenkass.

Techniker Krankenkasse. (2015b). *TK Campus Kompass*. Hamburg: Techniker Krankenkasse.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties arepresent in one ranking. *Indagationes Mathematicae, 14*(3), 327-333.

Thoits, P. A. (1986). Social support as coping assistance. *Journal of Consulting and Clinical Psychology, 54*(4), 416-423. doi:10.1037/0022-006X.54.4.416

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology, 30*(2), 96-101.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions II: The estimation of magnitudes. *Psychological Review, 8*(4), 384-395. doi:10.1037/h0071280

Tian, X., Wei, D., Du, X., Wang, K., Yang, J., Liu, W., . . . Qiu, J. (2016). Assessment of trait anxiety and prediction of changes in state anxiety using functional brain imaging: A test–retest study. *NeuroImage, 133*, 408-416.

Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, *21*(1), 11-27. doi:10.1024/1010-0652.21.1.11

Twist, L. (2012). The impact of PIRLS in England. *Progress in Reading Literacy in National and International Context. The Impact of PIRLS 2006 in 12 Countries*, 75.

Tynjälä, P. (1999). Towards expert knowledge? A comparison between a constructivist and a traditional learning environment in the university. *International Journal of Educational Research, 31*(5), 357-442. doi:10.1016/S0883-0355(99)00012-9

Ullén, F., Hambrick, D. Z., & Mosing, M. A. (2015). Rethinking expertise: A multifactorial gene–environment interaction model of expert performance. *Psychological bulletin, 142*(4), 427.

UNESCO. (2012). *International Standard Classification of Education 2011 (ISCED 2011)*. Montreal, Quebec, Canada: UNESCO Institute for Statistics

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior research methods,* *45*(2), 576-594.

Van Der Ploeg, E., & Kleber, R. J. (2003). Acute and chronic job stressors among ambulance personnel: predictors of health symptoms. *Occupational and environmental medicine, 60*(suppl 1), i40-i46.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: Schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience, 26*(10), 2250-2261. doi:10.1162/jocn_a_00630

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1-48.

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Reserach Synthesis methods, 1*(2), 112-125. doi:10.1002/jrsm.11

Vogel, S., Fernández, G., Joëls, M., & Schwabe, L. (2016). Cognitive adaptation under stress: a case for the mineralocorticoid receptor. *Trends in cognitive sciences, 20*(3), 192-203.

Vogel, S., Kluen, L. M., Fernández, G., & Schwabe, L. (2018). Stress affects the neural ensemble for integrating new information and prior knowledge. *NeuroImage, 173*, 176-187. doi:https://doi.org/10.1016/j.neuroimage.2018.02.038

Vogel, S., & Schwabe, L. (2016). Learning and memory under stress: implications for the classroom. *npj Science of Learning, 1*, 16011.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45-69.

Vosniadou, S. (2008). *International handbook for research on conceptual change*. New York, NY: Routledge.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive psychology, 24*(4), 535-585.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 11*(4-5), 381-419.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The Framework Theory Approach to the Problem of Conceptual Change. In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change* (pp. 3-34). New York, NY: Routledge.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Walberg, H. J., & Tsai, S.-L. (1983). Matthew effect in education. *American educational* *research journal, 20*, 359-373.

Wang, J., Lesage, A., Schmitz, N., & Drapeau, A. (2008). The relationship between work stress and mental disorders in men and women: findings from a population-based study. *Journal of Epidemiology and Community Health, 62*(1), 42-47.

Warttig, S. L., Forshaw, M. J., South, J., & White, A. K. (2013). New, normative, English-sample data for the short form Perceived Stress Scale (PSS-4). *Journal of health psychology, 18*(12), 1617-1628.

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., . . . Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*(1), 264-271. doi:10.1016/j.concog.2013.01.004

Weinert, F. E. (1999). Definition and selection of competencies: Concepts of Competence. *Organization for Economic Co-operation and Development*.

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual review of psychology, 43*(1), 337-375.

Whitman, N. A. (1985). Student Stress: Effects and Solutions. ERIC Digest 85-1.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*(1), 10-18.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.

Winne, P. H., & Nesbit, J. C. (2010). The Psychology of Academic Achievement. *Annual review of psychology, 61*(1), 653-678. doi:10.1146/annurev.psych.093008.100348

Winzer, R., Lindberg, L., Guldbrandsson, K., & Sidorchuk, A. (2018). Effects of mental health interventions for students in higher education are sustainable over time: a systematic review and meta-analysis of randomized controlled trials. *PeerJ, 6*, e4598.

Witter, R. A., Okun, M. A., Stock, W. A., & Haring, M. J. (1984). Education and subjective well-being: A meta-analysis. *Educational Evaluation and Policy Analysis, 6*(2), 165-173.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (2000). Negative transfer errors in sequential cognitive skills: Strong-but-wrong sequence application. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*(3), 601-625. doi:10.1037/0278-7393.26.3.601

Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte-Carlo study. *Frontiers in psychology, 5*, 920.

Wüst, S., Federenko, I., Hellhammer, D. H., & Kirschbaum, C. (2000). Genetic factors, perceived chronic stress, and the free cortisol response to awakening. *Psychoneuroendocrinology, 25*(7), 707-720.

Xu, X., Villafane, S. M., & Lewis, J. E. (2013). College students' attitudes toward chemistry, conceptual knowledge and achievement: structural equation model analysis. *Chemistry Education Research and Practice, 14*(2), 188-200.

Yeung, P.-s., Ho, C. S.-h., Chan, D. W.-o., & Chung, K. K.-h. (2016). Orthographic skills important to Chinese literacy development: The role of radical representation and orthographic memory of radicals. *Reading and Writing, 29*(9), 1935-1958. doi:10.1007/s11145-016-9659-3

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477-501.

Ystgaard, M., Tambs, K., & Dalgard, S. O. (1999). Life stress, social support and psychological distress in late adolescence: a longitudinal study. *Social Psychiatry and Psychiatric Epidemiology, 34*(1), 12-19. doi:10.1007/s001270050106

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: Evidence from eyetracking. *Frontiers in psychology, 3*, Article 148. doi:10.3389/fpsyg.2012.00148

Yusufov, M., Nicoloro-SantaBarbara, J., Grey, N. E., Moyer, A., & Lobel, M. (2018). Meta-analytic evaluation of stress reduction interventions for undergraduate and graduate students. *International Journal of Stress Management*, No Pagination Specified-No Pagination Specified. doi:10.1037/str0000099

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quartely, 27*(3), 447-457. doi:10. 1O16/j.ccrcsq.20 12.02.004

Zajacova, A., Lynch, S. M., & Espenshade, T. J. (2005). Self-efficacy, stress, and academic success in college. *Research in higher education, 46*(6), 677-706.

Zimmermann, D. W., & Williams, R. H. (1998). Reliability of gain scores under realistic assumptions about properties of pre-test and post-test scores. *Mathematical and Statistical Psychology, 51*, 343-351. doi:10.1111/j.2044-8317.1998.tb00685.x

Appendix A: Supplementary Materials (SM) of Study 1

SM 1: Studies Included in the Meta-Analysis

Aarnoutse, C., Van Leeuwe, J., & Verhoeven, L. (2005). Early Literacy From a Longitudinal Perspective. *Educational Research and Evaluation, 11*, 253-275. doi:10.1080/08993400500101054

Acuna, S. R., Rodicio, H. G., & Sanchez, E. (2011). Fostering active processing of instructional explanations of learners with high and low prior knowledge. *European Journal of Psychology of Education, 26*, 435-452. doi:10.1007/s10212-010-0049-y

Alao, S., & Guthrie, J. T. (1999). Predicting conceptual understanding with cognitive and motivational variables. *Journal of Educational Research, 92*, 243-254. doi: 10.1080/00220679909597602

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders, 36,* 312-333. doi:10.1097/TLD.0000000000000102

Amadieu, F., Tricot, A., & Marine, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior, 25,* 381-388. doi:10.1016/j.chb.2008.12.017

Amadieu, F., Tricot, A., & Marine, C. (2010). Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation. *Interacting with computers, 22,* 88-97. doi:10.1016/j.intcom.2009.07.001

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Marine, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19,* 376-386. doi:10.1016/j.learninstruc.2009.02.005

Aram, D., Korat, O., & Hassunah-Arafat, S. (2013). The contribution of early home literacy activities to first grade reading and writing achievements in Arabic. *Reading and Writing, 26*, 1517-1536. doi:10.1007/s11145-013-9430-y

Armand, F. (2001). Learning from expository texts; Effects of the interaction of prior knowledge and text structure on responses to different question types. *European Journal of Psychology of Education, 16*, 67-86. doi: 10.1007/BF03172995

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. *Journal of Educational Psychology, 96*, 699-713. doi:10.1037/0022-0663.96.4.699

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science, 17*, 775-785. doi:10.1111/desc.12155

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science, 25*. doi:10.1177/0956797614547539

Balch, W. R. (2015). Related General-Vocabulary Knowledge Transfers to Learning Technical Terms. *Teaching of Psychology, 42*, 254-259. doi:10.1177/0098628315589495

Bergey, B. W., Cromley, J. G., Kirchgessner, M. L., & Newcombe, N. S. (2015). Using diagrams versus text for spaced restudy: Effects on learning in 10th grade biology classes. *British Journal of Educational Psychology, 85*, 59-74. doi:10.1111/bjep.12062

Blayney, P. J., Kalyuga, S., & Sweller, J. (2015). Using Cognitive Load Theory to Tailor Instruction to Levels of Accounting Students' Expertise. *Educational Technology & Society, 18*, 199-210. Retrieved from http://www.jstor.org/stable/jeductechsoci.18.4.199

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016-1031. doi:10.1111/j.1467-8624.2008.01173.x

Braasch, J. L. G., & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes, 47*, 447-479. doi:10.1080/01638530903420960

Braten, I., Anmarkrud, O., Brandmo, C., & Stromso, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction, 30*, 9-24. doi:10.1016/j.learninstruc.2013.11.002

Bråten, I., Johansen, R.-P., & Strømsø, H. I. (2017). Effects of different ways of introducing a reading task on intrinsic motivation and comprehension. *Journal of Research in Reading, 40*, 17-36. doi:10.1111/1467-9817.12053

Braten, I., & Stromso, H. I. (2011). Measuring strategic processing when students read multiple texts. *Metacognition and Learning, 6*, 111-130. doi:10.1007/s11409-011-9075-7

Bryant, P., Nunes, T., & Barros, R. (2014). The connection between children's knowledge and use of grapho-phonic and morphemic units in written text and their learning at school. *British Journal of Educational Psychology, 84*, 211-225. doi:1 0.1 1 1 l/bjep.12030

Bulu, S. T., & Pedersen, S. (2012). Supporting problem-solving performance in a hypermedia learning environment: The role of students' prior knowledge and metacognitive skills. *Computers in Human Behavior, 28*, 1162-1169. doi:10.1016/j.chb.2012.01.026

Burgess, S. R. (2002). The influence of speech perception, oral language ability, the home literacy environment, and pre-reading knowledge on the growth of phonological sensitivity: A one-year longitudinal investigation. *Reading and Writing, 15*, 709-737. doi: 10.1023/A:1020954606695

Butcher, K. R., & Sumner, T. (2011). How does prior knowledge impact students' online learning behaviors? *International Journal of Cyber Behavior, Psychology and Learning, 1*(4), 1-18. doi:10.4018/ijcbpl.2011100101

Byrnes, J. P. (1992). The conceptual basis of procedural learning. *Cognitive Development, 7*, 235-257. doi:10.1016/0885-2014(92)90013-H

Callahan, D., & Drum, P. A. (1984). Reading ability and prior knowledge as predictors of eleven and twelve year olds' text comprehension. *Reading Psychology, 5*, 145-154. doi:10.1080/0270271840050117

Caravolas, M., Hulme, C., & Snowling, M. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751-774. doi:10.1006/jmla.2000.2785

Carroll, J. M. (2004). Letter knowledge precipitates phoneme segmentation, but not phoneme invariance. *Journal of Research in Reading, 27*, 212-225. doi:10.1111/j.1467-9817.2004.00228.x

Carroll, J. M., Snowling, M., Stevenson, J., & Hulme, C. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*, 913-923. doi:10.1037/0012-1649.39.5.913

Carson, K., Boustead, T., & Gillon, G. (2015). Content validity to support the use of a computer-based phonological awareness screening and monitoring assessment (Com-PASMA) in the classroom. *International Journal of Speech-Language Pathology, 17*, 500-510. doi:10.3109/17549507.2015.1016107

Cavallo, A. M. L., & Schafer, L. E. (1994). Relationships between students' meaningful learning orientation and their understanding of genetics topics. *Journal of Research in Science Teaching, 31*, 393-418. doi:10.1002/tea.3660310408

Chambers, S. K., & Andre, T. (1997). Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching, 34*, 107-123. doi:10.1002/(SICI)1098-2736(199702)34:2<107::AID-TEA2>3.0.CO;2-X

Chandran, S., Treagust, D. F., & Tobin, K. (1987). The role of cognitive factors in chemistry achievement. *Journal of Research in Science Teaching, 24*, 145-160. doi:10.1002/tea.3660240207

Chen, C.-H., & Wu, I.-C. (2012). The interplay between cognitive and motivational variables in a supportive online learning system for secondary physical education. *Computers & Education, 58*, 542-550. doi:10.1016/j.compedu.2011.09.012

Cheng, Y., Zhang, J., Wu, X., Liu, H., & Li, H. (2016). Cross-lagged relationships between morphological awareness and reading comprehension among chinese children. *Frontiers in Psychology, 7*, 1379. doi:10.3389/fpsyg.2016.01379

Chliounaki, K., & Bryant, P. (2007). How children learn about morphological spelling rules. *Child Development, 78,* 1360-1373. doi:10.1111/j.1467-8624.2007.01070.x

Cho, J.-R., Chiu, M. M., & McBride-Chang, C. (2011). Morphological awareness, phonological awareness, and literacy development in Korean and English: A 2-year longitudinal study. *Scientific Studies of Reading, 15*, 383-408. doi:10.1080/10888438.2010.487143

Chu, F. W., vanMarle, K., & Geary, D. C. (2016). Predicting children's reading and mathematics achievement from early quantitative knowledge and domain-general cognitive abilities. *Frontiers in Psychology, 7,* 775. doi:10.3389/fpsyg.2016.00775

Coiro, J. (2011). Predicting reading comprehension on the internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*, 352-392. doi:10.1177/1086296X11421979

Connor, C., Rice, D. C., Canto, A. I., Southerland, S. A., Underwood, P., Kaya, S., Morrison, F. J. (2012). Child characteristics by science instruction interactions in second and third grade and their relation to students' content-area knowledge, vocabulary, and reading skill gains. *Elementary School Journal, 113,* 52-75. doi:10.1086/665815

Conrad, N. J., & Deacon, S. H. (2016). Children's orthographic knowledge and their word reading skill: testing bidirectional relations. *Scientific Studies of Reading, 20*, 339-347. doi:10.1080/10888438.2016.1183128

Crawford, J., Dancer, J., & Pittenger, J. (1986). Initial performance level on a speechreading task as related to subsequent improvement after shortterm training. *Volta Review, 88*, 101-105.

Crevecoeur, Y. C., Coyne, M. D., & McCoach, D. B. (2014). English language learners and English-only learners' response to direct vocabulary instruction. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 30*, 51-78.

doi:10.1080/10573569.2013.758943

Dahlstrom, O. (2012). Learning during a collaborative final exam. *Education Research and Evaluation, 18*, 321-332. doi:10.1080/13803611.2012.674689

Dandache, S., Wouters, J., & Ghesquière, P. (2014). Development of reading and phonological skills of children at family risk for dyslexia: A longitudinal analysis from kindergarten to sixth grade. *Dyslexia, 20*, 305-329. doi:10.1002/dys.1482

Darabi, A., Arrastia-Lloyd, M. C., Nelson, D. W., Liang, X., & Farrell, J. (2015). Learning how the electron transport chain works: independent and interactive effects of instructional strategies and learners' characteristics. *Advances in Health Sciences Education, 20*, 1135-1148. doi:10.1007/s10459-015-9592-2

Davis, M. H., & Guthrie, J. T. (2015). Measuring reading comprehension of content area texts using an assessment of knowledge organization. *Journal of Educational Research, 108*, 148-164. doi:10.1080/00220671.2013.863749

Diakidoy, I.-A. N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction, 41*, 60-69. doi:10.1016/j.learninstruc.2015.10.002

Diakidoy, I.-A. N., Mouskounti, T., & Ioannides, C. (2011). Comprehension and learning from refutation and expository texts. *Reading Research Quarterly, 46*, 22-38. doi:10.1598/RRQ.46.l.2

Diana, E. M., & Webb, J. M. (1997). Using geographic maps in classrooms: The conjoint influence of individual differences and dual coding on learning facts. *Learning and Individual Differences, 9*(3), 195-214. doi:10.1016/S1041-6080(97)90006-6

Dittman, C. K. (2016). The impact of early classroom inattention on phonological processing and word-reading development. *Journal of Attention Disorders, 20*(8), 653-664. doi:10.1177/1087054713478979

Dresel, M., Ziegler, A., Broome, P., & Heller, K. A. (1998). Gender differences in science education: The double-edged role of prior knowledge in physics. *Roeper Review, 21*(2), 101-106. doi:10.1080/02783199809553939

Ercetin, G. (2010). Effects of topic interest and prior knowledge on text recall and annotation use in reading a hypermedia text in the L2. *ReCALL: Journal of Eurocall, 22*(2), 228-246. doi:10.1017/S0958344010000091

Fan, Y.-C., Wang, T.-H., & Wang, K.-H. (2011). A web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education, 57*(2), 1727-1740. doi:10.1016/j.compedu.2011.03.006

Fesel, S. S., Segers, E., Clariana, R. B., & Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: Evidence from pathfinder networks. *Computers in Human Behavior, 49*, 135-146. doi:10.1016/j.chb.2015.01.014

Florez, I. R. (2011). Case-based instruction in early childhood teacher preparation: Does it work? *Journal of Early Childhood Teacher Education, 32*(2), 118-134. doi:10.1080/10901027.2011.572227

Florit, E., Roch, M., & Levorato, M. C. (2014). Listening text comprehension in preschoolers: A longitudinal study on the role of semantic components. *Reading and Writing*, *27*(5), 793-817. doi:10.1007/s11145-013-9464-1

Fricke, S., Szczerbinski, M., Fox-Boyer, A., & Stackhouse, J. (2016). Preschool predictors of early literacy acquisition in German-speaking children. *Reading Research Quarterly, 51*(1), 29-53. doi:10.1002/rrq.116

Fricke, S., Szczerbinski, M., Stackhouse, J., & Fox-Boyer, A. V. (2008). Predicting individual differences in early literacy acquisition in German: The role of speech and language processing skills and letter knowledge. *Written Language and Literacy, 11*(2), 103-146. doi:10.1075/wll.11.2.02fri

Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading, 32*(3), 275-292. doi:JO.l l l l!j.1467-9817.2009.01393.x

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*(1), 82-97. doi:10.1037/edu0000053

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104*(4), 1094-1108. doi:10.1037/a0028389

Garcia-Madruga, J. A., & Fernandez-Corte, T. (1999). Learning complex historical knowledge at high school: The role of working memory. *Analise Psicologica, 17*(2), 241-252.

García, J. R., Bustos, A., & Sánchez, E. (2015). The contribution of knowledge about anaphors, organisational signals and refutations to reading comprehension. *Journal of Research in Reading, 38*(4), 405-427. doi:10.1111/1467-9817.12021

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. *Journal of Educational Psychology*, *78*(3), 225-227. doi:

10.1037/0022-0663.78.3.225

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing, 25*(2), 321-346. doi:10.1007/s11145-010-9271-x

Gobel, S. M., Watson, S. E., Lervag, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science, 25*(3), 789-798. doi:10.1177/0956797613516471

Goodrich, J. M., Lonigan, C. J., & Farver, J. M. (2013). Do early literacy skills in children's first language promote development of skills in their second language? An experimental evaluation of transfer. *Journal of Educational Psychology, 105*(2), 414-426. doi:10.1037/a0031780

Greene, J. A., Costa, L.-J., Robertson, J., Pan, Y., & Deekens, V. M. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education, 55*(3), 1027-1043. doi:10.1016/j.compedu.2010.04.013

Griggs, R. A., & Jackson, S. L. (1988). A reexamination of the relationship of high school psychology and natural science courses to performance in a college introductory psychology class. *Teaching of Psychology, 15*(3), 142-144. doi:10.1207/s15328023top1303_3

Guo, J.-p., & Pang, M. F. (2011). Learning a mathematical concept from comparing examples: the importance of variation and prior knowledge. *European Journal of Psychology of Education, 26*(4), 495-525. doi:10.1007/s10212-011-0060-y

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology, 84*(2), 182-192. doi:10.1037/0022-0663.84.2.182

Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*(1), 83-104. doi:10.1080/01638539409544860

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in mathematics: A structural model. *Educational Psychology, 28*(1), 59-71. doi:10.1080/01443410701413753

Hall, V. C., & Edmondson, B. (1992). Relative importance of aptitude and prior domain knowledge on immediate and delayed posttests. *Journal of Educational Psychology, 84*(2), 219-223. doi:10.1037/0022-0663.84.2.219

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition, 31*(6), 902-917. doi:10.3758/BF03196444

Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science, 42*(4), 639-657. doi:10.1007/s11251-013-9302-5

Helle, L., Nivala, M., Kronqvist, P., Ericsson, K. A., & Lehtinen, E. (2010). Do prior knowledge, personality and visual perceptual ability predict student performance in microscopic pathology? *Medical Education, 44*(6), 621-629. doi:10.1111/j.1365-2923.2010.03625.x

Hollingsworth, P. M., & Reutzel, D. R. (1990). Prior knowledge, content-related attitude, reading comprehension: Testing Mathewson's affective model of reading. *The Journal of Educational Research, 83*(4), 194-199. doi:10.1080/00220671.1990.10885956

Homer, B. D., & Plass, J. L. (2010). Expertise reversal for iconic representations in science visualizations. *Instructional Science, 38*(3), 259-276. doi:10.1007/s11251-009-9108-7

Hou, H.-T., & Li, M.-C. (2014). Evaluating multiple aspects of a digital educational problem-solving-based adventure game. *Computers in Human Behavior, 30*, 29-38. doi:10.1016/j.chb.2013.07.052

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23*(6), 572-577. doi:10.1177/0956797611435921

Huntsinger, C. S., Jose, P. E., & Luo, Z. (2016). Parental facilitation of early mathematics and reading skills andknowledge through encouragement of home-based activities. *Early Cildhood Research Quarterly, 37*, 1-15. doi:10.1016/j.ecresq.2016.02.005

Ise, E., Arnoldi, C. J., & Schulte-Korne, G. (2014). Development of orthographic knowledge in German-speaking children: A 2-year longitudinal study. *Journal of Research in Reading, 37*(3), 233-249. doi:10.1111/j.1467-9817.2012.01535.x

Jared, D., Cormier, P., Levy, B. A., & Wade-Woolley, L. (2011). Early predictors of biliteracy development in children in French immersion: A 4-year longitudinal study. *Journal of Educational Psychology, 103*(1), 119-139. doi:10.1037/a0021284

Johnson, A. M., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*(2), 97-115. doi:10.1111/jcal.12078

Johnson, M. A., & Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of Research in Science Teaching, 35*(1), 89-103. doi:10.1002/(SICI)1098-2736(199801)35:1<89::AID-TEA6>3.0.CO;2-J

Johnston, M. B., & Whitman, T. L. (1987). Enhancing math computation through variations in training format and instructional content. *Cognitive Therapy and Research, 11*(3), 381-397. doi:10.1007/BF01186288

Johnston, P. H. (1984). Prior knowledge and reading comprehension test bias. *Reading Research Quarterly, 19*(2), 219-239. doi:10.2307/747364

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology, 106*(4), 1096-1106. doi:10.1037/a0037100

Kahn-Horwitz, J., Sparks, R. L., & Goldstein, Z. (2012). English as a foreign language spelling development: A longitudinal study. *Applied Psycholinguistics, 33*(2), 343-363. doi:10.1017/S0142716411000397

Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior, 24*(3), 852-861. doi:10.1016/j.chb.2007.02.018

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition, 35*(7), 1567-1577. doi:10.3758/BF03193491

Kerfoot, B. P., Shaffer, K., McMahon, G. T., Baker, H., Kirdar, J., Kanter, S., Armstrong, E. G. (2011). Online "spaced education progress-testing" of students to confront two upcoming challenges to medical schools. *Academic Medicine, 86*(3), 300-306. doi:10.1097/ACM.0b013e3182087bef

Kirby, J. R., & Boulter, D. R. (1999). Spatial ability and transformational geometry. *European Journal of Psychology of Education*, *14*(2), 283-294. doi:10.1007/BF03172970

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*(2), 310-321. doi:10.1037/0022-0663.100.2.310

Klem, M., Hagtvet, B., Hulme, C., & Gustafsson, J.-E. (2016). Screening for language delay: Growth trajectories of language ability in low- and high-performing children. *Speech, Language, and Hearing Research, 59*, 1035-1045. doi:10.1044/2016_JSLHR-L-15-

0289

Klem, M., Melby-Lervag, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.-E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science, 18*(1), 146-154. doi:10.1111/desc.12202

Kobayashi, K. (2009). The influence of topic knowledge, external strategy use, and college experience on students' comprehension of controversial texts. *Learning and Individual Differences, 19*(1), 130-134. doi:10.1016/jl inclif.2008.06.001

Kopcha, T. J., & Sullivan, H. (2008). Learner preferences and prior knowledge in learner-controlled computer-based instruction. *Educational Technology Research and Development, 56*(3), 265-286. doi:10.1007/s11423-007-9058-1

Kramarski, B., Weiss, I., & Sharon, S. (2013). Generic versus context-specific prompts for supporting self-regulation in mathematical problem solving among students with low or high prior knowledge. *Journal of Cognitive Education and Psychology, 12*(2), 197-214. doi:10. 1891/1945-8959.12 .2. 197

Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158-170. doi:10.1016/j.learninstruc.2008.03.003

Kruk, R. S., Mayer, J., & Funk, L. (2014). The predictive relations between non-alphanumeric rapid naming and growth in regular and irregular word decoding in at-risk readers. *Journal of Research in Reading, 37*(1), 17-35. doi:10.1111/jrir.12005

Kyun, S. A., & Lee, H. (2009). The effects of worked examples in computer-based instruction: Focus on the presentation format of worked examples and prior knowledge of learners. *Asia Pacific Education Review, 10*(4), 495-503. doi:10.1007/s12564-009-9044-x

Lambiotte, J. G., & Dansereau, D. F. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *Journal of Experimental Education, 60*(3), 189-201. doi:10.1080/00220973.1992.9943875

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150-161. doi:10.1037/0022-0663.100.1.150

Lawson, A. E. (1983). Predicting science achievement: The role of developmental level, disembedding ability, mental capacity, prior knowledge, and beliefs. *Journal of Research in Science Teaching, 20*(2), 117-129. doi:10.1002/tea.3660200204

Lawson, A. E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowlege, prior belief and religious commitment. *Journal of Research in Science Teaching, 29*(2), 143-166. doi:10.1002/tea.3660290205

Le Bigot, L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445-470. doi:10.1080/10862960701675317

Lee, C.-Y., & Chen, M.-J. (2014). The impacts of virtual manipulatives and prior knowledge on geometry learning performance in junior high school. *Journal of Educational Computing Research, 50*(2), 179-201. doi:10.2190/EC.50.2.b

Lee, C. H., & Kalyuga, S. (2011). Effectiveness of on-screen pinyin in learning Chinese: An expertise reversal for multimedia redundancy effect. *Computers in Human Behavior, 27*(1), 11-15. doi:10.1016/j.chb.2010.05.005

Lee, K.-W. L., Goh, N.-K., Chia, L.-S., & Chin, C. (1996). Cognitive variables in problem solving in chemistry: A revisited study. *Science Education, 80*(6), 691-710. doi:10.1002/(SICI)1098-237X(199611)80:6<691::AID-SCE4>3.0.CO;2-E

Lee, Y., & Nelson, D. W. (2005). Viewing or visualising--Which concept map strategy works best on problem-solving performance? *British Journal of Educational Technology, 36*(2), 193-203. doi:10.1111/j.1467-8535.2005.00452.x

LeFevre, J.-A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology, 4*, ArtID 641. doi:10.3389/fpsyg.2013.00641

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science, 41*(5), 811-831. doi:10.1007/s11251-012-9256-z

Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of grade 3 reading comprehension. *Reading Research Quarterly, 51*(4), 373-390. doi:10.1002/rrq.145

Lepola, J., Lynch, J., Laakkonen, E., Silven, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4-6-year-old children. *Reading Research Quarterly, 47*(3), 259-282. doi:10.1002/rrq.020

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, D. (2005). Development of and Relationship Between Phonological and Motivational Processes and Naming Speed in

Predicting Word Recognition in Grade 1. *Scientific Studies of Reading, 9*(4), 367-399. doi:10.1207/s1532799xssr0904_3

Lerkkanen, M.-K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J.-E. (2005). Mathematical performance predicts progress in reading comprehension among 7-year olds. *European Journal of Psychology of Education, 20*(2), 121-137. doi:10.1007/BF03173503

Lerner, M. D., & Lonigan, C. J. (2016). Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: A growth curve analysis of the relation between two code-related skills. *Journal of Experimental Child Psychology, 144*, 166-183. doi:10.1016/j.jecp.2015.09.023

Lervag, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry, 51*(5), 612-620. doi:10.1111/j.1469-7610.2009.02185.x.

Lervag, A., & Hulme, C. (2010). Predicting the growth of early spelling skills: Are there heterogeneous developmental trajectories? . *Scientific Studies of Reading, 14*(6), 485-513. doi:10.1080/10888431003623488

Leslie, L., & Cooper, P. (1993). Assessing the predictive validity of prior-knowledge assessmen. *National Reading Conference Yearbook, 42*, 93-99.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly, 50*(1), 37-59. doi:10.1002/rrq.85

Levin, I., Ravid, D., & Rapaport, S. (2001). Morphology and spelling among Hebrew-speaking children: From kindergarten to first grade. *Journal of Child Language, 28*(3), 741-772.

Li, T., McBride-Chang, C., Wong, A., & Shu, H. (2012). Longitudinal predictors of spelling and reading comprehension in Chinese as an L1 and English as an L2 in Hong Kong Chinese children. *Journal of Educational Psychology, 104*(2), 286-301. doi:10.1037/a0026445

Liebfreund, M. D. (2015). Success with informational text comprehension: An examination of underlying factors. *Reading Research Quarterly, 50*(4), 387-392. doi:10.1002/rrq.109

Liebfreund, M. D., & Conradi, K. (2016). Component skills affecting elementary students' informational text comprehension. *Reading and Writing, 29*(6), 1141-1160. doi:10.1007/s11145-016-9629-9

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*(3), 482-494. doi:10.1037/0022-0663.95.3.482

Linnenbrink-Garcia, L., Pugh, K. J., Koskey, K. L. K., & Stewart, V. C. (2012). Developing conceptual understanding of natural selection: The role of interest, efficacy, and basic prior knowledge. *The Journal of Experimental Education, 80*(1), 45-68. doi:10.1080/00220973.2011.559491

Liu, H.-C., & Chuang, H.-H. (2011). Investigation of the impact of two verbal instruction formats and prior knowledge on student learning in a simulation-based learning environment. *Interactive Learning Environments, 19*(4), 433-446. doi:10 .1080/10494820903356940

Liu, T.-C., Lin, Y.-C., & Paas, F. (2014). Effects of prior knowledge on learning from different compositions of representations in a mobile learning environment. *Computers & Education, 72*, 328-338. doi:10.1016/j.compedu.2013.10.019

Liu, Y., Lin, D., & Zhang, X. (2016). Morphological awareness longitudinally predicts counting ability in Chinese kindergarteners. *Learning and Individual Differences, 47*(2016), 215-221. doi:10.1016/j.lindif.2016.01.007

Lo, L.-y., Yeung, P.-s., Ho, C. S.-H., Chan, D. W.-o., & Chung, K. (2015). The role of stroke knowledge in reading and spelling in Chinese. *Journal of Research in Reading., 00*(00), 1-22. doi:10.1111/1467-9817.12058

Lonigan, C. J., Burgees, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596-613. doi:10.1037//OOI2-1649.36.5.596

McBride-Chang, C., & Ho, C. S.-H. (2005). Predictors of Beginning Reading in Chinese and English: A 2-Year Longitudinal Study of Chinese Kindergartners. *Scientific Studies of Reading, 9*(2), 117-144. doi:10.1207/s1532799xssr0902_2

McGeown, S. P., Johnston, R. S., & Medford, E. (2012). Reading instruction affects the cognitive skills supporting early reading development. *Learning and Individual Differences, 22*(3), 360-364. doi:10.1016/j.lindif.2012.01.012

Moos, D. C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning, 4*(3), 197-216. doi:10.1007/s11409-009-9045-5

Morris, D., Bloodgood, J. W., Lomax, R. G., & Perney, J. (2003). Developmental steps in learning to read: A longitudinal study in kindergarten and first grade. *Reading Research Quarterly, 38*(3), 302-328. doi:10.1598/RRQ.38.3.1

Mou, Y., Li, Y., Hoard, M. K., Nugent, L. D., Chu, F., W., Rouder, J. N., & Geary, D. C. (2016). Developmental foundations of children's fraction magnitude knowledge. *Cognitive Development, 39*(2016), 141-153. doi:10.1016/j.cogdev.2016.05.002

Müller-Kalthoff, T., & Möller, J. (2003). The Effects of Graphical Overviews, Prior Knowledge, and Self-Concept on Hypertext Disorientation and Learning Achievement. *Journal of Educational Multimedia and Hypermedia, 12*(2), 117-134. Retrieved February 16, 2018 from https://www.learntechlib.org/p/14576/.

Müller-Kalthoff, T., & Möller, J. (2006). Browsing while reading: Effects of instructional design and learners' prior knowledge. *ALT-J Association for Learning Technology Journal, 14*(2), 183-198. doi:10.1080/09687760600668602

Murphy, K. P., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *Journal of Experimental Education*, *70*(3), 197-214. doi:10.1080/00220970209599506

Muter, V., Hulme, C., Snowling, M., & Stevenson, J. (2004). Phonemes, Rimes, Vocabulary, and Grammatical Skills as Foundations of Early Reading Development: Evidence From a Longitudinal Study. *Developmental Psychology, 40*(5), 665-681. doi:10.1037/0012-1649.40.5.665

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65*(3), 370-396. doi:10.1006/jecp.1996.2365

Muter, V., & Snowling, M. (1997). Grammar and phonology predict spelling in middle childhood. *Reading and Writing, 9*(5-6), 407-425. doi:10.1007/978-94-017-3054-9_6

Naslund, J. C., & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology*, *62*(1), 30-59. doi:10.1006/jecp.1996.0021

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly, 36*(2016), 550-560. doi:10.1016/j.ecresq.2016.02.003

Nielsen, A.-M. V., & Juul, H. (2016). Predictors of early versus later spelling development in Danish. *Reading and Writing, 29*(2), 245-266. doi:10.1007/s11145-015-9591-y

Nithart, C., Demont, E., Metz-Lutz, M.-N., Majerus, S., Poncelet, M., & Leybaert, J. (2011).

Early contribution of phonological awareness and later influence of phonological memory throughout reading acquisition. *Journal of Research in Reading, 34*(3), 346-363. doi:10.1111/j.1467-9817.2009.01427.x

Nunes, T., Bryant, P., & Bindman, M. (2006). The effects of learning to spell on children's awareness of morphology. *Reading and Writing, 19*(7), 767-787. doi:10.1007/s11145-006-9025-y

O'Brien, L. M., Meszaros, B., & Pulliam, W. E. (1985). Effects of teachers' use of objectives on student achievement in social studies. *Theory and Research in Social Education, 13*(3), 57-65. doi:10.1080/00933104.1985.10505506

O'Brien, L. M., Paratore, J. R., Leighton, C. M., Cassano, C. M., Krol-Sinclair, B., & Green, J. G. (2014). Examining differential effects of a family literacy program on language and literacy growth of English language learners with varying vocabularies. *Journal of Literacy Research, 46*(3), 383-415. doi:10.1177/1086296X14552180

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91-121. doi:10.1080/10888438.2010.529219

Ohst, A., Fondu, B., M.E. , Glogger, I., Nückles, M., & Renkl, A. (2014). Preparing learners with partly incorrect intuitive prior knowledge for learning. *Frontiers in Psychology, 5*(664), 1-11. doi:10.3389/fpsyg.2014.00664

Ohst, A., Glogger, I., Nückles, M., & Renkl, A. (2015). Helping preservice teachers with inaccurate and fragmentary prior knowledge to acquire conceptual understanding of psychological principles. *Psychology Learning & Teaching, 14*(1), 5-25. doi:10.1177/1475725714564925

Oortwijn, M., Boekaerts, M., & Vedder, P. (2008). The impact of the teacher's role and pupils' ethnicity and prior knowledge on pupils' performance and motivation to cooperate. *Instructional Science, 36*(3), 251-268. doi:10.1007/s11251-007-9032-7

Östergren, R., & Träff, U. (2013). Early number knowledge and cognitive ability affect early arithmetic ability. *Journal of Experimental Child Psychology, 115*(3), 405-421. doi:10.1016/j.jecp.2013.03.007

Pampush, J. D., & Petto, A. J. (2011). Familiarity with Latin and Greek anatomical terms and course performance in undergraduates. *Anatomical Sciences Education, 4*(1), 9-15. doi:10.1002/ase.189

Pan, J., McBride-Chang, C., Shu, H., Liu, H., Zhang, Y., & Li, H. (2011). What is in the naming? A 5-year longitudinal study of early rapid naming and phonological sensitivity in relation to subsequent reading skills in both native Chinese and English as a second language. *Journal of Educational Psychology, 103*(4), 897-908. doi:10.1037/a0024344

Park, S. I., Lee, G., & Kim, M. (2009). Do students benefit equally from interactive computer simulations regardless of prior knowledge levels? *Computers & Education, 52*(3), 649-655. doi:10.1016/j.compedu.2008.11.014

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*(2), 165-184. doi:10.1016/j.cogdev.2006.09.001

Philips, B. M., Lonigan, C. J., & Wyatt, M. A. (2009). Predictive validity of the Get Ready to Read! screener: Concurrent and long-term relations with reading-related skills. *Journal of Learning Disabilities, 42*(2), 133-147. doi:10.1177/0022219408326209

Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology, 104*(4), 945-958. doi:10.1037/a0027757

Piquard-Kipffer, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading-skills: A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8). *L'Annee Psychologique, 113*(4), 491-521. doi:10.4074/S0003503313014012

Poe, M. D., Burchinal, M. R., & Roberts, J. E. (2004). Early language and the development of children's reading skills. *Journal of School Psychology*, *42*(4), 315-332. doi:10.1016/j.jsp.2004.06.001

Priebe, S. J., Keenan, J. M., & Miller, A. C. (2012). How prior knowledge affects word identification and comprehension. *Reading and Writing, 25*(1), 131-149. doi:10.1007/s11145-010-9260-0

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86*(1), 159-175. doi:10.1111/cdev.12292

Randler, C., Ilg, A., & Kern, J. (2005). Cognitive and Emotional Evaluation of an Amphibian Conservation Program for Elementary School Students. *Journal of Environmental Education, 37*(1), 43-52. doi:10.3200/JOEE.37.1.43-52

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology, 80*(1), 16. doi:10.1037/0022-

0663.80.1.16

Ree, M. J., Carretta, T. R., & Teachout, M. S. (1995). Role of ability and prior knowledge in complex training performance. *Journal of Applied Psychology, 80*(6), 721-730. doi:10.1037/0021-9010.80.6.721

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction, 16*(2), 92-103. doi:10.1016/j.learninstruc.2006.02.008

Retelsdorf, J., & Köller, O. (2014). Reciprocal effects between reading comprehension and spelling. *Learning and Individual Differences, 30*, 77-83. doi:10.1016ij.lindif.2013.11 JJ07

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24*(1), 87-92. doi:10.1177/0956797612450883

Rittle-Johnson, B., & Koedinger, K. (2009). Iterating between lessons on concepts and procedures can improve mathematics knowledge. *British Journal of Educational Psychology, 79*(3), 483-500. doi:10.1348/000709908X398106

Sadi, Ö., & Cakiroglu, J. (2014). Relations of cognitive and motivational variables with students' human circulatory system achievement in traditional and learning cycle classrooms. *Educational Sciences: Theory & Practice, 14*(5), 1997-2012. doi:10.12738/estp.2014.5.2037

Salmeron, L., Kintsch, W., & Canas, J. J. (2006). Reading strategies and prior knowledge in learning from hypertext. *Memory & Cognition, 34*(5), 1157-1171. doi:10.3758/BF03193262

Samuelstuen, M. S., & Braten, I. (2005). Decoding, knowledge, and strategies in comprehension of expository text. *Scandinavian Journal of Psychology, 46*(2), 107-117. doi:10.1111/j.1467-9450.2005.00441.x

Sanchez, M., Magnan, A., & Ecalle, J. (2012). Knowledge about word structure in beginning readers: What specific links are there with word reading and spelling? *European Journal of Psychology of Education, 27*(3), 299-317. doi:10.1007/s10212-011-0071-8

Sander, E., & Heis, A. (2014). Interactive computer-supported learning in mathematics: A comparison of three learning programs on trigonometry. *Journal of Educational Computing Research, 50*(1), 45-65. doi:10.2190/EC.50.1.c

Schaap, L., Schmidt, H. G., & Verkoeijen, P. P. (2012). Assessing knowledge growth in a psychology curriculum: Which students improve most? *Assessment & Evaluation in Higher Education, 37*(7), 875-887. doi:10.1080/02602938.2011.581747

Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference? . *Learning and Individual Differences, 26*, 42-54. doi:10.1016/j.lindil".2013.04.003

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten Prediction of Reading Skills: A Longitudinal Comparative Analysis. *Journal of Educational Psychology, 96*(2), 265-282. doi:10.1037/0022-0663.96.2.265

Schmidt, H. K., Rothgangel, M., & Grube, D. (2015). Prior knowledge in recalling arguments in bioethical dilemmas. *Frontiers in Psychology, 6*, 1292. doi:10.3389/fpsyg.2015.01292

Schneider, W., & Korkel, J. (1989). The knowledge base and text recall: Evidence from a short-term longitudinal study. *Contemporary Educational Psychology, 14*(4), 382-393. doi:10.1016/0361-476X(89)90023-4

Schnotz, W. (1993). Adaptive construction of mental representations in understanding expository texts. *Contemporary Educational Psychology, 18*(1), 114-120. doi:10.1006/ceps.1993.1011

Schnotz, W., & Heis, A. (2009). Semantic scaffolds in hypermedia learning environments. *Computers in Human Behavior, 25*(2), 371-380. doi:10.1016/j.chb.2008.12.016

Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10*(4), 291-308. doi:10.1016/S1041-6080(99)80124-1

Schwarz, C. V., & White, B. Y. (2005). Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling. *Cognition and Instruction, 23*(2), 165-205. doi:10.1207/s1532690xci2302_1

Senechal, M., Ouellette, G., & Young, L. (2004). Testing the concurrent and predictive relations among articulation accuracy, speech perception, and phoneme awareness. *Journal of Experimental Child Psychology, 89*(3), 242-269. doi:10.1016/j.jecp.2004.07.005

Shapiro, A. M. (2004). How Including Prior Knowledge As a Subject Variable May Change Outcomes of Learning Research. *American Educational Research Journal, 41*(1), 159-189. doi:10.3102/00028312041001159

Shatil, E., Share, D. L., & Levin, I. (2000). On the contribution of kindergarten writing to grade 1 literacy: A longitudinal study in Hebrew. *Applied Psycholinguistics, 21*(1), 1-21. doi:10.1017/S0142716400001016

Shen, B., & Chen, A. (2006). Examining the interrelations among knowledge, interests, and learning strategies. *Journal of Teaching in Physical Education, 25*(2), 182-199.

Shintani, N. (2017). The effects of the timing of isolated FFI on the explicit knowledge and written accuracy of learners with different prior knowledge of the linguistic target *Studies in Second Language Acquisition, 39*(1), 129-166. doi:10.1017/S0272263116000127

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691-697. doi:10.1177/0956797612440101

Silverman, R. (2013). Investigating video as a means to promote vocabulary for at-risk children. *Contemporary Educational Psychology, 38*(3), 170-179. doi:10.1016/j.cedpsych.2013.03.001

Singh, L., Reznick, S. J., & Liang Xuehua. (2012). Infant word segmentation and childhood vocabulary development: A longitudinal analysis. *Developmental Science, 15*(4), 482-295. doi:10.1111/j.1467-7687.2012.01141.x

Snow, R. E., Tiffin, J., & Seibert, W. F. (1965). Individual differences and instructional film effects. *Journal of Educational Psychology, 56*(6), 315-326. doi:10.1037/h0022788

Sparks, R. L., Patton, J., & Murdoch, A. (2014). Early reading success and its relationship to reading achievement and reading volume: Replication of '10 years later'. *Reading and Writing, 27*(1), 189-211. doi:10.1007/s11145-013-9439-2

Stahl, S. A., Hare, V. C., Sinatra, R., & Gregory, J. F. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension: The retiring of number 4. *Journal of Reading Behavior*, *23*(4), 487-508. doi:10.1080/10862969109547755

Stahl, S. A., & Jacobson, M. G. (1986). Vocabulary difficulty, prior knowledge, and text comprehension. *Journal of Reading Behavior, 18*(4), 309-323. doi:10.1080/10862968609547578

Stahl, S. A., Jacobson, M. G., Davis, C. E., & Davis, R. L. (1989). Prior knowledge and difficult vocabulary in the comprehension of unfamiliar text. *Reading Research Quarterly, 24*(1), 27-43. doi:10.2307/748009

Stock, P., Desoete, A., & Roeyers, H. (2007). Early markers for arithmetic difficulties. *Educational and Child Psychology, 24*(2), 28-39.

Stock, P., Desoete, A., & Roeyers, H. (2009). Predicting arithmetic abilities: The role of preparatory arithmetic markers and intelligence. *Journal of Psychoeducational Assessment, 27*(3), 237-251. doi:10.1177/0734282908330587

Stuart, M. (1995). Prediction and qualitative assessment of five- and six-year-old children's reading: A longitudinal study. *British Journal of Educational Psychology, 65*(3), 287-296. doi:10.1111/j.2044-8279.1995.tb01150.x

Taboada, A., & Guthrie, J. T. (2006). Contributions of Student Questioning and Prior Knowledge to Construction of Knowledge From Reading Information Text. *Journal of Literacy Research, 38*(1), 1-35. doi:10.1207/s15548430jlr3801_1

Taft, M. L., & Leslie, L. (1985). The effects of prior knowledge and oral reading accuracy on miscues and comprehension. *Journal of Reading Behavior, 17*(2), 163-179. doi:10.1080/10862968509547537

Tarchi, C. (2010). Reading comprehension of informative texts in secondary school: A focus on direct and indirect effects of reader's prior knowledge. *Learning and Individual Differences, 20*(5), 415-420. doi:10.1016/j.Enctif. 2010.04.002

Tarchi, C. (2017). Comprehending expository texts: the role of cognitive and motivational factors. *Reading Psychology, 38*(2), 154-181. doi:10.1080/02702711.2016.1245229

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology, 30*(2), 96-101. doi:10.1207/S15328023TOP3002_02

Thompson, R. A., & Zamboanga, B. L. (2004). Academic Aptitude and Prior Knowledge as Predictors of Student Achievement in Introduction to Psychology. *Journal of Educational Psychology, 96*(4), 778-784. doi:10.1037/0022-0663.96.4.778

Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly, 62*(2), 179-206. doi:10.13110/merrpalmquar1982.62.2.0179

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading, 16*(2), 122-140.

Van Blankenstein, F. M., Dolmans, D. H. J. M., Van der Vleuten, C. P. M., & Schmidt, H. G. (2013). Relevant prior knowledge moderates the effect of elaboration during small group discussion on academic achievement. *Instructional Science, 41*(4), 729-744. doi:10.1007/s11251-012-9252-3

van Dijk, A. M., Eysink, T. H. S., & de Jong, T. (2016). Ability-related differences in performance of an inquiry task: The added value of prompts. *Learning and Individual Differences, 47*(2016), 145-155. doi:10.1016/j.lindif.2016.01.008

Van Joolingen, W. R., Aukes, A. V. A., Gijlers, H., & Bollen, L. (2015). Understanding elementary astronomy by making drawing-based models. *Journal of Science Education and Technology, 24*(2-3), 256-264. doi:10.1007/s10956-014-9540-6

Verhagen, W., Aarnoutse, C., & Van Leeuwe, J. (2008). Phonological awareness and naming speed in the prediction of Dutch children's word recognition. *Scientific Studies of Reading, 12*(4), 301-324. doi:10.1080/10888430802132030

Verhagen, W., Aarnoutse, C., & Van Leeuwe, J. (2009). The predictive power of phonemic awareness and naming speed for early Dutch word recognition. *Educational Research and Evaluation, 15*(1), 93-116. doi:10.1080/13803610802663912

Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development, 85*(4), 1461-1476. doi:10.1111/cdev.12218

Walcott, C. M., Scheemaker, A., & Bielski, K. (2010). A longitudinal investigation of inattention and preliteracy development. *Journal of Attention Disorders, 14*(1), 79-85. doi:10.1177/1087054709333330

Waniek, J., & Schafer, T. (2009). The role of domain and system knowledge on text comprehension and information search in hypermedia. *Journal of Educational Multimedia and Hypermedia, 18*(2), 221-240. Retrieved February 16, 2018 from https://www.learntechlib.org/p/28259/.

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*(6), 1892-1907. doi:10.1111/cdev.12416

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352-360. doi:10.3102/0013189X14553660

Webb, S., & Chang, A. C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition, 37*(04), 1-25 (651-675). doi:10.1017/S0272263114000606

Wecker, C., Rachel, A., Heran-Dorr, E., Waltner, C., Wiesner, H., & Fischer, F. (2013). Presenting theoretical ideas prior to inquiry activities fosters theory-level knowledge. *Journal of Research in Science Teaching, 50*(10), 1180-1206. doi:10.1002/tea.21 106

Weinert, F. E., & Helmke, A. (1998). The neglected role of individual differences in theoretical models of cognitive development. *Learning and Instruction, 8*(4), 309-323. doi:10.1016/S0959-4752(97)00024-8

Wetzels, S. A. J., Kester, L., Van Merrienboer, J. J. G., & Broers, N. J. (2011). The influence of prior knowledge on the retrieval-directed function of note taking in prior knowledge activation. *British Educational Research Journal, 81*(2), 274-291. doi:10.1348/000709910X517425

Wilkins, J. L. M., & Ma, X. (2002). Predicting student growth in mathematical content knowledge. *The Journal of Educational Research, 95*(5), 288-298. doi:10.1080/00220670209596602

Winters, F. I., & Azevedo, R. (2005). High-school students' regulation of learning during computer-based science inquiry. *Journal of Educational Computing Research, 33*(2), 189-217. doi:10.2190/F7HM-9JN5-JUX8-4BM9

Wolfe, M. B. W., & Mienko, J. A. (2007). Learning and memory of factual content from narrative and expository text. *British Journal of Educational Psychology, 77*(3), 541-564. doi:10.1348/000709906X143902

Wolfe, M. B. W., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology, 80*(3), 341-362. doi:10.1348/000709910X485700

Wu, X., Lowyck, J., Sercu, L., & Elen, J. (2013). Task complexity, student perceptions of vocabulary learning in EFL, and task performance. *British Journal of Educational Psychology, 83*(1), 160-181. doi:10.1111/j.2044-8279.2011.02061.x

Yeung, P.-s., Ho, C. S.-h., Chan, D. W.-o., & Chung, K. K.-h. (2013). Modeling the relationships between cognitive-linguistic skills and writing in Chinese among elementary grades students. *Reading and Writing, 26*(7), 1195-1221. doi:10.1007/s11145-012-9411-6

Yeung, P.-s., Ho, C. S.-H., Chan, D. W.-o., & Chung, K. K.-h. (2016). Orthographic skills important to Chinese literacy development: the role of radical representation and orthographic memory of radicals. *Reading and Writing, 29*(9), 1935-1958. doi:10.1007/s11145-016-9659-3

Zacharia, Z. C., Loizou, E., & Papaevripidou, M. (2012). Is physicality an important aspect of learning through science experimentation among kindergarten students? *Early Childhood Research Quartely, 27*(3), 447-457. doi:10. 1O16/j.ccrcsq.20 12.02.004

Zhang, J., McBride-Chang, C., Wagner, R. K., & Chan, S. (2014). Uniqueness and overlap: Characteristics and longitudinal correlates of native Chinese children's writing in

English as a foreign language. *Bilingualism: Language and Cognition, 17*(2), 347-363. doi:10.1017/S1366728913000163.

Zhang, X., Koponen, T., Rasanen, P., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development, 85*(3), 1091-1107. doi:10.1177/0956797613519111

Zhang, Y., Tardif, T., Shu, H., Li, H., Liu, H., McBride-Chang, C., Zhang, Z. (2013). Phonological skills and vocabulary knowledge mediate socioeconomic status effects in predicting reading outcomes for Chinese children. *Developmental Psychology, 49*(4), 665-671. doi:10.1037/a0028612

Zhu, L., & Grabowski, B. L. (2006). Web-Based Animation or Static Graphics: Is the Extra Cost of Animation Worth It? *Journal of Educational Multimedia and Hypermedia, 15*(3), 329-347.

Zumbach, J., Schmitt, S., Reimann, P., & Starkloff, P. (2006). Learning life sciences: Design and development of a virtual molecular biology learning lab. *Journal of Computers in Mathematics and Science Teaching, 25*(3), 281-300.

SM 2: Number of coded studies and effect sizes on the levels of each moderator

Table S 1

Number of coded studies (j) and effect sizes (k) on the levels of each moderator, separate for the correlations of prior knowledge with posttest knowledge, absolute knowledge gains, and normalized knowledge gains

	Correlation with posttest knowledge		Correlation with absolute knowledge gains		Correlation with normalized knowledge gains	
	j	*k*	*j*	*k*	*j*	*k*
Knowledge characteristics						
Knowledge type[a]						
Mixed	74	645	5	16	3	7
Declarative	159	2125	14	40	9	19
Procedural	10	51	2	8	1	6
Knowledge subtype[a]						
Declarative: facts	41	268	5	11	2	4
Declarative: concepts	128	1244	13	22	8	15
Procedural: cognitive skill	10	51	2	8	1	6
Broad content area[a]						
STEM	87	578	13	40	9	23
Language	99	3055	4	18	2	6
Humanities	5	19	0	0	0	0
Social sciences	17	107	3	4	2	3
Health sciences	2	9	0	0	0	0

Sports	5	54	0	0	1	1
Content domain[a]						
Mathematics	37	416	5	21	3	11
Physics	16	55	2	3	2	3
Chemistry	6	20	0	0	0	0
Biology	21	62	4	10	4	9
Geosciences	6	19	2	6	0	0
Computer sciences	1	1	0	0	0	0
Medicine and nursing	2	9	0	0	0	0
Psychology	14	99	3	4	2	3
First language	94	2715	1	10	0	0
Second language	10	53	3	8	2	6
Sports	5	54	1	1	1	1
History	3	10	0	0	0	0
Other	3	20	0	0	0	0
Similarity of prior knowledge and learning outcome						
Similarity of the knowledge type						
Low	117	1398	0	0	1	1
High	210	2821	20	62	13	32
Similarity of the content domain						
Low	72	688	0	0	0	0

High	209	3535	20	62	14	33
Similarity of the physical context						
Low	7	94	0	0	0	0
High	228	4071	20	62	14	33
Similarity of the temporal context						
Low	184	3974	14	46	10	23
High	55	248	6	16	4	10
Time between assessments	171	3908	16	53	8	21
Similarity of the functional context						
Low	4	12	1	1	1	1
High	230	4204	19	61	13	32
Similarity of the social context						
Low	15	185	0	0	1	1
High	222	3909	19	56	13	32
Similarity of the modality						
Low	46	650	0	0	1	1
High	205	3064	18	55	13	32
Learner characteristics						
Age	136	3362	8	30	5	15
Educational level						

Daycare	5	139	0	0	0	0
Kindergarten/preschool	57	1790	1	10	1	1
Primary education	71	1676	4	15	3	11
Secondary education	49	228	4	10	4	5
Higher education	57	358	7	16	4	11
Continued education	5	22	3	9	2	5
Several	3	10	1	2	0	0
Environmental characteristics						
Country[b]						
Australia	3	29	0	0	0	0
Austria	1	30	0	0	0	0
Belgium	2	80	0	0	0	0
Canada	5	115	0	0	0	0
Peoples Republic of China	5	300	0	0	0	0
Hong Kong	6	168	0	0	0	0
Taiwan	3	9	0	0	0	0
Denmark	2	7	0	0	0	0
Finland	8	322	0	0	0	0
France	2	122	0	0	0	0
Germany	9	253	0	0	0	0
Greece	1	27	0	0	0	0
Israel	4	49	0	0	0	0

Italy	1	36	0	0	0	0
Korea	1	45	0	0	0	0
Netherlands	5	163	1	1	0	0
New Zealand	2	42	0	0	0	0
Norway	3	172	0	0	0	0
Singapore	1	8	0	0	0	0
Spain	2	5	0	0	0	0
Sweden	1	12	0	0	0	0
Turkey	1	2	0	0	0	0
United Kingdom	12	627	0	0	0	0
USA	52	904	1	6	0	0
Intervention Setting						
No intervention	14	150	0	0	0	0
School instruction only	78	2772	2	7	0	0
School instruction and other intervention	57	867	5	20	3	5
Other intervention only	85	397	12	33	11	28
Intervention Duration						
0–2 hours	73	352	10	29	9	26
2–24 hours	7	29	1	4	0	0
2–7 days	5	16	0	0	0	0
> 1 week	57	889	7	24	5	7
Cognitive demands of						

intervention						
Lower	39	100	8	13	8	12
Higher	38	96	7	11	6	10
Instructional methods in intervention						
Written instruction						
No	23	226	3	9	4	10
Yes	101	463	13	33	9	21
Oral instruction						
No	85	495	11	30	8	20
Yes	36	253	4	12	4	11
Multimedia instruction						
No	70	489	5	19	4	14
Yes	43	149	9	21	7	15
Practice						
No	75	440	7	18	4	10
Yes	47	302	8	23	9	21
Constructive activities						
No	81	501	9	27	7	23
Yes	45	158	7	14	6	7
Technology						
No	78	521	5	19	5	18
Yes	41	127	9	20	6	10

Feedback						
No	94	561	9	24	7	17
Yes	18	65	6	15	5	11
Collaborative learning						
No	98	569	13	38	10	27
Yes	22	87	2	3	3	3
Problem-based learning						
No	96	583	14	39	11	28
Yes	17	93	3	3	2	2
Methodological study characteristics						
Study design						
Group differences in prior knowledge (quasi-experimental or experimental design)	83	403	15	39	13	32
Individual differences in prior knowledge (correlational design)	152	3820	5	23	1	1
Randomized controlled trial						
No	226	4179	20	62	14	33
Yes	9	44	0	0	0	0
Number items T1	186	3053	15	44	12	30
Number items T2	189	3088	14	44	11	29
Response format[c]						

Open	101	1368	6	10	3	5
Fill-in	7	34	2	8	2	8
Single or multiple choice	66	508	7	16	4	9
Rating	1	2	0	0	0	0
Behaviour	1	1	0	0	0	0
Other	6	30	0	0	0	0
Various	18	99	3	9	3	9
Retention test						
No	235	4159	20	62	14	33
Yes	13	64	0	0	0	0
Same test for prior knowledge and learning outcome						
No	201	3606	5	17	2	4
Yes	104	615	17	45	12	29
Measures at T2						
Outcome						
Knowledge	228	4169	19	61	13	32
Achievement	10	54	1	1	1	1
Domain specificity						
Specific domain	232	4179	20	62	14	33
Various domains	8	44	0	0	0	0

[a] Only studies assessing the same specific knowledge characteristic at T1 and T2.

[b] Only coded for studies including school instruction.

[c] Only studies using the same response format at T1 and T2.

SM 3: Equations of statistical analyses

Computation of effect sizes

Cohen's *d* for posttest knowledge. To calculate Cohen's *d* for group comparisons at T2, we used the following formula (Cumming, 2012):

$$d_p = \frac{X_{HPK,T2} - X_{LPK,T2}}{\frac{SD^2_{HPK,T2} + SD^2_{LPK,T2}}{2}}, \tag{1}$$

in which the numerator is the difference between the means of the high prior knowledge group (HPK) and the low prior knowledge group (LPK), and the denominator is the pooled standard deviation for the means of these groups.

Cohen's *d* for knowledge gains. For the calculation of Cohen's *d* for knowledge gains, we calculated the change of knowledge across time as pretest-adjusted posttest gain difference effect sizes (absolute gain scores) and pretest-adjusted posttest standardized gain difference effect sizes (normalized gain scores). We calculated the pretest-adjusted posttest gain difference effect sizes as follows:

$$d_{AG} = \left[\frac{(X_{HPK,T2} - X_{HPK,T1}) - (X_{LPK,T2} - X_{LPK,T1})}{SD_{pooled,T1,gain}}\right], \tag{2}$$

in which the numerator is the difference in absolute gains for the high prior knowledge group, and the low prior knowledge group and the denominator represent the pooled standard deviation for the pretest scores in those groups (using the pooled pretest standard deviation).

We calculated the pretest-adjusted posttest standardized gain difference effect sizes as follows (derived from Morris, 2008, p. 369):

$$d_{NG} = \left[\frac{G_{NG,HPK} - G_{NG,LPK}}{SD_{pooled,T1,gain}}\right], \tag{3}$$

in which the numerator is the difference in normalized gains for the high prior knowledge group, and the low prior knowledge group and the denominator represent the pooled standard deviation for the pretest scores in those groups (using the pooled pretest standard deviation). The normalized difference in gains for each group can be obtained using:

$$G_{NG} = \frac{M_{T2} - M_{T1}}{G_{max}} \tag{4}$$

with

$$G_{max} = M_{max} - M_{T1}\ , \tag{5}$$

in which the nominator is the difference between the obtained scores at the first and second measurement point, and the denominator is the maximal possible gain calculated by the difference of the highest possible score of the test and the mean at T1.

The standard deviations of the normalized gain scores can be calculated as (derived from

Morris, 2008, Formula 9) follows:

$$SD_{T1,gain} = \sqrt{\frac{(n_{HPK}-1)\left(\frac{SD_{T1,HPK}}{G_{max,HPK}}\right)^2+(n_{LPK}-1)\left(\frac{SD_{T1,LPK}}{G_{max,LPK}}\right)^2}{n_{HPK}+n_{LPK}-2}}. \tag{6}$$

We transformed all Cohen's *d*s to *r*s using the following formula (Borenstein, Hedges, Higgins, & Rothenstein, 2009, p. 48):

$$r = \frac{d}{\sqrt{d^2+a}} \tag{7}$$

with

$$a = \frac{(n_{HPK}+ n_{LPK})^2}{n_{HPK}\times n_{LPK}}. \tag{8}$$

Meta-analytic model

To estimate the overall strength of the correlation between prior knowledge and learning outcomes, we estimated a simple RVE meta-regression model:

$$y_{ij} = \beta_0 + u_j + e_{ij}, \tag{9}$$

where y_{ij} is the *i*th correlation effect size in the *j*th study, β_0 is the average population effect of the correlation, u_j is the study level random effect such that Var(u_j) = τ^2 is the between-study variance component, and e_{ij} is the residual for the *i*th effect size in the *j*th study. To estimate the variability in the effect size due to moderator variables, we estimated a mixed-effects RVE meta-regression model:

$$y_{ij} = \beta_0 + \beta_1(Moderator_1)_{ij} + \cdots + \beta_k(Moderator_k)_{ij} + u_j + e_{ij}. \tag{10}$$

In this model, each moderator represents a continuous variable or specific dummy coded level of an included moderator variable (e.g., age as a continuous predictor or the two levels CT or longitudinal as predictors pertaining to the moderator study design). Continuous moderator variables were log-transformed to obtain normal distribution. We set the parameter ρ, that is, the assumed intercorrelation of dependent effect sizes, to the default value of .80. Sensitivity analyses revealed that the results remained the same across different values of ρ (0–1).

To estimate R^2 for the multiple regression models, we calculated the sum across the product of the correlation between the dependent variable with the independent variable and the respective standardized regression coefficient (Tabachnick & Fidell, 2014, p. 167, Fomula 4):

$$R^2 = \sum_{i=1}^{k} r_{yi}\beta_i. \tag{11}$$

We calculated the standardized regression coefficients from the unstandardized regression coefficients reported by the robumeta package with (Bring, 1994, pp. 210, formula 2.3):

$$\beta_i = b_i \times \frac{SD\,(Moderator_i)}{SD\,(r^+)}. \tag{12}$$

SM 4: Mean Effect Sizes and Moderator Analyses for the Correlations of Prior Knowledge with Absolute and Normalized Knowledge Gains

Table S 2

Mean Effect Sizes and Moderator Analyses for the Correlations of Prior Knowledge with Absolute and Normalized Knowledge Gains

	Correlation with absolute knowledge gains								Correlation with normalized knowledge gains							
	j	k	r_{AG}^{+}	CI r_{AG}^{+} 95%	τ^2	I^2	Moderator		j	k	r_{NG}^{+}	CI r_{NG}^{+} 95%	τ^2	I^2	Moderator	
							Sign.	R^2							Sign.	R^2
Overall	20	62	-.210	[-.437, .043]	.380	96.99			14	33	-.083	[-.478, .342]	.580	94.83		
Knowledge characteristics																
Similarity of prior knowledge and learning outcome																
Similarity of the temporal context							*ns*	.000							-	-

Low	14	46	-.217	[-.503, .113]	.416	97.83	Ref		10	23	-.170	[-.635, .384]	.381	93.39	-	
High	6	16	-.167	[-.597, .336]	.228	81.11	*ns*		4	10	.220	-	-	-	-	
Time between assessments	16	53					*ns*	.000	8	21					*ns*	.135
Learner characteristics																
Age	8	30					*ns*	.036	5	15					*ns*	.244
Environmental characteristics																
Intervention																
Intervention Duration							*ns*	.016							-	-
0–2 hours	10	29	-.181	[-.489, .168]	.237	86.38	*ns*		9	26	-.069	[-.445, .328]	.326	90.96	-	
> 1 week	7	21	-.359	[-.774, .277]	.405	94.27	Ref		5	7	-.057	-	-	-	-	
Cognitive demands of							*	.281							*	.366

intervention																
Lower	8	13	-.582	[-.851, -.068]	.502	92.94	Ref		8	12	-.478	[-.804, .064]	.451	93.90	Ref	
Higher	7	11	.106	[-.207, .399]	.086	67.34	*		6	10	.252	[-.130, .569]	.081	81.91	*	
Instructional methods in intervention																
Practice							*ns*	.017							-	-
No	7	18	-.380	[-.786, .250]	.639	93.69	Ref		4	10	-.513	-	-	-	-	
Yes	8	23	-.177	[-.523, .219]	.230	86.08	*ns*		9	21	-.078	[-.481, .352]	.326	90.99	-	
Constructive activities							*ns*	.000								.046
No	9	27	-.287	[-.633, .155]	.387	87.81	*ns*		7	23	.038	[-.461, .518]	.488	89.99	*ns*	
Yes	7	14	-.297	[-.708,	.413	92.68	Ref		6	7	-.350	[-.789,	.442	93.36	Ref	

				.265]								.324]				
Feedback							*ns*	.011							-	-
No	9	24	-.338	[-.686, .135]	.569	92.61	Ref		7	17	-.344	[-.746, .225]	.514	92.65		
Yes	6	15	-.270	[-.778, .449]	.374	88.41	*ns*		5	11	.027	-	-	-	-	
Multivariate effect of all instructional methods	14	39					*ns*	.018								
Methodological study characteristics																
Number of items in prior knowledge measure	15	44					*ns*	.021	12	30					*ns*	.106
Number of items in learning outcome measure	14	44					*ns*	.020	11	29					*ns*	.117
Response format[a]							*ns*	.031							-	-
Open	6	10	-.259	[-.750, .418]	.327	94.44	*ns*		3	5	-.639	-	-	-	-	

Fill-in	2	8	-.281	-	-	-	-		2	8	.446	-	-	-	-	
Single or multiple choice	7	16	-.478	[-.766, -.029]	.247	79.89	Ref		4	9	-.263	-	-	-	-	
Various	3	9	-.048	-	-	-	-		3	9	.310	-	-	-	-	
Same test for prior knowledge and learning outcome							*	.110							-	-
No	5	17	.208	[.038, .367]	.513	93.35	**		2	4	-.345	-	-	-	-	
Yes	17	45	-.318	[-.549, -.042]	.267	90.77	Ref		12	29	-.029	[-.498, .453]	.749	94.96		

[a] Only for effects in which the moderator had the same level at T1 and T2.

[b] Only coded for studies including school instruction.

Note. * $p < .05$, ** $p < .01$, *ns* = nonsignificant. Cells without results indicate that the *df* were too small for valid interpretation. Moderation analyses were only calculated for the levels of the moderating variables where the confidence intervals are presented, respectively

Appendix B: Supplementary Materials of Study 2

SM 5: Latent Transition Probabilities Based On the Estimated Model

Table S 3

Latent Transition Probabilities Based On the Estimated Model: T1 Classes (Rows) by T2 Classes (Columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	.709	.000	.070	.221
C2	Fragmented profile	.000	.389	.611	.000
C3	Indecisive profile	.000	.000	.319	.681
C4	Scientific profile	.000	.000	.000	1.00

Table S 4

Latent Transition Probabilities Based On the Estimated Model: T2 Classes (Rows) by T3 Classes (Columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	1.00	.000	.000	.000
C2	Fragmented profile	.049	.786	.165	.000
C3	Indecisive profile	.000	.000	.856	.144

C4	Scientific profile	.000	.000	.000	1.00

Table S 5

Latent Transition Probabilities Based On the Estimated Model: T3 Classes (Rows) by T4 Classes (Columns)

Class		C1	C2	C3	C4
C1	Misconceptions profile	.630	.000	.037	.000
C2	Fragmented profile	.000	.867	.133	.000
C3	Indecisive profile	.000	.039	.765	.196
C4	Scientific profile	.000	.000	.000	1.00

SM 6: Example Task „Chunking“:

Wild horses

Two groups of children participate in the study.

The children in one of these groups (Group A) are twelve year-olds, who know little about horses in general.

The children in the other group (Group B) are eight year-olds, who know a lot about horses in general.

The two groups do not differ in terms of the children's intelligence and the number of boys and girls in the group.

All participants read a simple text about wild horses that teaches them what kinds of species exist, where they are found and what they need to live. Afterwards, the children are asked to answer questions assessing their knowledge and comprehension, based on the text from memory.

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect Definitely correct
Group A will perform markedly better than Group B because they have four more years of practice in reading and remembering from texts than the younger children.	O O O O O O O
Group B will perform markedly worse than Group A, because the children are on a lower stage of cognitive development and therefore cannot process information so well.	O O O O O O O
Group B will perform markedly better than Group A because they have more prior knowledge and therefore can store information from the text in a more structured way.	O O O O O O O
Group A will perform markedly worse than Group B because the memory of twelve year-olds is partly impaired already due to hormonal changes during puberty.	O O O O O O O
Both groups will perform almost equally well because they have read	O O O O O O O

the same text with the result that each person has memorized the same information.	
There barely will be any difference between the two groups because their intelligence is similar and therefore they can process information almost equally well.	O O O O O O O

SM 7: Example Task „Interference“:

Font Sizes

In one study the participants are presented with 50 pairs of simple numbers successively on a computer screen, for example 4 and 9. One number is always presented in a smaller font size and the other one in a larger font size. The participants' task for each pair of numbers is to answer as quickly as possible whether the left or the right number is presented in a larger font size.

For half of the participants (Group A) the number with the larger numeric value (e.g. 9) is always presented in a larger font size and the number with the smaller numeric value (e.g. 4) in a smaller font size.

For the other half of the participants (Group B) the number with the smaller numeric value (e.g. 4) is always presented in a larger font size and the number with the larger numeric value (e.g. 9) in a smaller font size.

It is analyzed how many milliseconds participants in Group A and Group B need on average to evaluate which one of the two numbers is presented in a larger font size.

How sure are you that each of the following statements that is correct or incorrect?

	Definitely incorrect ... Definitely correct
Group A will be markedly faster than Group B because the numeric values and font sizes in Group B provide the brain with contradicting information.	O O O O O O O
Group B will be markedly slower than Group A because people are used to larger numeric values being presented in larger font sizes than smaller numeric values in everyday life.	O O O O O O O

Group B will be markedly faster than Group A because it is more distinct when numeric values and font sizes do not match and therefore it is remembered more easily.	O O O O O O O
Group A will be markedly slower than Group B because the task of Group B is more unusual and therefore more interesting and motivating.	O O O O O O O
Both groups will perform almost equally fast because the brain processes numeric values and font sizes independently.	O O O O O O O
There barely will be any difference between the two groups because the task is about font sizes only and the participants will pay no attention to the numbers.	O O O O O O O

SM 8: Example Task „Source Monitoring“:

Europe

Participants in a study know little about the European economic system. They read a text on the topic “Should all Europeans be allowed to live and work anywhere in the European Union without restrictions?” Participants are told that the author of the text is a participant in a casting show for singing talents and has little expertise in politics. The arguments in the text are partly true and partly incorrect.

The participants rate right after reading the text, how convincing the arguments are (Test 1). One month later, they are presented with the arguments a second time without announcement and again, they are asked again how convincing they find the arguments (Test 2).

How sure are you that each of the following statements is correct or incorrect?

	Definitely incorrect — Definitely correct
The arguments will be judged to be clearly more convincing in Test 1 than in Test 2 because they will be more available in memory during Test 1.	O O O O O O O
The arguments will be judged to be clearly less convincing in Test 2	O O O O O O O

than in Test 1 because the participants will be less impressed by the author's fame.	
The arguments will be judged to be clearly more convincing in Test 2 because participants will have spent more time thinking about the topic and will assume a more positive attitude.	o o o o o o o
The arguments in Test 1 will be judged to be clearly less convincing than in Test 2 because participants will be more likely to remember in Test 1 that the arguments were from an author with little expertise in politics.	o o o o o o o
The arguments are judged in a comparable way in both tests because the arguments did not change in the meantime.	o o o o o o o
The judgment of the arguments will barely differ between the two tests because attitudes towards political topics are embedded deeply in memory.	o o o o o o o

Appendix C: Supplementary Materials of Study 3

SM 9: Assessment of Student Resources in Study 3a

Table S 6

Item Texts, Means (standard deviations in parentheses) of the Items Measuring Student Resources in Study 3a

	Study 3a	
	T1	T2
Social Family-Related Resources		
When I have trouble or worries, I can always turn to my family.	4.140 (1.125)	4.240 (1.097)
My family does not support my university education.*	4.310 (1.076)	4.370 (1.032)
I think my parents are interested in my university program.	3.750 (1.137)	3.720 (1.112)
The separation from my family bothers me.*	3.690 (1.169)	3.760 (1.135)
I don't miss my family in my everyday life.	3.120 (1.250)	3.060 (1.252)
I feel uprooted since I have moved from my home town.*	4.070 (1.115)	3.970 (1.149)
Social Peer-Related Resources		
My fellow students and I support each other in our studies.	3.900 (1.032)	4.010 (1.024)
There is a lot of competition among my fellow students.*	3.380 (1.278)	3.400 (1.254)

I discuss difficult learning content with my fellow students, even beyond classes.	3.670 (1.123)	3.680 (1.162)
Some of my fellow students would support me even if I had troubles in other areas of life than university.	4.100 (1.030)	4.270 (.916)
I couldn't ask my fellow students to help me move (to another residence).*	4.280 (1.106)	4.400 (1.023)
I can rely upon my fellow students when I need help in everyday life issues.	3.960 (1.102)	4.080 (1.060)
I miss my friends from school a lot.*	3.390 (1.200)	3.360 (1.172)
It doesn't bother me that I don't see my friends from school as often during the semester.	3.310 (1.251)	3.230 (1.250)
I would like to see my friends from school more often than it is possible due to my studies.*	2.860 (1.262)	2.700 (1.280)
Interest-Related Resources		
I am deeply interested in the jobs for which my university program prepares me.	3.830 (.980)	3.760 (.965)
I am losing interest in the jobs for which my university program prepares me. *	4.210 (.955)	4.070 (1.006)
The jobs for which my university program prepares me barely have any boring aspects.	3.560 (1.008)	3.480 (.973)
The contents of my university program interest me a lot.	3.900 (.928)	3.760 (.963)
A lot of the contents of my university program are boring.*	3.430 (1.011)	3.340 (1.078)

A lot of my classes are exciting.	3.670 (.965)	3.630 (.983)
My university program is my choice.	4.120 (1.050)	4.070 (1.076)
I'd rather take classes in another discipline.*	4.020 (1.186)	4.010 (1.226)
From the beginning, I haven't been assured that I chose the right university program.*	3.660 (1.311)	3.670 (1.273)
Competence-Related Resources		
The demands that my program places upon me are as I expected.	3.240 (1.093)	3.240 (1.051)
The contents of my program have not fit my expectations from the beginning.*	3.190 (1.146)	3.200 (1.106)
My program is about as difficult as expected.	3.190 (1.094)	3.190 (1.062)
I am sure that I will successfully graduate.	4.070 (.840)	4.170 (.958)
I will probably fail my university program.*	4.260 (.841)	4.290 (.902)
I expect that I will earn good grades at university.	3.630 (.965)	3.630 (1.074)
Context-Related Resources		
I'm satisfied with my current financial situation.	2.950 (1.335)	3.160 (1.343)
I'm afraid that my financial situation will be difficult after graduating from university.*	3.450 (1.203)	3.400 (1.278)
Currently, I don't have money troubles.	3.220 (1.363)	3.460 (1.356)
I commute to university on a daily basis.*	4.400 (1.293)	4.460 (1.264)
I live in my university town.	4.240 (1.441)	4.300 (1.384)

I have stayed at my university town for the last eight weeks.	2.510 (1.637)	2.410 (1.615)
Hours of work per week*	34.041 (7.209)	32.911(7.525)

*reverse coded (higher scores indicate higher level of resources).

SM 10: Assessment of Demands in Study 3b

Table S 7

Assessment of Demands in Study 3b

Academic demands

1. Time pressure
2. Long study time
3. Monotony, mental underload
4. Perfectionism
5. Mental overload
6. Frequent disturbances during classes
7. Little interest in classes
8. Lapse of concentration
9. Conflicts with other students, teachers etc.
10. Having to travel a long distance to university
11. Other demands

Demands during leisure time

1. Not enough leisure time
2. I can't enjoy my leisure time
3. Too much leisure time
4. Not enough sleep
5. Boredom
6. Sleeping too much
7. Organization of one's leisure time on weekends is hard
8. Daily errands
9. Anxiety interferes with my leisure time activities
10. Going to administrative offices
11. Demanding extracurricular work
12. Lack of ideas and opportunities for leisure activities
13. Financial difficulties
14. Caring for loved ones / children
15. Other demands

Social demands

1. Too many social activities
2. Frequent feelings of loneliness
3. Not enough social activities
4. Social insecurity
5. Conflicts with family members
6. Problems in romantic relationships
7. Other demands
8. Frequent conflicts with others

Health-related demands

1. Pain
2. Underweight
3. Frequent infections
4. Irregular meals
5. Chronic Diseases (Diabetes etc.)
6. Sleeping disorders
7. Other demands
8. Overweight

Authorship and Publication Status

Chapters 8 to 10 present research articles that are or will soon be submitted for publication. Hereinafter the authors and publications statuses of the articles are enlisted.

1. Study 1 (Chapter 8): Simonsmeier, B. A., **Flaig, M.**, Deiglmayr, A., Schalk, L., & Schneider, M. (manuscript in preparation). Domain-specific prior knowledge and learning: A meta-analysis.
2. Study 2 (Chapter 9): **Flaig, M.**, Simonsmeier, B. A., Mayer, A.-K., Rosman, T., Gorges, J., & Schneider, M. (2018). Conceptual change and knowledge integration as learning processes in higher education: A latent transition analysis. *Learning and Individual Differences, 62*, 49-61.
3. Study 3 (Chapter 10): **Flaig, M.**, Domes, G., & Schneider, M. (manuscript in preparation). State and Trait Differentially Predict Health and Achievement in Higher Education Students.

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. This dissertation was not previously presented to another examination board in order to obtain an academic title.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass die vorliegende Dissertationsschrift von mir selbständig angefertigt wurde und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet wurden. Zudem wurde die Arbeit an keiner anderen Universität zur Erlangung eines akademischen Grades eingereicht.

Trier, ____________________________ ___________________________________

Date/ Datum Maja Sophie Flaig