ACADEMIC BOREDOM DEVELOPMENT IN EARLY SECONDARY EDUCATION

by

Markus Daniel Feuchter

A Dissertation Submitted to

Faculty I - Department of Psychology

University of Trier

In Partial Fulfillment of the Requirements

for the Degree Dr. rer. nat. (doctor rerum naturalium)

Submitted at University of Trier: March 2023

Defended at University of Trier: July 2023

Publication at University of Trier: September 2023

Supervisors:

Prof. Dr. Franzis Preckel

Prof. Dr. Roland Neumann

Location of Dissertation: Trier

MANY FACES OF BOREDOM – A COLLAGE[1]

[1] For individual picture sources, see Appendix 1

Danksagung

Zunächst möchte ich mich herzlich bei Frau Prof. Dr. Preckel für die fachliche Betreuung und Begutachtung meiner Dissertation, ihre professionelle Anleitung als Abteilungsleiterin sowie ihre unterstützenden Beiträge als Co-Autorin im Zuge meiner wissenschaftlichen Projekte bedanken.

Des Weiteren danke ich Herrn Prof. Dr. Neumann vielmals für die Begutachtung meiner Dissertation in der Rolle des Zweitbetreuers.

Ebenfalls ausdrücklich bedanken möchte ich mich bei meinen lieben Kolleg*innen der Abteilung Hochbegabtenforschung und -förderung, insbesondere bei Sevo, Moritz, Julian und Julia für den fachlichen Austausch und viele tolle Momente abseits des Unialltags.

Ein besonderer Dank gebührt meiner Freundin, Eli, die mich zu jeder Zeit fachlich wie persönlich unterstützt und auffängt. Neben privatem Glück fand ich in ihr ebenfalls ein großes wissenschaftliches Vorbild.

Zuletzt bedanke ich mich herzlich bei meiner Familie, insbesondere bei meinen Eltern, Petra und Bernhard, für die bedingungslose und unentwegte Unterstützung meines beruflichen und privaten Werdegangs.

Euch und Ihnen allen: Vielen lieben Dank!

Abstract

The following dissertation contains three studies examining academic boredom development in five high-track German secondary schools (AVG-project data; Study 1: *N* = 1,432; Study 2: *N* = 1,861; Study 3: *N* = 1,428). The investigation period spanned 3.5 years, with four waves of measurement from grades 5 to 8 (T1: 5th grade, after transition to secondary school; T2: 5th grade, after mid-term evaluations; T3: 6th grade, after mid-term evaluations; T4: 8th grade, after mid-term evaluations). All three studies featured cross-sectional and longitudinal analyses, separating, and comparing the subject domains of mathematics and German.

Study 1 provided an investigation of academic boredom's factorial structure alongside correlational and reciprocal relations of different forms of boredom and academic self-concept. Analyses included reciprocal effects models and latent correlation analyses. Results indicated separability of boredom intensity, boredom due to underchallenge and boredom due to overchallenge, as separate, correlated factors. Evidence for reciprocal relations between boredom and academic self-concept was limited.

Study 2 examined the effectiveness and efficacy of full-time ability grouping for as a boredom intervention directed at the intellectually gifted. Analyses included propensity score matching, and latent growth curve modelling. Results pointed to limited effectiveness and efficacy for full-time ability grouping regarding boredom reduction.

Study 3 explored gender differences in academic boredom development, mediated by academic interest, academic self-concept, and previous academic achievement. Analyses included measurement invariance testing, and multiple-indicator-multi-cause-models. Results showed one-sided gender differences, with boys reporting less favorable boredom development compared to girls, even beyond the inclusion of relevant mediators.

Findings from all three studies were embedded into the theoretical framework of control-value theory (Pekrun, 2006; 2019; Pekrun et al., 2023). Limitations, directions for future research, and practical implications were acknowledged and discussed.

Overall, this dissertation yielded important insights into boredom's conceptual complexity. This concerned factorial structure, developmental trajectories, interrelations to other learning variables, individual differences, and domain specificities.

Keywords: Academic boredom, boredom intensity, boredom due to underchallenge, boredom due to overchallenge, ability grouping, gender differences, longitudinal data analysis, control-value theory

List of Tables

List of Figures

Table of Contents

Chapter 1 – Topical Introduction

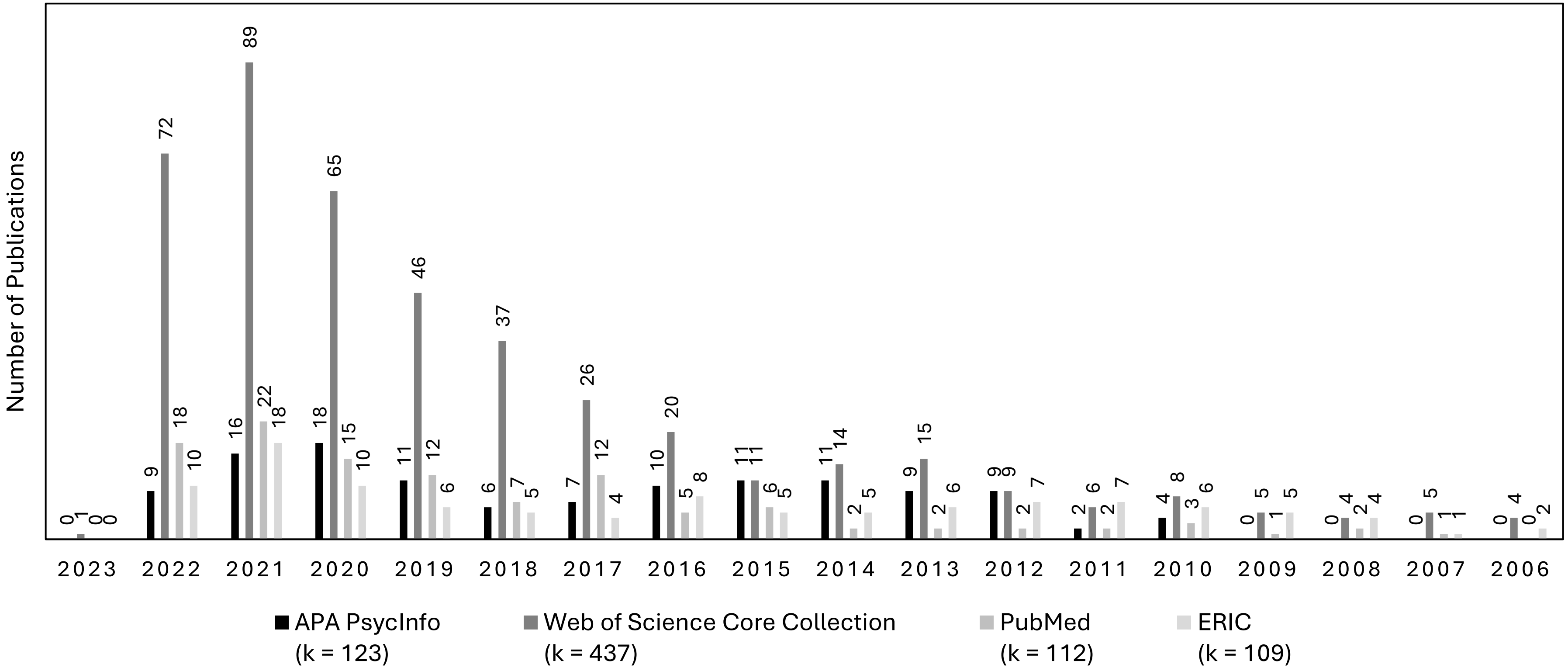
This dissertation concerns experiences of boredom in secondary school. I will detail its theoretical base and research questions on the pages to come. Before this, I first illustrate the relevance of this research topic by elaborating on the gravity of academic boredom.

First off, there is no escaping boredom at school. It affects all students alike and is easily the most frequent emotional experience in modern-day classrooms (Goetz et al., 2020; Nett et al., 2011). Second, boredom feels bad. It is an intense emotion, even when compared to supposedly more averse ones, like anxiety (Goetz & Nett, 2012; Haag & Goetz, 2012). Neuroscientific evidence even points to boredom literally hurting (Willis, 2014). Third, boredom at school is hard to grasp. Boredom research suffers from theoretical heterogeneity (Goetz et al., 2019). To make matters even more complex, boredom and other academic emotions have been shown to be experienced domain-specifically (Goetz et al., 2007). This implies that subject matter is likely linked to differential emotional experience. Fourth, boredom at school can lead to diminished academic achievement and even to drop-out (Camacho-Morles et al., 2021; Robinson, 1975; Tze et al., 2016). This tendency is exacerbated as poor academic performance also increases subsequent boredom, forming a dangerous downward spiral (Pekrun et al., 2017). Fifth, boredom's negative consequences are not exclusive to the scholar environment. Boredom, for instance, correlates with various unhealthy behaviors, such as alcohol and drug abuse or binge eating (Bench & Lench, 2013). Finally, and perhaps most gravely, boredom increases over time, especially during adolescence and secondary education (Spaeth et al., 2015; Vierhaus et al., 2016; Weybright et al., 2020). Given all this alarming information on academic boredom, it is hard to believe that direct boredom interventions are still lacking in educational practice.

On a positive note, ever since boredom was formally addressed as an achievement emotion for the first time (Pekrun, 2006), research interest in academic boredom has continuously grown stronger (see Figure 1.1). Especially in recent years, a steep increase in publication numbers with a peak in 2021 is evident. Research interest remains strong to date. One could argue that academic boredom research is at an all-time high. Interestingly, this recent period of acceleration in publication numbers coincides with the present dissertation's process (i.e., 2019 – 2023). Before detailing its contents, however, I will (try to) establish academic boredom's theoretical foundations, starting with defining boredom.

Figure 1.1

Scientific Data Base Entries for "Academic Boredom" Since 2006



Note. Results for simple all field-searches using the search string "academic boredom". Searches were conducted on January 19th, 2023.

k = total number of publications since 2006 for respective scientific data bases.

What is Boredom?

Pinpointing the boredom experience in scientific terms is an ongoing task in emotion research. Studied in various contexts – e.g., in academia (Pekrun, 2006), on the job (Reijseger et al., 2013), in relationships (Harasymchuk & Fehr, 2012), in leisure time (Ragheb & Merydith, 2001), or even in art (Elpidorou, 2017) – many questions remain on boredom's phenomenology. In many ways, boredom has always been enigmatic, and highly different from other emotions (e.g., van Tilburg & Igou, 2017). That is why I dedicate the following columns to describing boredom in detail, as far as scientific consensus is established.

Defining Boredom

Through the years there have been a plethora of definitions. Probably the most cited one, Eastwood and colleagues (2012) call boredom "the aversive state of wanting, but being unable, to engage in satisfying activity" (p. 483). Roughly 20 years earlier, Mikulas and Vodanovich (1993) stated that "boredom is a state of relatively low arousal and dissatisfaction, which is attributed to an inadequately stimulating situation" (p. 3). While both definitions appear to be different at first, they do share the idea of boredom as an emotional *state*. At this point, we could enter another ongoing discussion in the boredom literature: is boredom solely experienced in the moment, or are there habitual components to it – putting different individuals at differential risk of experiencing boredom in comparable situations? In other words: is boredom a state, a trait, or both? I will address this question briefly, before returning to contemporary definitions of boredom.

In personality research, boredom has been conceptualized as a *trait* in the past. Within Zuckerman's (1979) sensation seeking concept, the boredom susceptibility scale (ZBS) forms an important subscale. Similarly, Farmer and Sundberg (1986) developed the boredom proneness scale (BPS) to capture dispositional boredom. However, both measures have been critiqued in terms of their psychometric properties. For instance, ZBS- and BPS-scores are only weakly correlated (*r* = .17; Mercer-Lynn et al., 2014). The ZBS displayed reliability problems (Mercer-Lynn et al., 2013). The BPS has been questioned in terms of validity. Its assumed factorial structure has not been clearly replicated (Melton & Schulenberg, 2009). Moreover, BPS-sores did not show constructs stability over extended amounts of time (Gana et al., 2019). In sum, it appears that neither scale, albeit conceptually related, functions as an adequate marker of trait boredom (Vodanovich & Watt, 2016). This leaves the contemporary literature hanging in the balance when it comes to trait-ness of boredom. It would be wisest to assume a temporal continuum regarding boredom experience, with momentarily occurring boredom being targeted in most definition attempts.

Moving on from the two most historically popular definitions of boredom to two more timely ones: Westgate (2020) and Elpidorou (2021) both provide theoretical reviews, addressing conceptual heterogeneity in the boredom literature. While Elpidorou's text puts boredom's function at its center, Westgate's article focuses on avenues to boredom regulation. Even though neither review explicitly proposes a definition of boredom, there are quotable conclusions drawn in both works as to what boredom encapsulates. According to Westgate (2020), boredom is "an affective indicator of unsuccessful attentional engagement in valued goal-congruent activities" (p. 34). Elpidorou (2021) summarizes that "boredom is a sign of the presence of an unfulfilled desire to engage with our situation in a satisfactory manner and also a motivation to fulfill that desire by doing something other than what we are currently doing" (p. 506).

All four definitions are united by boredom's unsatisfactory nature. Boredom hence appears to occur whenever something holds us back, hinders us, or stops us in any way from pursuing our goals. Second, all four definitions view boredom as situation dependent. Mikulas and Vodanovich (1993), for instance, highlight an adequate amount of stimulation, while Westgate (2020) emphasises the content of the situation at hand as being of high value and relevancy. Third, all four definitions agree on boredom being an emotion as they attribute salient affective properties to it. Eastwood et al. (2012) call it an “aversive state” (p. 483), Westgate (2020) an “affective indicator” (p. 34), and Elpidorou (2021) a “sign” (p. 506). The latter account further stresses boredom’s motivational role as it makes us strive to change activities whenever possible.

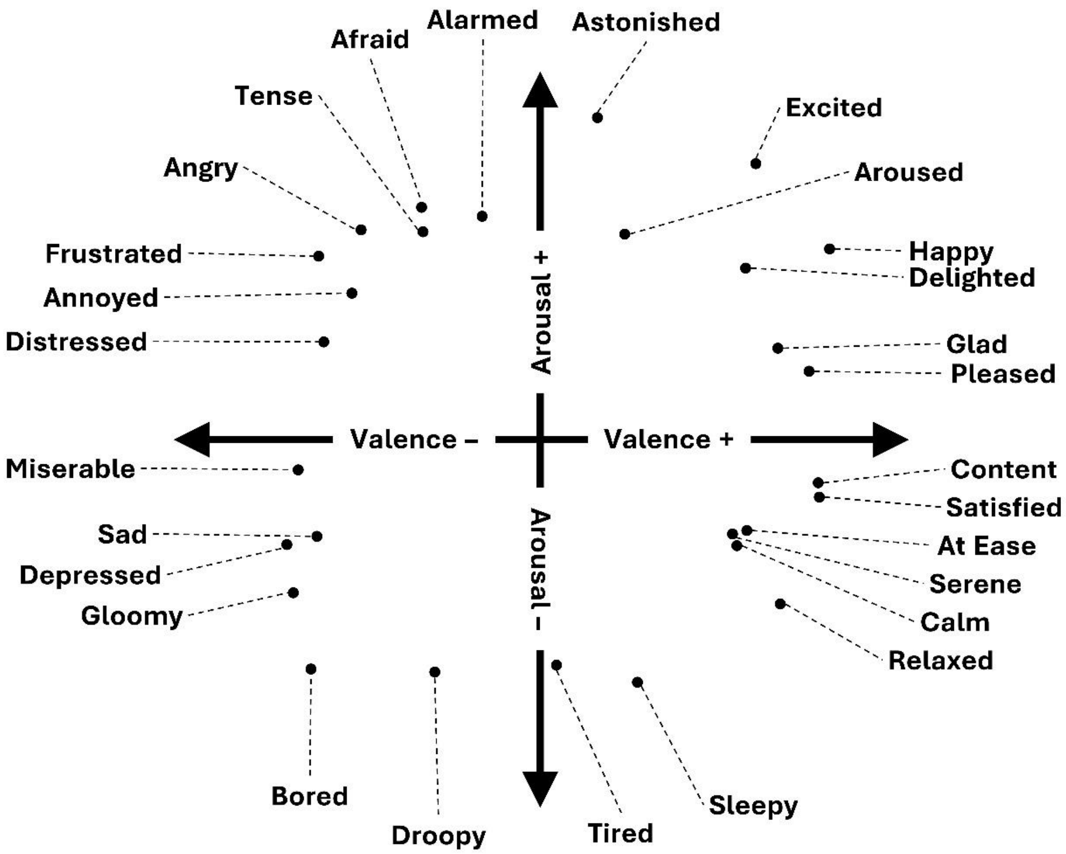
To sum up, boredom definitions, albeit different in wording, show considerable conceptual overlap. Nevertheless, theoretical models are needed beyond verbal definitions to specify the emotional experience of feeling bored.

Components of Boredom

Theories of emotion take different stances on boredom. Despite being absent from some of the most important theories of emotion in the psychological literature (e.g., Ekman, 1992; Izard, 1992; Weiner, 1985), being bored is part of Russell’s (1980) circumplex model of affect. Herein, different affective states are placed inside a valence-by-arousal-continuum, describing a circular shape (see Figure 1.2).

Figure 1.2

Circumplex Model of Affect



Note. Results of unidimensional scaling of 28 affect words conducted by Russell (1980, for original figure, see p. 1169).

The resulting grid contains four different subsections: affective states of either negative valence and low arousal (e.g., bored), negative valence and high arousal (e.g., angry), positive valence and low arousal (e.g., relaxed), or positive valence and high arousal (e.g., happy). While there is consensus on boredom's negative valence, its low-arousal status needs to be revised. Boredom has been found to correlate with low, as well as high arousal (Raffaelli et al., 2018; van Hooft & van Hooff, 2018). Because of inconclusive results in this arousal debate, it is argued that defining boredom in terms of arousal should be avoided altogether (Elpidorou, 2021). Consequently, other components need to be examined to specify boredom experiences.

Scherer's (2009) component process model of emotion proposes five components to describe and distinguish emotional experience: affect, physiology, cognition, expression, and motivation. Starting with the *affective* component, boredom is experienced as negative and averse (see above) with accompanying feelings of inertia, emptiness, anger or impulsivity (Sommers & Vodanovich, 2000; Titz, 2001). From a *physiological* standpoint, inconclusive results on arousal loom large, with boredom being associated, for instance, with both sleepiness and restlessness (Goetz & Frenzel, 2006; Sundberg & Bisno, 1983). Regarding *cognitive* processes, feeling bored is accompanied by concentration difficulties, task-unrelated thoughts, perceptions of time dilatation and mind-wandering (Eastwood et al., 2012; Raffaelli et al., 2018; Westgate & Wilson, 2018). For the *expressive* component, I kindly refer the reader to the collage at the beginning of this dissertation. Finally, boredom decidedly *motivates* the pursuit of novel activity (Bench & Lench, 2013; Elpidorou, 2021).

To sum up, when feeling bored, arousal is either high or low, goal-directed cognitive functions are impeded, and aversiveness prevails. Definitive consensus is limited to boredom's core function of motivating changes in the current activity. After detailing the complexity of boredom as an emotion (and failing at establishing a common ground), I will narrow my focus to experiences of boredom in achievement situations.

The Control-Value Theory of Achievement Emotions

In 2006, Pekrun introduced the control-value theory (CVT) of achievement emotions (see also Pekrun, 2019; Pekrun et al., 2023). This marked a turning point in educational emotion research, as emotions linked to achievement settings were embedded in a theoretical framework for the first time. There are other achievement settings beyond academia – for instance, in sports, competitive job environments, or the performing arts. Herein, however, I will focus on the academic realm. CVT drew from many theories historically rooted in motivational psychology (e.g., Atkinson's risk preference model, 1964; flow theory, Csikszentmihalyi, 1990; or Weiner's attributional theory, 1985). Its innovation lay in combining existing approaches and tailoring its premises towards achievement situations.

Taxonomy of Achievement Emotions

Before CVT's rise, test anxiety was the only emotion relevant for educational research (Mandler & Sarason, 1952). Nowadays, 20 distinct emotions experienced in achievement-related contexts can be systematically distinguished (see Table 1.1). These *achievement emotions* are defined as "emotions tied directly to achievement activities or achievement outcomes" (Pekrun, 2006, p. 317). This definition contains the first important distinction within achievement emotions: there are *activity emotions* (e.g., boredom) and there are *outcome emotions* (e.g., hope or pride). They differ in temporal object focus. Activity emotions are experienced during achievement activities (e.g., while preparing a presentation) whereas outcome emotions relate to future (i.e., *prospective* outcome emotions) or past events (i.e., *retrospective* outcome emotions). Boredom is an activity emotion, underlining its momentary nature stressed also in previously introduced construct definitions (see above).

Table 1.1

Taxonomy of Achievement Emotions

			+		–	
Object Focus	Control	Value	Activating	Deactivating	Activating	Deactivating
Activity Emotions	High	Positive	Enjoyment, Excitement	Relaxation°		
	High	Negative			Anger	
	Low	Positive/Negative			Frustration	
	High/Low	None				Boredom
Prospective Outcome Emotions	High	Positive (Success)	Anticipatory Joy	Assurance*		
	Medium		Hope			
	Low					Hopelessness
	High	Negative (Failure)		Anticipatory Relief[X]		
	Medium				Anxiety	
	Low					Hopelessness
Retrospective Outcome Emotions	Irrelevant	Positive (Success)	Retrospective Joy	Relief°		
	Self		Pride	Contentment		
	Other		Gratitude			
	Irrelevant	Negative (Failure)				Sadness
	Self				Shame/Guilt	
	Other				Anger	Disappointment°

Note. For original taxonomy see Pekrun (2006, p. 320). For updated taxonomy see Pekrun et al. (2023, p. 149).

* added in 2023. ° added in later version (cf. Pekrun, 2017). [X] removed in 2023.

Besides object focus, achievement emotions can also be distinguished in terms of CVT's name-giving control and value appraisals. In case of prospective outcome emotions, control appraisals regarding future achievement situations (e.g., an upcoming biology exam) were assumed to either be low (e.g., hopelessness), medium (e.g., hope) or high (e.g., anticipatory joy). For retrospective outcome emotions, control appraisals revert to attributions of success or failure in a past achievement situation (e.g., who is to blame for me not passing the recent biology exam?). Retrospectively, control was either exerted by oneself (e.g., shame), or located in the outside circumstances (e.g., gratitude for the exam having been composed of rather easy questions). Locus of control might also have been irrelevant for the resulting emotion (e.g., joy or sadness). Control appraisals in activity emotions are usually either low (e.g., frustration) or high (e.g., enjoyment). For boredom, however, both can apply. That way, boredom is experienced in underchallenging (i.e., high control) as well as overchallenging (i.e., low control) achievement situations (cf. Acee et al., 2010; see also Chapter 2).

Value appraisals are less complex, compared to control appraisals. They can either be positive or negative, regardless of object focus. They entail how important is it to an individual to succeed in a given achievement situation. Again, boredom begs to differ, as it is the only achievement emotion for which subjective value is neither positive nor negative but nonexistent. That means that boredom is expected to occur in achievement situations that we do not care about. This notion is contradictory, however, to Westgate's (2020) boredom definition, where it is precisely stated that boredom pertains to "valued goal-congruent activities" (p. 34).

In newer accounts of CVT (e.g., Pekrun et al., 2023) the structure of achievement emotions was slightly altered. Instead of separating emotions by object focus, control and value, they are now taxonomized in terms of object focus, valence, and arousal (cf. Russell, 1980). The valence dimension, i.e., whether an emotion is positive or negative, largely overlaps with value appraisals. Boredom is now listed as a negative emotion, which aligns with its aversive nature (cf. Eastwood et al., 2012). The arousal dimension divides emotions into two categories: activating (e.g., anger) and deactivating (e.g., boredom) emotions. With the inconclusiveness in boredom's arousal debate illustrated above, labelling it a deactivating emotion can certainly be discussed. Even in this newer taxonomy, boredom appears to not quite fit in.

To sum up, boredom is a negative deactivating activity achievement emotion. According to control-value theory, boring achievement situations are characterized by either high or low subjective control and an absence of subjective value.

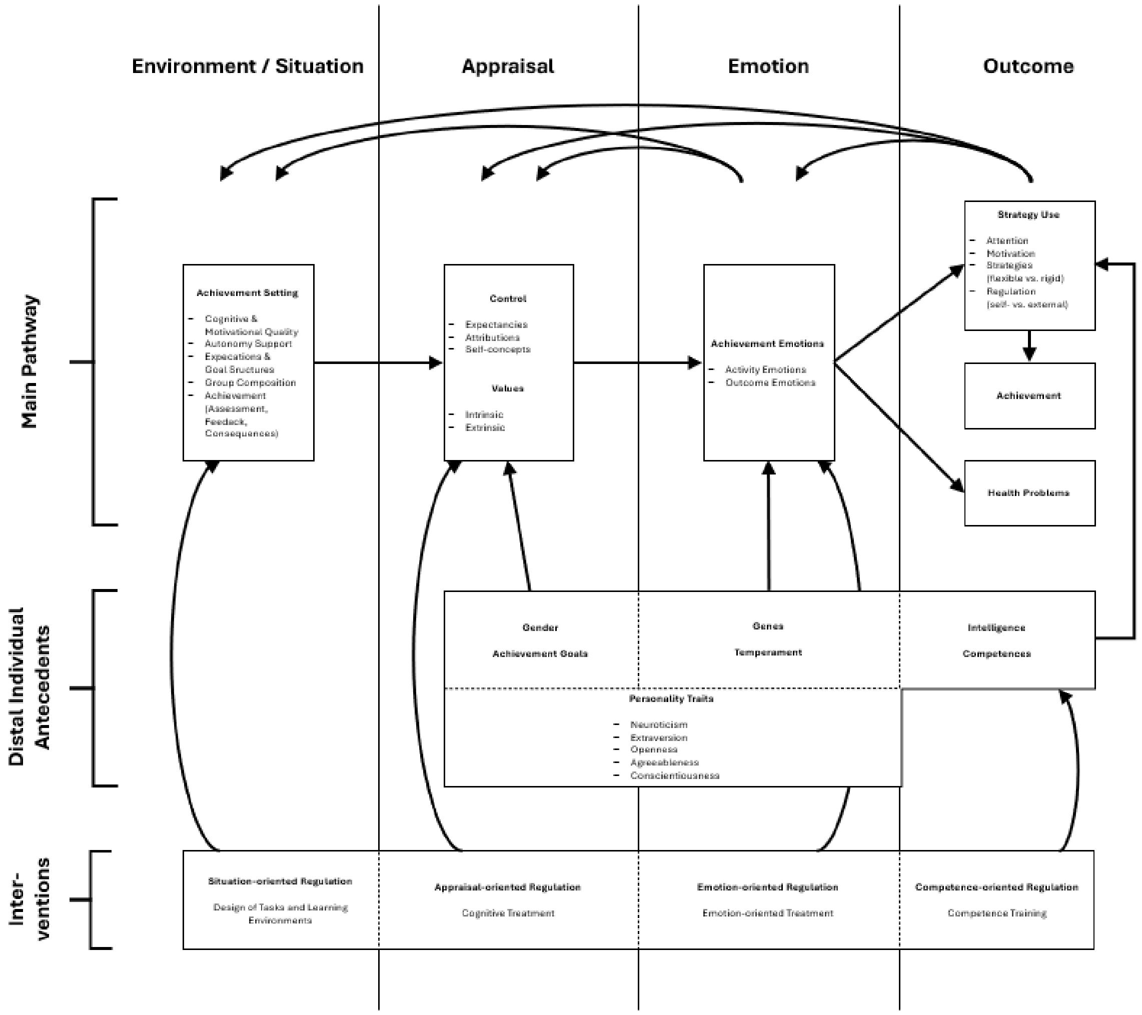
Theoretical Framework

Besides offering a continuously updated taxonomy of achievement emotions, CVT also features a theoretical framework serving as base for many empirical studies. Herein, a motivational process is outlined that establishes a chain of events leading to emotion elicitation, which, in turn, affects academic outcomes. With emotion at its core, different antecedents as well as consequences to emotional experience in achievement settings are proposed. In-between different steps of this process, feedback loops are possible, however. Still, I will focus on the main pathway of emotional development. The motivational process implied by CVT is depicted in Figure 1.3[2].

[2] It is worth noting that the theoretical framework has been altered significantly in CVT's latest amendment (see Pekrun et al., 2023). For example, regulative processes, serving as starting points for classroom-based interventions, have been excluded in the 2023 version. Other variables, such as personality traits or health problems, have gained in importance. However, the central pathway of emotional development remained rather stable.

Figure 1.3

Control-Value Theory's Theoretical Framework



Note. Hybrid model of CVT's core propositions (2019-2023). For original models, see Pekrun (2019, p. 147) and Pekrun et al. (2023, p. 147).

Regarding the learning *environment*, CVT emphasizes cognitive and motivational quality of instruction, clear expectations and goal structures, and a climate of autonomy as favorable attributes of the classroom. However, learning environments are not solely encountered at school. They also include students' private and family life (Robinson, 1975). Learning environments promoting boredom have been described as monotonous, with tasks low in complexity and variety, resulting in reduced intellectual stimulation (Pekrun et al., 2010). Lack of clarity and structure in class is a source of boredom, as well (Goetz, 2004). Other teaching factors also play their part, as a supportive presentation style can reduce boredom, while excessive demands may enhance it (Goetz et al., 2020). Teachers' enthusiasm and humor have been shown to protect against boredom (Bieg et al., 2019). On the other hand, boredom on the teacher's part can promote boredom among students, if recognized (Tam et al., 2020). This is especially important, as teachers often fail to view themselves as responsible for boredom experiences in their classrooms (Daschmann et al., 2014).

Besides environmental factors, *appraisals* are the primary and proximal antecedents to emotional experience in CVT. I already introduced control and value appraisals and their potential results (see above). However, appraisals can come in different forms (cf. Pekrun, 2019). Control appraisals, for instance, subsume expectancies, attributions, and self-concepts. Value appraisals can refer to both intrinsic and extrinsic aspects. Regarding boredom, subjective control shows an inconclusive pattern. Befitting boredom's theoretical appraisal profile, subjective control has been both positively (e.g., Goetz et al., 2020) and negatively associated with later boredom (e.g., Clem et al., 2021). Other studies found no relation (e.g., Putwain et al., 2018). When different forms of boredom are distinguished, academic self-concept has been shown to negatively predict boredom due to overchallenge and positively predict boredom due to underchallenge in mathematics (Goetz & Frenzel, 2010). Subject value, on the other hand, has consistent negative links to boredom experiences, with higher effect sizes, compared to subjective control (Forsblom et al., 2021; Goetz et al., 2020; Pekrun et al., 2010). Within different value appraisals, intrinsic value and motivation are most predictive of boredom (Putwain et al., 2018; Sutter-Brandenberger et al., 2018).

Besides proximal appraisal antecedents, CVT proposes person characteristics as distal antecedents to achievement emotions. These should primarily affect appraisals but can also influence emotion directly. Examples of such distal individual antecedents are demographic characteristics (e.g., gender) or cognitive abilities (e.g., intelligence). Regarding gender, for example, some studies point to higher boredom in male students (e.g., Pekrun et al., 2017; Raccanello et al., 2019; Zaccoletti et al., 2020), while others did not find significant gender effects (e.g., Forsblom et al., 2021). When distinguishing between boredom due to under- and overchallenge, however, girls report more boredom because of overchallenge, in mathematics, with boys stating more boredom because of underchallenge (Daschmann et al., 2011; Goetz & Frenzel, 2010). Not much research exists on intelligence-boredom relations. Gifted students displaying high cognitive abilities, for instance, have been found to experience less (Gjesme, 1977), more (Larson & Richards, 1991; Stambaugh, 2017) and equal amounts of boredom (Feldhusen & Kroll, 1991; Hornstra et al., 2017; Preckel et al., 2010), in comparison to age-appropriate control samples, on different occasions.

After emotions flare up, they can lead to different consequences. Among various *outcomes*, academic achievement is of prime relevance to educators and educational scientists alike. Empirical evidence unequivocally points to boredom leading to reduced academic performance. This connection has been established in singular studies (Forsblom et al., 2021; Hunter & Eastwood, 2021; Lichtenfeld et al., 2022; Pekrun et al., 2017; Putwain et al., 2022), as well as in meta-analyses (Camacho-Morles et al., 2021; Tze et al., 2016).

Beyond achievement, health problems have received increased recognition in modern-day CVT (Pekrun et al., 2023). Boredom is associated with a list of problematic health outcomes (see Bench & Lench, 2013, for a review), including reduced psychological well-being (Schwartze et al., 2021), depressive symptoms (LePera, 2011), and substance abuse (Freund et al., 2021). Aside from negative effects of boredom, positive consequences are rarely mentioned in the academic literature. Therefore, assumed benefits in relaxation and creative incubation need further investigation (Elpidorou, 2014).

To sum up, boredom has been widely researched within the CVT framework. Results indicate that the learning environment is of central importance to classroom-bound boredom. Regarding appraisal antecedents, boredom is associated with negative value, while control patterns may vary. Negative associations with academic performance are firmly established. Let us now delve into the present dissertational project, which also uses CVT as a theoretical framework.

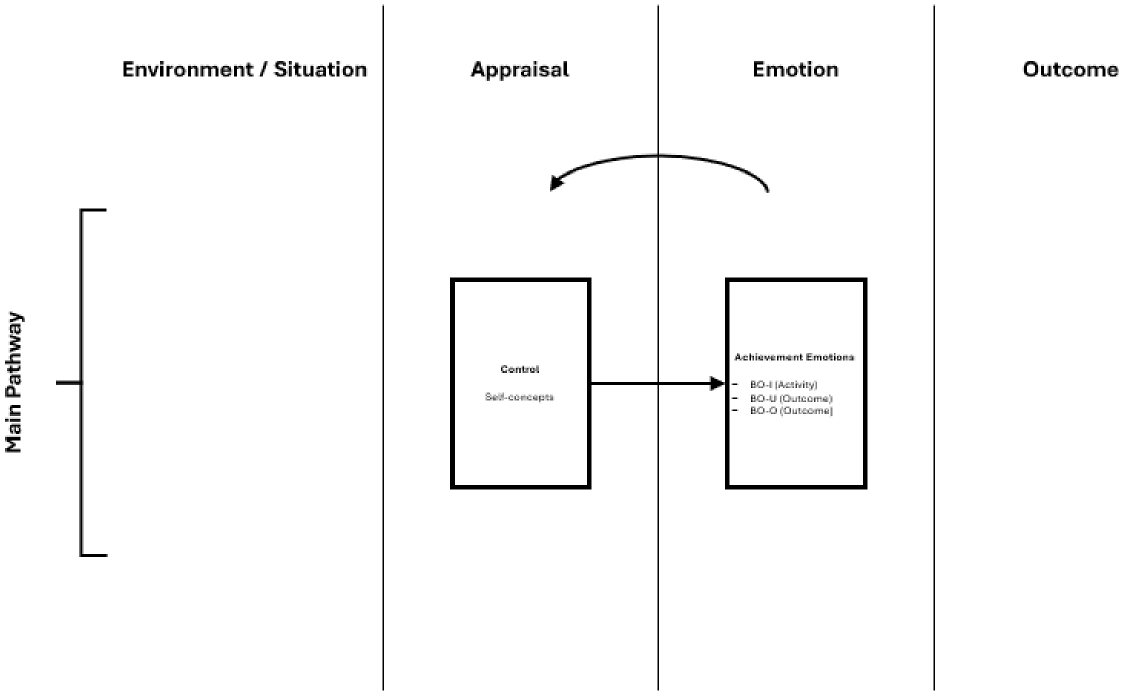
The Present Dissertation

Objectives

I compiled three studies (Studies 1-3) into the present dissertation. Each work tackles different research questions that can be projected onto CVT's theoretical framework. The purpose of Study 1 was the empirical separation of different forms of boredom. There is ongoing debate on boredom being a singular vs. a facetted construct. Functional accounts (Bench & Lench, 2013; Elpidorou, 2018, 2021) emphasise boredom's unity through its motivational component. Other authors have found different types of boredom, using factor and profile analyses (Acee et al., 2010; Daschmann et al., 2014; Goetz et al., 2014). Moreover, boredom and its different forms have been shown to differentially relate to academic self-concept and task difficulty perceptions (Krannich et al., 2019; Westgate & Wilson, 2018). Therefore, studying interrelations between boredom and academic self-concept as control appraisal and important individual characteristic related to academic challenge perception formed an addition to this construct validation approach. Precise research questions, hypotheses, methods, analytical results, and implications from Study 1 are laid out in chapter 2 of this dissertation. Figure 1.4 depicts the portions of CVT examined.

Figure 1.4

Portions of Control-Value Theory's Theoretical Framework Examined in Study 1

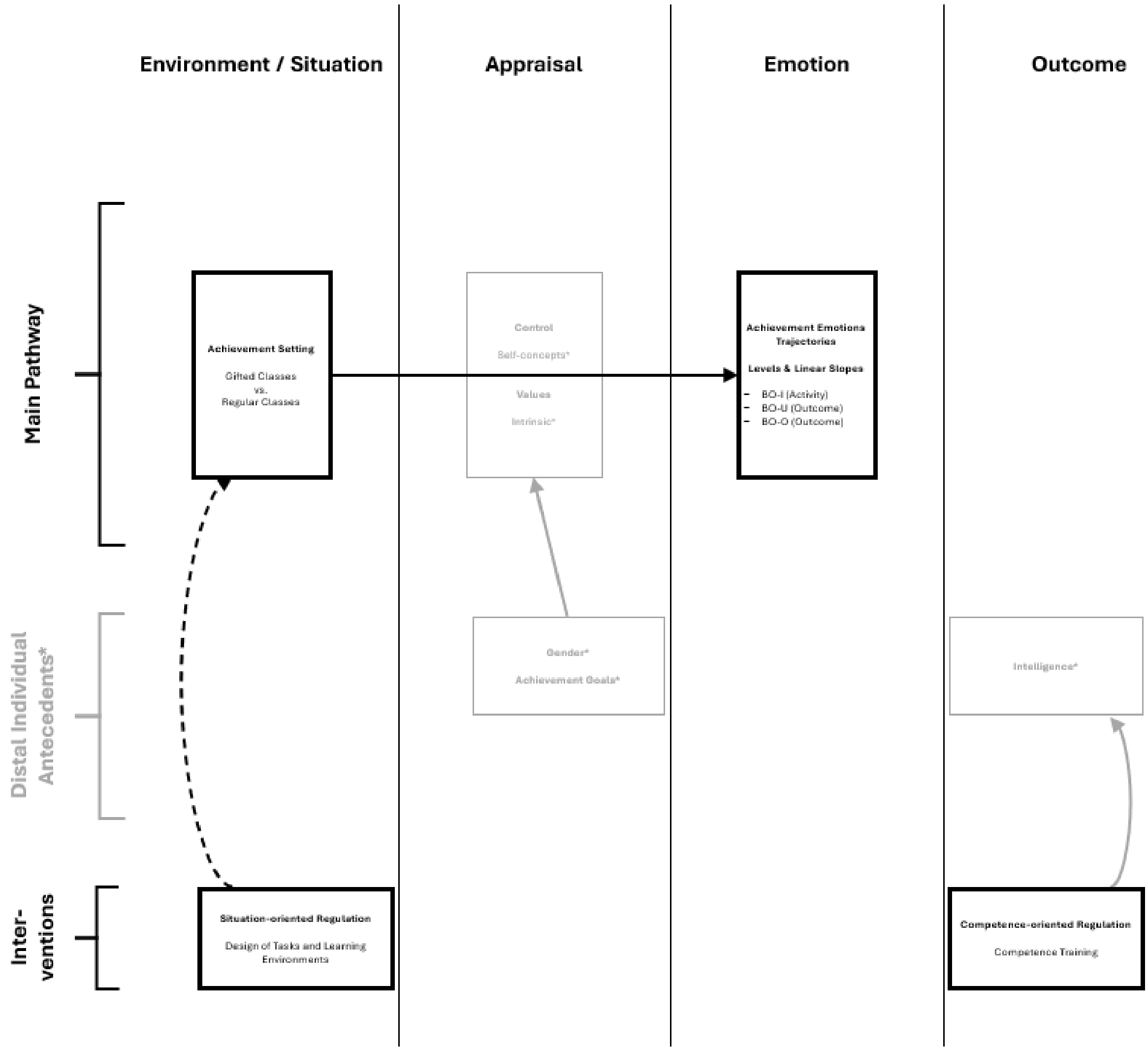


Note. For details on exact implementation, see chapter 2. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge.

Study 2 focused on intellectually gifted populations. It is often argued that high-ability students are at an increased risk of experiencing underchallenge and boredom at school. This proposition is also known as the *boredom hypothesis* (Feldhusen & Kroll, 1991). Grouping intellectually gifted students into special classrooms has often been advocated as an ample compensation strategy (Bar-On, 2007; Plucker et al., 2004). However, neither the boredom hypothesis nor the proposed effectiveness of gifted classes has been tested using convincing methodological approaches. Study 2 closes this gap, providing a systematic approach to evaluating ability grouping as a classroom-based boredom intervention. To isolate class type-effects, propensity score matching was applied (Rosenbaum & Rubin, 1983). Study 2 is outlined in chapter 3 (see also Figure 1.5).

Figure 1.5

Portions of Control-Value Theory's Theoretical Framework Examined in Study 2

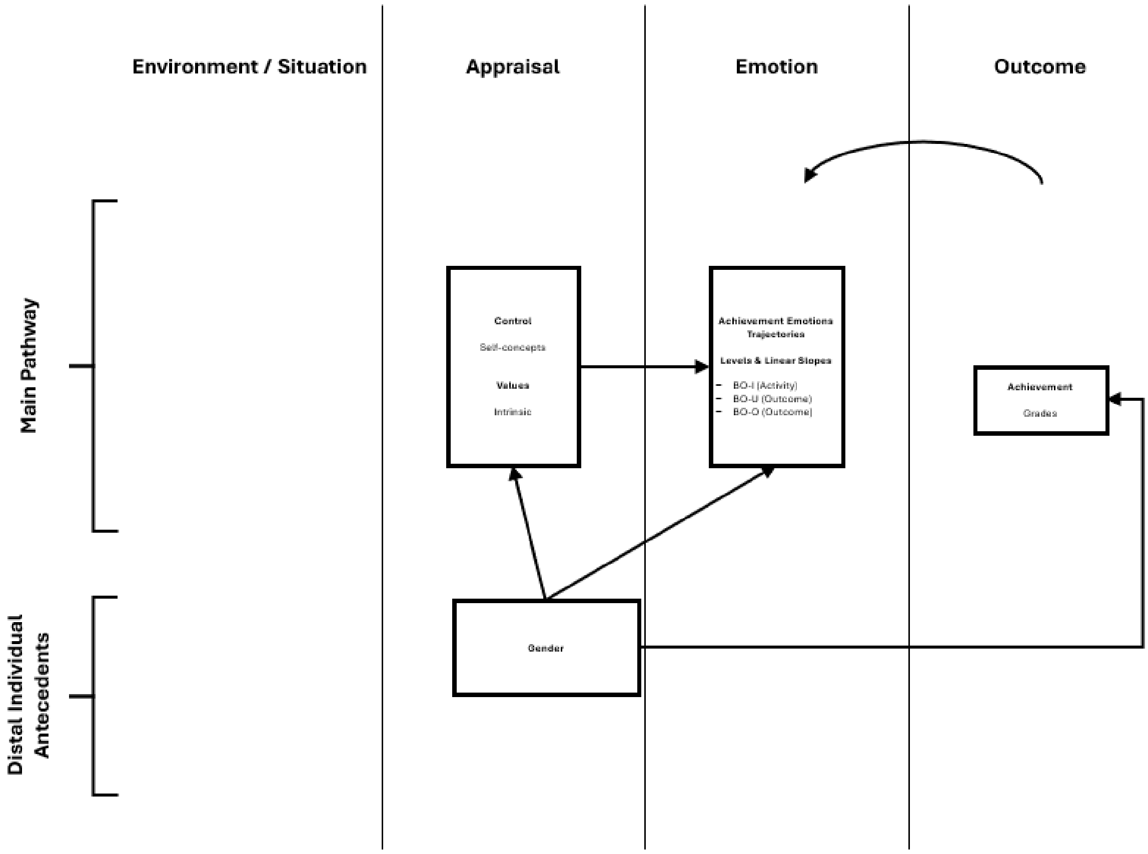


Note. For details on exact implementation, see chapter 3. Ability grouping as an educational intervention, possess attributes of situation- and competence-oriented regulation (cf. Pekrun, 2019). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. * Variables controlled via propensity score matching (grey).

Study 3 examined gender differences in academic boredom development. Few longitudinal studies exist on this topic (e.g., Weybright et al., 2020). Cross-sectional findings indicate higher subject-specific boredom for boys vs. girls (for the mathematics domain, see Daschmann et al., 2011; Goetz & Frenzel, 2010; for the verbal domain, see Raccanello et al., 2019; Zaccoletti et al., 2020). Study 3 covered the largest portion of the CVT framework, as gender, control and value appraisals, and academic achievement, were all embedded into the same analytical model (see Figure 1.6). Thereby, I was able to concurrently examine gender differences in boredom trajectories, as well as possible explanations for them. Study 3 is detailed in chapter 4.

Figure 1.6

Portions of Control-Value Theory's Theoretical Framework Examined in Study 3



Note. For details on exact implementation, see chapter 4. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge.

Methodological Directions

All studies from this dissertation are characterized by a longitudinal research approach, analysing data provided from multiple schools over several years of research cooperation.[3] As all relevant data had already been acquired, conducting optimal methodological analyses to extract the most important results and promote new insights

[3] Data for this dissertation stem from the AVG-project (for a research report, see https://www.uni-trier.de/fileadmin/fb1/prof/PSY/HBF/Sachbericht_2013_Klassen5-8_final_12-11-2014.pdf). This longitudinal study was conducted from 2005 – 2020, examining motivational and emotional development in five successive cohorts from five German schools, over the course of secondary education and beyond. Methodological details are issued in chapters 2-4 as individual studies pertaining to this work are addressed.

became my top priority in the process. For instance, I examined self-report data with structural equation modelling (SEM; Jöreskog, 1970) techniques. That way, I could effectively account for measurement error, while also providing a confirmation framework to test some of CVT's most important propositions (see Pekrun, 2006, 2019; Pekrun et al., 2023). Furthermore, I tested for different types of measurement invariance (cf. Byrne, 2012; Geiser, 2011; Jöreskog, 1971; Kleinke et al., 2017) as necessary prerequisites to investigating group differences (cf. Studies 2 and 3) and long-term developmental processes (e.g., reciprocal causation, cf. Study 1 or latent growth, cf. Studies 2 and 3). Moreover, I was able to improve methodological precision within each study, by using techniques, such as propensity score matching (PSM; Rosenbaum & Rubin, 1983, 1984, 1985, cf. Study 2), the inclusion of control variables (cf. Study 3), or reciprocal effects models (REM; Hamaker et al., 2015; Usami et al., 2019, cf. Study 1).

Project Overview

Before I close this topical introduction, I kindly refer the reader to take note of Table 1.2. This project overview serves as an advanced organizer to the present dissertation. It summarizes the scientific status quo for Studies 1-3 at the time this dissertational text is written. Thereby, I cared to provide transparency for my process, while also highlighting the scientific effort behind each study.

Table 1.2

Project Overview

				Publication Progress			
Study	Title	Chapter	Appendix	Conference Contribtions[a]	Submission Attempts (Journals)	Manuscipt Status (Journals)	Article Reference (URL)
1	Academic Boredom and Self-Concept in Secondary School – Subject-specific Correlations and Reciprocal Effects	2	2	AERA Annual Meeting 2023 (Poster)[b]	1	Submitted to British Journal of Educational Psychology	n/a
2	Reducing Boredom in Gifted Education – Evaluating the Effects of Full-Time Ability Grouping	3	3	- 2. Interdisziplinäre Graduiertenkonferenz (Poster) - JURE 2021 (Paper Presentation) - PAEPSY 2021 (Paper Presentation as Part of a Symposium)	1	Accepted / Published in Journal of Educational Psychology	https://doi.org/10.1037/edu0000694
3	Gender Differences in Academic Boredom and Its Development in Secondary School	4	4	- 3. Interdisziplinäre Graduiertenkonferenz (Paper Presentation) - AERA Annual Meeting 2022 (Paper Presentation) - DGPs-Kongress (Poster)	6	Ready for Submission to Learning and Individual Differences	n/a

Note. Current state as this dissertation is written (March, 27th 2023).

[a] See references for details (Feuchter & Preckel, 2020, 2021a, 2021b, 2021c, 2022a, 2022c, 2023).

[b] A 2,000 word summary of Study 1 has been accepted for the AERA Annual Meeting 2023, scheduled to take place April 16-20, 2023.

[c] Prof. Dr. Franzis Preckel supervised this dissertational project and contributed to all studies listed here as sole co-author.

AERA = American Educational Research Association. JURE = Junior Researchers of the European Association on Learning and Instruction (EARLI). DGPs = Deutsche Gesellschaft für Psychologie.

Chapter 2 – Study 1

Academic Boredom and Self-Concept in Secondary School – Subject-specific Correlations and Reciprocal Effects

Markus D. Feuchter and Franzis Preckel

Department of Psychology, University of Trier

Abstract

In this longitudinal investigation in German secondary school (N = 1,432, grades 5-8, $M_{\text{age at T1}}$ = 10.24 years), intensity of boredom, boredom due to overchallenge, boredom due to underchallenge, and academic self-concept were investigated in mathematics and German. Results supported the separability of different forms of boredom. Latent correlations revealed increasing dissociations between boredom due to over- and underchallenge, intensity of boredom and boredom due to underchallenge, and stable relations between intensity of boredom and boredom due to overchallenge. Evidence for reciprocal effects between boredom and self-concepts was limited. Findings support the 3-factor-model of boredom and point to domain specificities in the development of boredom-self-concept-relations.

Keywords: Academic boredom, Academic Self-concept, Underchallenge, Overchallenge, Reciprocal effects

Academic Boredom and Self-Concept in Secondary School – Subject-specific Correlations and Reciprocal Effects

Despite growing research interest, boredom remains a complicated emotion not yet grasped by researchers in a widely agreed upon fashion. Many authors state that there are different forms of boredom, for example, due to under- vs. overchallenge (Acee et al., 2010; Daschmann et al., 2011; Westgate & Wilson, 2018, Feuchter & Preckel, 2022b). Others view boredom as a unitary phenomenon (Elpidorou, 2021). Boredom's structure as a psychological construct needs to be understood more deeply. Moreover, antecedents to differential boredom experiences are understudied, especially regarding boredom due to under- and overchallenge (Daschmann et al., 2011). Academic self-concepts form theoretically important appraisal antecedents to boredom (Pekrun, 2019) and have been demonstrated to differentially relate to experiences of under- and overchallenge (Krannich et al., 2019; Preckel et al., 2010). Furthermore, long-term development in boredom and its antecedents needs further investigation. This concerns feedback loops and reciprocities proposed by theory (Pekrun, 2019). Related longitudinal studies covering several years of secondary education are rare (Clem et al., 2021; Forsblom et al., 2021).

This present work aims to provide a first construct validation study for a previously established three-factor-model of boredom (3F-model of boredom; Feuchter & Preckel, 2022b). Herein, we provide cross-sectional (i.e., construct correlations) and longitudinal evidence (i.e., reciprocal relations) from 3.5 years of German secondary education. We further investigated the role of academic self-concept as a crucial conceptual antecedent to boredom experiences. In addition, we investigated our research questions in two subject domains (namely, mathematics and German). Findings of our study thus entail differentiated evidence on academic boredom's construct validity, add to existing theories on boredom development, and yield important implications for teaching practice.

Boredom in Secondary School

In general, boredom is "an affective indicator of unsuccessful attentional engagement in valued goal-congruent activities" (Westgate, 2020, p. 34). Within Pekrun et al.'s (2023) taxonomy of achievement emotions, boredom is a negative deactivating activity emotion. Beyond its high prevalence in classrooms (Goetz et al., 2020), boredom does not constitute a mere state of indifference or reduced interest but is experienced as highly aversive and emotional (Haag & Goetz, 2012). Considering its consequences, boredom is associated with many problematic outcomes in academia (e.g., reduced achievement or decreased motivation, see Camacho-Morles et al., 2021; Tze et al., 2016), as well as in personal life (e.g., reduced psychological well-being or increased unhealthy behaviors, see Schwartze et al., 2021; Weybright et al., 2015).

For secondary education, increasing boredom trajectories are reported frequently and across educational systems (e.g., Feuchter & Preckel, 2022b; Weybright et al., 2020). Boredom has also been shown to have negative reciprocal relations with academic achievement (Forsblom et al., 2021; Pekrun et al., 2017). This downward spiral is already evident at the primary school stage (Lichtenfeld et al., 2022) and lingers beyond secondary school (Hunter & Eastwood, 2021). Many of these studies use reciprocal effects models (REMs) to study boredom's interplay with otherwise important learning outcomes (for a methods overview, see Usami et al., 2019). However, important subject specificities to classroom-based motivation and emotion (Goetz et al., 2010) are often neglected. Research on subject-specific boredom is mostly carried out for the mathematics domain, with other domains, such as verbal ones, being underrepresented.

Different Forms of Academic Boredom

The phenomenological nature of boredom as a psychological construct is unclear. Although boredom's internal processes (e.g., negative affect, perception of time dilatation, see Eastwood et al., 2012) and its functionality (i.e., signalling situational discomfort and motivating activity changes, see Elpidorou, 2021) have been established, there is a plurality of boredom conceptualizations in the contemporary educational literature. Whether boredom is a unitary or faceted emotion remains open to debate. In the following, we list a variety of approaches to different forms of boredom generated in educational research.

Acee and colleagues (2010), as well as Daschmann et al. (2011), expanded upon the idea that boredom is elicited by nonoptimal subjective challenge (Csikszentmihalyi, 1975). Using confirmatory factor analyses, both teams provided psychometrical distinctions between boredom due to under- and overchallenge, captured in the Academic Boredom Scale (ABS; Acee et al., 2010), and the Precursors to Boredom Scales (PBS; Daschmann et al., 2011). Along a similar line of reasoning, Westgate and Wilson (2018) distinguish between two attentional subtypes of boredom in their meaning and attentional components (MAC) model: attentional boredom resulting from under- and overchallenge. Most recently, Feuchter and Preckel (2022b) featured a three-factor-model of boredom (3F-model), combining different short scale assessments. Herein, boredom intensity, measured by items from the Achievement Emotions Questionnaire (AEQ-M; Pekrun et al., 2005), and challenge-related boredom due to either under- or overchallenge are used as separate but correlated latent constructs. Preckel et al. (2010) used a similar threefold approach to different forms of boredom, analysing manifest scale data. All these models share the idea of separate, challenge-related forms of boredom (see also Goetz & Frenzel, 2010) but differ in factorial structure and domain specificity (Acee et al., 2010: one vs. two factors, depending on under- vs. overchallenge, analysed without separating domains; Daschmann et al., 2011: eight factors analysed for mathematics; Westgate & Wilson, 2018: theoretical distinction yet to be examined psychometrically; Feuchter & Preckel, 2022b: three subject-specific factors (mathematics vs. German)).

Interrelations of different forms of boredom are understudied overall, with verbal domains being neglected altogether. For mathematics and subject-unspecific assessments, general boredom has been found to be positively correlated with boredom due to underchallenge (r = .64, Acee et al., 2010; $r_{\text{Mathematics}}$ = .10, Daschmann et al., 2011; .18 ≤ $r_{\text{Mathematics}}$ ≤ .22, Preckel et al., 2010) and overchallenge (.45 ≤ r ≤ .48, Acee et al., 2010; $r_{\text{Mathematics}}$ = .29, Daschmann et al., 2011; .17 ≤ $r_{\text{Mathematics}}$ ≤ .43, Preckel et al., 2010), with both challenge-related forms being negatively correlated with each other for mathematics (-.27 ≤ r ≤ -.35, Goetz & Frenzel, 2010; r = -.49, Daschmann et al., 2011; -.30 ≤ r ≤ -.55, Preckel et al., 2010). Findings for longitudinal relations of the different forms of boredom are lacking.

Antecedents to Academic Boredom

For the academic realm, Pekrun's (2006, 2019) control-value theory (CVT) forms a prominent process model of achievement emotions such as academic boredom. In CVT, boredom is experienced on-task whenever success is deemed irrelevant (i.e., low subjective value) and either too easily (i.e., high subjective control) or too uncertainly achieved (i.e., low subjective control). As both extremes regarding subjective control are expected to promote boredom, different control-based forms of boredom become plausible (i.e., boredom due to underchallenge and boredom due to overchallenge). Empirically, academic boredom and subjective control predominantly showed negative correlations (e.g., Frenzel et al., 2007; Pekrun et al., 2010). Still, research relating different forms of boredom to subjective control is lacking, potentially neglecting boredom experiences reverting to perceptions of disproportionately high subjective control.

Beyond control and value appraisals, academic boredom is distally affected by characteristics of the learning environment (e.g., monotony, Hill & Perkins, 1985; teaching characteristics, Goetz et al., 2020; teacher characteristics, Bieg et al., 2019) and by personal characteristics (e.g., dispositional boredom, Farmer & Sundberg, 1986). Furthermore, situation and person characteristics interact as additional boredom antecedents in a person-environment-fit sense (see Goetz et al., 2019). An established antecedent to boredom is a nonoptimal level of individual challenge (Raffaelli et al., 2018). Originally thought to be exclusive to underchallenging situations (Csikszentmihalyi, 1975), boredom has also been found to occur in overchallenging situations (Acee et al., 2010; Daschmann et al., 2011). Overall, evidence for differentiated challenge-related boredom experience has rarely been linked to its antecedents in past educational studies (e.g., Krannich et al., 2019).

Academic Self-Concept and Boredom

When discussing different forms of academic boredom in terms of differential subjective control, academic self-concept is an important variable to consider in educational settings. Academic self-concept refers to a person's self-evaluation of one's ability in general (e.g., "I am good at school") or in a specific academic domain (e.g., "I am good at Math") (Brunner et al., 2010). Academic self-concepts are relevant control appraisals (cf. Pekrun, 2019) and have been used to operationalize subjective control in various studies (e.g., Boehme et al., 2017). Empirically, academic self-concepts have widely shown negative associations with generalized boredom measures in secondary school (e.g., Clem et al., 2021; Forsblom et al., 2021; Krannich et al., 2019; Preckel et al., 2010). Negative correlations were larger for mathematics than for verbal domains ($r_{\text{Mathematics}}$ = -.23, r_{German} = -.11, Krannich et al., 2019; -.17 ≤ $r_{\text{Mathematics}}$ ≤-.30, -.11 ≤ r_{Literacy} ≤ -.20, Clem et al., 2021). Regarding challenge-related forms of boredom, Goetz and Frenzel (2010) found academic self-concept to positively predict boredom due to underchallenge (β = .58, r = .46) and negatively predict boredom due to overchallenge (β = -.71, r = -.60) in mathematics. Examining the inverse relation, Krannich et al. (2019) found perceptions of underchallenge to positively predict later boredom ($\beta_{\text{Mathematics}}$ = .22, β_{German} = .29) and self-concept ($\beta_{\text{Mathematics}}$ = .20, β_{German} = .22).

Being overchallenged, on the other hand, predicted self-concept negatively ($\beta_{\text{Mathematics}} = -.56$, $\beta_{\text{German}} = -.34$) while predicting boredom positively ($\beta_{\text{Mathematics}} = .37$, $\beta_{\text{German}} = .29$). More recent research, spanning several years of secondary education (grades 6-7, Clem et al., 2021; grades 5-10; Forsblom et al., 2021), featured reciprocal relations of subject-specific academic self-concepts and boredom. Both studies found academic self-concept to negatively predict later boredom for mathematics ($-.08 \leq \beta \leq -.12$, Clem et al., 2021; $-.13 \leq \beta \leq -.14$, Forsblom et al., 2021) but not for literacy. Conversely, boredom did not predict academic self-concept in either domain. Although implied by CVT, reciprocal effects of different forms of boredom and academic self-concept have not been examined domain-specifically.

The Present Study

In the present study, we explored reciprocal relations of different forms of subject-specific academic boredom and academic self-concept over the course of early secondary education in a German high-track student sample (*N* = 1,432, grades 5-8). We contributed to the debate on different forms of academic boredom and their factorial structure by further examining the 3F-model. As domain-specific assessments of emotion and motivation constructs are often neglected or carried out solely for the mathematics domain, we sought to close this research gap by exploring reciprocal effects of boredom intensity, boredom due to underchallenge, boredom due to overchallenge, and academic self-concept for two scholarly subjects (i.e., mathematics and German). To our knowledge, no educational study of reciprocal causation has systematically and simultaneously featured different forms of boredom and their interplay. However, this constitutes an important construct validation attempt to different forms of academic boredom while considering their long-term development in secondary school. Including academic self-concept and examining relations to different forms of boredom further strengthens this validation effort while addressing the lack of research relating academic self-concept to different forms of boredom.

Succinctly stated, we tested the following assumptions:

Hypothesis 1 (H1): Three forms of academic boredom, namely, boredom intensity (BO-I), boredom due to underchallenge (BO-U) and boredom due to overchallenge (BO-O), can be empirically separated.

Hypothesis 2 (H2): Subject-specific BO-I, BO-U, and BO-O are cross-sectionally correlated with each other and with academic self-concept (for additional assumptions on expected effect sizes, see Table 2.1).

Open Research Question 1 (RQ1): Do BO-I, BO-U, BO-O, and academic self-concept show reciprocal relations over time?

Table 2.1

Assumptions on Correlational Effect Sizes for Study Variables

	BO-I	BO-U	BO-O
BO-U	$r_{Mathematics} \geq .10$, $r_{\text{German}} = ?$		
BO-O	$r_{Mathematics} \approx .30$, $r_{\text{German}} = ?$	$r_{Mathematics} > -.30$, $r_{\text{German}} = ?$	
ASC	$r_{Mathematics} > -.10^{a}$, $r_{\text{German}} > -.10$	$r_{Mathematics} > .30$, $r_{\text{German}} = ?$	$r_{Mathematics} > -.50$, $r_{\text{German}} = ?$

Note. Expectations derived from previous studies (Clem et al., 2021; Daschmann et al., 2011; Forsblom et al., 2021; Goetz & Frenzel, 2010; Krannich et al., 2019; Preckel et al., 2010). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

[a] Compared to verbal domains, effect sizes in mathematics are expected to be larger.

Method

Data from this study were acquired in the AVG-project, a five-cohort longitudinal study conducted in 43 different classrooms of five high-track German secondary schools located in two federal states. The AVG-project was sanctioned by the Supervision and Services Directorate of Rhineland-Palatinate (Aufsichts- und Dienstleistungsdirektion) (protocol number:32-03 405/29/05). For this longitudinal study, we focused on the first half of secondary school (grades 5-8) and the subject domains of mathematics and German. Our research design had four waves of questionnaire measurement, two in 5th grade (T1, shortly after transitioning to secondary school; T2, after mid-term evaluations), one in 6th grade (T3, after mid-term evaluations), and one in 8th grade (T4, after mid-term evaluations). Successive cohorts were followed from the school-term of 2005/2006 onward.

Participants

Our sample comprised N = 1,432 students (n = 761 identified as male, n = 667 identified as female, n = 4 without entry). Mean age in years was 10.24 (SD = 0.56 years) at T1, 10.54 (SD = 0.57) at T2, 11.56 (SD = 0.57) at T3, and 13.55 (SD = 0.57) at T4. Both parents’ highest educational degree was assessed to gauge socioeconomic status (sample statistics for mothers: 2.5% PhDs, 12.4% university graduates, 6.2% high-track secondary school graduates,10.3% high school graduates, 2.4% primary or middle school graduates, 0.1% without degree, 66.1% without entry). Only 10.4% stated another first language than German (65.9%; 23.7% without entry), with an average time of 7.63 years (SD = 2.31 years) of speaking German prior to the start of our investigation. With German as the study language, we did not expect any comprehension problems. Participant recruitment was voluntary with parental consent acquired beforehand.

Variables and Measures

We assessed three forms of academic boredom using 2-item short scales. In long-term assessments, short scales are frequently used because of economic reasons (i.e., limited testing time), and they have been proven viable alternatives to regular-sized scales considering construct representation for latent variables (Gogol et al., 2014). We measured boredom intensity (BO-I) with items from the *Achievement Emotions Questionnaire – Mathematics* (AEQ-M; Pekrun et al., 2005). We assessed boredom due to under- (BO-U) and overchallenge (BO-O) using self-designed items adapted from the PALMA-project (Pekrun et al., 2007). We assessed academic self-concept (ASC) by four items from the *Self-Description Questionnaire* (SDQ-II, short version; Marsh, 1990). All study variables were assessed separately for mathematics and German, with varied item wording in representation of domain-specific construct properties (Frenzel et al., 2007; Goetz et al., 2010). All self-report items were Likert-type, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). For descriptive information on all study variables, including item wording, see Table 2.2.

Table 2.2

Manifest Scale Descriptives and Reliabilities

		Descriptives[a]												Reliabilities[b]			
Variable / Scale	Number of items	*n*	% Missing	Min	Max	*M*	SE_{Mean}	*SD*	*Var*	Skewness	SE_{Skewness}	Kurtosis	SE_{Kurtosis}	*n*	% Missing	Spearman-Brown ρ	Cronbach's α
Mathematics																	
T1																	
BO-I	2	1,022	28.6	1	5	1.75	0.03	0.93	0.86	1.40	0.08	1.58	0.15	1,005	29.8	.70	-
BO-U	2	1,025	28.4	1	5	2.02	0.03	0.99	0.98	0.90	0.08	0.26	0.15	999	30.2	.70	-
BO-O	2	1,020	28.8	1	5	1.66	0.03	0.89	0.79	1.56	0.08	2.36	0.15	1,008	29.6	.77	-
ASC	4	1,025	28.4	1	5	3.93	0.03	0.88	0.77	-0.64	0.08	-0.07	0.15	963	32.8	-	.88
T2																	
BO-I	2	1,029	28.1	1	5	1.83	0.03	0.98	0.95	1.31	0.08	1.27	0.15	1,014	29.2	.73	-
BO-U	2	1,031	28.0	1	5	2.08	0.03	0.98	0.97	0.90	0.08	0.39	0.15	1,019	28.8	.69	-
BO-O	2	1,027	28.3	1	5	1.78	0.03	0.93	0.87	1.18	0.08	0.77	0.15	1,009	29.5	.75	-
ASC	4	1,033	27.9	1	5	3.76	0.03	0.97	0.93	-0.44	0.08	-0.65	0.15	998	30.3	-	.88
T3																	
BO-I	2	1,000	30.2	1	5	2.08	0.04	1.13	1.27	0.98	0.08	0.11	0.15	985	31.2	.79	-
BO-U	2	997	30.4	1	5	2.07	0.03	0.99	0.99	0.98	0.08	0.64	0.15	984	31.3	.72	-
BO-O	2	996	30.4	1	5	1.96	0.03	1.02	1.03	0.96	0.08	0.17	0.15	985	31.2	.79	-
ASC	4	1,001	30.1	1	5	3.47	0.03	1.08	1.17	-0.21	0.08	-0.89	0.15	967	32.5	-	.90
T4																	
BO-I	2	986	31.1	1	5	2.55	0.04	1.13	1.27	0.48	0.08	-0.54	0.16	963	32.8	.76	-
BO-U	2	975	31.9	1	5	2.25	0.03	1.05	1.10	0.80	0.08	0.08	0.16	963	32.8	.74	-
BO-O	2	979	31.6	1	5	2.32	0.04	1.18	1.39	0.65	0.08	-0.54	0.16	968	32.4	.82	-
ASC	4	986	31.1	1	5	3.19	0.04	1.13	1.28	-0.06	0.08	-0.94	0.16	944	34.1	-	.90

German																	
T1																	
BO-I	2	1,019	28.8	1	5	1.81	0.03	0.94	0.89	1.29	0.08	1.33	0.15	1,006	29.7	.77	-
BO-U	2	1,019	28.8	1	5	2.00	0.03	0.98	0.97	0.85	0.08	0.14	0.15	1,005	29.8	.74	-
BO-O	2	1,018	28.9	1	5	1.61	0.03	0.86	0.73	1.68	0.08	2.91	0.15	1,008	29.6	.79	-
ASC	4	1,021	28.7	1	5	3.77	0.03	0.88	0.78	-0.38	0.08	-0.40	0.15	986	31.1	-	.88
T2																	
BO-I	2	1,027	28.3	1	5	1.96	0.03	1.02	1.05	1.12	0.08	0.77	0.15	1,007	29.7	.76	-
BO-U	2	1,024	28.5	1	5	2.05	0.03	0.99	0.98	0.85	0.08	0.30	0.15	1,011	29.4	.72	-
BO-O	2	1,022	28.6	1	5	1.71	0.03	0.93	0.87	1.35	0.08	1.36	0.15	1,007	29.7	.84	-
ASC	4	1,029	28.1	1	5	3.59	0.03	0.93	0.86	-0.26	0.08	-0.46	0.15	991	30.8	-	.87
T3																	
BO-I	2	994	30.6	1	5	2.25	0.04	1.12	1.25	0.76	0.08	-0.22	0.15	981	31.5	.82	-
BO-U	2	993	30.7	1	5	2.24	0.03	1.01	1.01	0.62	0.08	-0.15	0.16	965	32.6	.70	-
BO-O	2	988	31.0	1	5	1.77	0.03	0.89	0.79	1.17	0.08	0.97	0.16	975	31.9	.76	-
ASC	4	995	30.5	1	5	3.44	0.03	0.96	0.92	-0.08	0.08	-0.62	0.16	956	33.2	-	.88
T4																	
BO-I	2	976	31.8	1	5	2.45	0.04	1.14	1.29	0.54	0.08	-0.48	0.16	960	33.0	.82	-
BO-U	2	967	32.5	1	5	2.38	0.03	0.99	0.98	0.52	0.08	-0.23	0.16	952	33.5	.75	-
BO-O	2	965	32.6	1	5	1.76	0.03	0.87	0.75	1.28	0.08	1.61	0.16	959	33.0	.73	-
ASC	4	977	31.8	1	5	3.40	0.03	0.97	0.94	-0.14	0.08	-0.60	0.16	948	33.8	-	.88

Note: Manifest scale raw data as item means. Descriptive statistics computed in SPSS Statistics 29 (IBM Corporation, 2022) with missing values either based on pairwise ([a]) or listwise deletion ([b]). Item responses took values from 1(*strongly disagree*) to 5 (*strongly agree*). Item wordings were altered depending on subject domain:

BO-I: I find [mathematics / German] to be boring (Item 1). I find it hard to stay awake during [mathematics / German] class out of sheer boredom (Item 2).

BO-U: When I'm bored in [mathematics / German] class, this is because the subject matter is so easy (Item 1). When I'm bored in [mathematics / German] class, this is because the teacher goes on about trivial points (Item 2).

BO-O: When I'm bored in [mathematics / German] class, this is because I cannot follow the teacher (Item 1). When I'm bored in [mathematics / German] class, this is because the [mathematics / German] subject matter is too difficult for me (Item 2).

ASC: I get good grades in [mathematics / German]. (Item 1). [Mathematics / German] is one of my best subjects (Item 2). I've always been good at [mathematics / German] (Item 3). In [mathematics / German] I learn fast (Item 4).

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

Methods of Analysis

Confirmatory Factor Analyses

We conducted subject-specific confirmatory factor analyses (CFAs) for the 3F-model of boredom (Feuchter & Preckel, 2022b) and ASC at T1-T4. Analyses were conducted in Mplus version 8.4 (Muthén & Muthén, 1998-2019) using the MLR estimator and the “type is complex”- and FIML options to account for nonnormality, nestedness, and missingness in our raw data. We used the effects coding method of model identification (Little et al., 2006) to obtain a comparable metric for latent variables and manifest scales. We evaluated model fit in reference to Hu and Bentler (1999), i.e., a given model fits the data well if CFI ≥ .95, RMSEA < .06, SRMR < .08. Additionally, we calculated McDonald’s ω (McDonald, 1999) as a marker of factor reliability. We compared CFA fit indexes for the 3F-model of boredom to a one-factor-model (1F-model, cf. H1). Comparative model fit was evaluated based on ΔCFI and ΔBIC, with higher CFI values and lower BIC values indicating superior fit for either modeling variant.

Correlations and Reciprocal Relations

Measurement Invariance Testing. Building on single-wave CFAs, we tested for measurement invariance over time in the 3F-model of boredom and in ASC, employing a step-up approach up to a level of strict invariance (Brown, 2015). Repeated item residuals were correlated over all waves of measurements to account for method variance in item indicators (i.e., correlated uniqueness, Geiser, 2011). We evaluated the apex level of invariance according to Chen (2007), with ΔCFI < .01 as the superordinate criterion for adequate invariance at a higher adjacent level. Subsequent analyses were conducted based on the highest level of measurement invariance for our study constructs to maximize statistical interpretations (Kleinke et al., 2017).

Reciprocal Effects Models (REMs). We computed subject-specific REMs (more specifically: factor cross-lagged panel models, Usami et al., 2019), spanning T1-T4. First, we estimated REMs based on the 3F-model of boredom (3-Variable-REMs). Second, we estimated REMs including three forms of boredom and ASC (4-Variable-REMs). Invariance constraints and correlated uniqueness were carried over from previous analyses and supplemented by autoregressive and cross-lagged relations for BO-I, BO-U, BO-O, and ASC.

Correlations. We extracted latent correlation matrices from 4-Variable-REMs of boredom and ASC to address correlational patterns (cf. H2). Latent correlational effect sizes were interpreted due to criteria by Gignac and Szodorai (2016; $\rho \geq .15$ = small, $\rho \geq .25$ = moderate, $\rho \geq .35$ = large).

Cross-lagged Relations. Given good model fit (Hu & Bentler, 1999), significant (i.e., $p < .05$) cross-lagged paths in 3- and 4-Variable REMs were compared (cf. RQ 1). Path coefficient effect size was evaluated by Keith's (2019) criteria ($\beta < .10$ = small, $\beta \leq .25$ = moderate, $\beta > .25$ = large).

Results

CFA

CFAs revealed good model fit for the 3F-model of boredom in German at each wave of measurement (i.e., CFI ≥ .95, RMSEA < .06, SRMR < .08, Hu & Bentler, 1999). The mathematics model also showed good approximation of the data at T1 and T2. At T3, the CFI indicated good model fit but the RMSEA statistic indicated a bad model fit (.084), just shy of an acceptable one (i.e., ≤ .08, Moosbrugger & Schermelleh-Engel, 2012). At T4, both CFI (.892) and RMSEA (.148) indicated inadequate data approximation. CFAs for ASC showed good model fits, except for the German scale at T2 and the mathematics scale at T4, both yielding acceptable fit by RMSEA (≤ .079).

The 1F-model lacked approximation of the underlying data for both subjects and at each wave of measurement (i.e., CFI ≤ .723, RMSEA ≥ .16, SRMR ≥ .10). For all wave-specific model comparisons, the 3F-model outperformed the 1F-model regarding CFI- and BIC values. See Table 2.3 for details.

Table 2.3

Fit Indexes and Factor Reliabilities for CFAs of Study Variables (T1-T4)

	Model Fit								Reliability	3F vs 1F	
Model	*n*	χ^2	*df*	SCF	CFI	RMSEA	SRMR	BIC	ω	ΔCFI	ΔBIC
T1											
Mathematics											
3F-model of boredom	1,025	11.290	6	1.6127	.994	.029 [.000-.055]	.016	16,258.295	.703 (BO-I) / .703 (BO-U) / .770 (BO-O)		
1F-model of boredom	1,025	284.020***	9	1.8765	.690	.173 [.156-.190]	.114	16,761.771	.728	-.304	503.476
ASC	1,025	2.759	2	1.4512	.999	.019 [.000-.068]	.006	9,618.219	.877		
German											
3F-model of boredom	1,021	11.731	6	2.0503	.994	.031 [.000-.056]	.014	15,501.447	.779 (BO-I) / .744 (BO-U) / .787 (BO-O)		
1F-model of boredom	1,021	252.642***	9	2.1068	.723	.163 [.146-.180]	.097	15,998.430	.796	-.271	496.983
ASC	1,021	1.985	2	2.5274	1.000	.000 [.000-.062]	.006	9,391.282	.887		
T2											
Mathematics											
3F-model of boredom	1,033	50.171***	6	1.4091	.961	.084 [.064-.107]	.034	16,774.529	.740 (BO-I) / .695 (BO-U) / .756 (BO-O)		
1F-model of boredom	1,033	372.603***	9	1.4801	.677	.198 [.181-.215]	.111	17,244.039	.739	-.284	469.510
ASC	1,033	0.879	2	1.8890	1.000	.000 [.000-.048]	.004	10,318.190	.884		
German											
3F-model of boredom	1,029	2.566	6	2.1646	1.000	.000 [.000-.022]	.007	16,095.792	.767 (BO-I) / .723 (BO-U) / .840 (BO-O)		
1F-model of boredom	1,029	309.244***	9	2.1004	.667	.180 [.163-.198]	.104	16,728.496	.800	-.333	632.704
ASC	1,029	14.205***	2	1.0238	.989	.077 [.043-.117]	.012	10,283.646	.871		
T3											
Mathematics											
3F-model of boredom	1,001	60.467***	6	1.3330	.961	.095 [0.74-.118]	.050	16,877.290	.807 (BO-I) / .727 (BO-U) / .791 (BO-O)		
1F-model of boredom	1,001	738.141***	9	1.1508	.479	.284 [.267-.302]	.150	17,634.965	.700	-.482	757.675
ASC	1,001	3.489	2	1.9648	.999	.027 [.000-.074]	.007	10,544.030	.898		
German											
3F-model of boredom	996	3.654	6	1.8285	1.000	.000 [.000-.030]	.013	16,179.632	.818 (BO-I) / .715 (BO-U) / .780 (BO-O)		

1F-model of boredom	996	287.977***	9	1.9795	.712	.176 [.159-.194]	.117	16,731.831	.741	-.288	552.199
ASC	995	0.880	2	1.6168	1.000	.000 [.000-.049]	.004	10,105.439	.877		
T4											
Mathematics											
3F-model of boredom	987	136.164***	6	0.8605	.892	.148 [.127-.170]	.059	17,168.885	.770 (BO-I) / .749 (BO-U) / .825 (BO-O)		
1F-model of boredom	987	670.064***	9	1.2162	.451	.273 [.255-.291]	.149	17,855.486	.689	-.441	686.601
ASC	986	14.266***	2	1.4521	.992	.079 [.044-.120]	.013	10,501.115	.903		
German											
3F-model of boredom	977	22.384**	6	1.2988	.986	.053 [.031-.077]	.025	15,859.787	.819 (BO-I) / .755 (BO-U) / .734 (BO-O)		
1F-model of boredom	977	616.505***	9	1.0708	.486	.263 [.245-.281]	.129	16,479.731	.709	-.500	619.944
ASC	977	5.867	2	1.0888	.996	.044 [.000-.088]	.008	9,959.006	.877		

Note. ** $p < .01$. *** $p < .001$. SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019). 90%-CI for RMSEA in brackets. Sample-size adjusted BIC. McDonald's ω calculated as $\frac{(\sum_{i=1}^{p} \lambda_{ij})^2}{(\sum_{i=1}^{p} \lambda_{ij})^2 + \sum_{i=1}^{p} e_i}$, where λ_{ij} is the standardized factor loading of item i on factor j, and e_{ij} is the standardized item residual for item i regarding factor j (Brunner et al., 2012; McDonald, 1999). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

Measurement Invariance Testing

Both the 3F-model of boredom and ASC fulfilled measurement invariance prerequisites over time for both mathematics and German. At each level, configural, metric, scalar, and strict individual model fit indexes aligned with Hu and Bentler's (1999) criteria. Furthermore, comparing models of adjacent measurement invariance did not lead to a critical reduction in model fit (i.e., $\Delta|CFI| > .01$, Chen, 2007). Therefore, we assumed a level of strict measurement invariance over time for subject-specific measurement models of academic boredom and ASC (see Table 2.4).

Table 2.4

Fit Indexes for Measurement Invariance- and Reciprocal Effects Models for 3 Forms of Academic Boredom

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Mathematics (*n* = 1,421)												
Measurement Invariance Over Time												
3F-model, configural	355.155***	150	1.2125	.969	.031 [.027-.035]	.041						
3F-model, metric	375.016***	159	1.2242	.967	.031 [.027-.035]	.043	configural vs. metric	20.110*	9	-.002	.000	+.002
3F-model, scalar	404.177***	168	1.2278	.964	.031 [.028-.035]	.043	metric vs. scalar	29.161***	9	-.003	.000	.000
3F-model, strict	417.208***	186	1.3151	.965	.030 [.026-.033]	.043	scalar vs. strict	24.612	18	+.001	-.001	.000
ASC, configural (*n* = 1,420)	219.844***	74	1.2433	.981	.037 [.032-.043]	.046						
ASC, metric (*n* = 1,420)	246.335***	83	1.2348	.979	.037 [.032-.043]	.053	configural vs. metric	26.476**	9	-.002	.000	+.007
ASC, scalar (*n* = 1,420)	262.876***	92	1.2240	.978	.036 [.031-.041]	.056	metric vs. scalar	15.640	9	-.001	-.001	+.003
ASC, strict (*n* = 1,420)	310.848***	104	1.3009	.973	.037 [.033- 042]	.063	scalar vs. strict	43.705***	12	-.005	+.001	+.007
Reciprocal Effects Model												
3-Variable-REM	463.111***	213	1.2969	.962	.029 [.025-.032]	.049						
4-Variable-REM	1,471.343***	674	1.2111	.953	.029 [.027-.031]	.058						
German (*n* = 1,417)												
Measurement Invariance Over Time												
3F-model, configural	189.964*	150	1.2797	.994	.014 [.006-.019]	.025						
3F-model, metric	197.622*	159	1.2992	.994	.013 [.006-.019]	.026	configural vs. metric	8.406	9	.000	-.001	+.001
3F-model, scalar	213.342*	168	1.3009	.993	.014 [.007-.019]	.026	metric vs. scalar	15.618	9	-.001	+.001	.000
3F-model, strict	255.756***	186	1.4025	.989	.016 [.011-.021]	.028	scalar vs. strict	29.520*	18	-.004	+.002	+.002
ASC, configural	133.430***	74	1.2140	.991	.024 [.017-.030]	.033						
ASC, metric	150.356***	83	1.1909	.990	.024 [.018-.030]	.038	configural vs. metric	17.059*	9	-.001	.000	+.005
ASC, scalar	181.430***	92	1.1786	.987	.026 [.021-.032]	.040	metric vs. scalar	32.647***	9	-.003	+.002	+.002
ASC, strict	246.094***	104	1.2756	.979	.031 [.026-.036]	.051	scalar vs strict	49.565***	12	-.008	+.004	+.009
Reciprocal Effects Model												
3-Variable REM	314.900***	213	1.3795	.984	.018 [.014-.023]	.041						

4-Variable REM	1,216.369***	674	1.2754	.963	.024 [.022-.026]	.051

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. SCF = Scaling correction factor for MLR estimator in Mplus. 90%-CI for RMSEA are provided in brackets. Additional $\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR and are reported for reasons of transparency.

Correlations

Mathematics Results

All latent cross-sectional correlations, except for $r_{\text{BO-U,BO-O}}$ (T3) and $r_{\text{BO-I,BO-U}}$ (T4), were statistically significant (see Table 2.5). $r_{\text{BO-I,BO-U}}$ showed a large effect size in grade 5 before strongly decreasing over the course of the study to a small nonsignificant effect in grade 8 (.80/.73/.45/.09). $r_{\text{BO-I,BO-O}}$ increased over time, moving from an expected effect size near .30 to larger ones at T2-4 (.38/.45/.45/.56). $r_{\text{BO-U,BO-O}}$ was unexpectedly small and positive in grade 5 before moving through an insignificant to an expected negative correlation of large size at T4 (.18/.19/-.06/-.35). $r_{\text{ASC,BO-I}}$ was negative and large and increased in size between grades 6 and 8 (-.35/-.36/-.38/-.54). $r_{\text{ASC,BO-U}}$ was positive and increased from small to large over the course of the investigation (.10/.11/.34/.61). $r_{\text{ASC,BO-O}}$ was expectedly negative and large in size (-.45/-.54/-.58/-.70).

German Results

All latent cross-sectional correlations, except for $r_{\text{ASC,BO-U}}$ (T2) and $r_{\text{BO-U,BO-O}}$ (T4), reached statistical significance (see Table 2.5). $r_{\text{BO-I,BO-U}}$ showed a large positive effect, which decreased throughout the investigation (.83/.80/.67/.45). Large effect sizes for $r_{\text{BO-I,BO-O}}$ did not show much variation throughout our investigation period (.44/.50/.37/.46). $r_{\text{BO-U,BO-O}}$ continuously decreased from large to small, showing an insignificant correlation in grade 8 (.43/.40/.15/.08). $r_{\text{ASC,BO-I}}$ was expectedly negative and of large size at all waves of measurement (-.41/-.36/-.38/-.48). Interestingly, effect sizes were comparable across both domains (cf. Table 2.1). $r_{\text{ASC,BO-U}}$ was small and negative in grade 5 before shifting to small positive correlations in grades 6 and 8 (-.11/-.04/.12/.12). $r_{\text{ASC,BO-O}}$ showed all-increasing large negative effect sizes (-.38/-.41/-.49/-.52).

Table 2.5

Latent Cross-sectional Correlations

	BO-I				BO-U				BO-O			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
						Mathematics						
BO-U	.80	.73	.45	.09$^{n.s.}$								
BO-O	.37	.45	.45	.56	.18	.19	-.05$^{n.s.}$	-.35				
ASC	-.35	-.36	-.38	-.54	.10	.11	.34	.61	-.45	-.54	-.58	-.70
						German						
BO-U	.83	.80	.67	.45								
BO-O	.44	.50	.37	.46	.43	.40	.15	.08$^{n.s.}$				
ASC	-.41	-.38	-.38	-.48	-.11	-.04$^{n.s.}$	.12	.12	-.38	-.41	-.49	-.52

Note. Taken from 4-Variable REMs. Green cells indicate correlations in expected direction, with hue indicating effect size (small vs. moderate vs. large). Red cells indicate (small) correlations in unexpected direction. Grey cells indicate nonsignificant correlations. For remaining cells, no effect size assumptions applied.

For a complete table of manifest and latent correlations for all study variables, see Table A2.1 of the Appendix.

Reciprocal Relations

The 3- and 4-Variable-REMs approximated the data well for both subject domains (see Table 2.4).

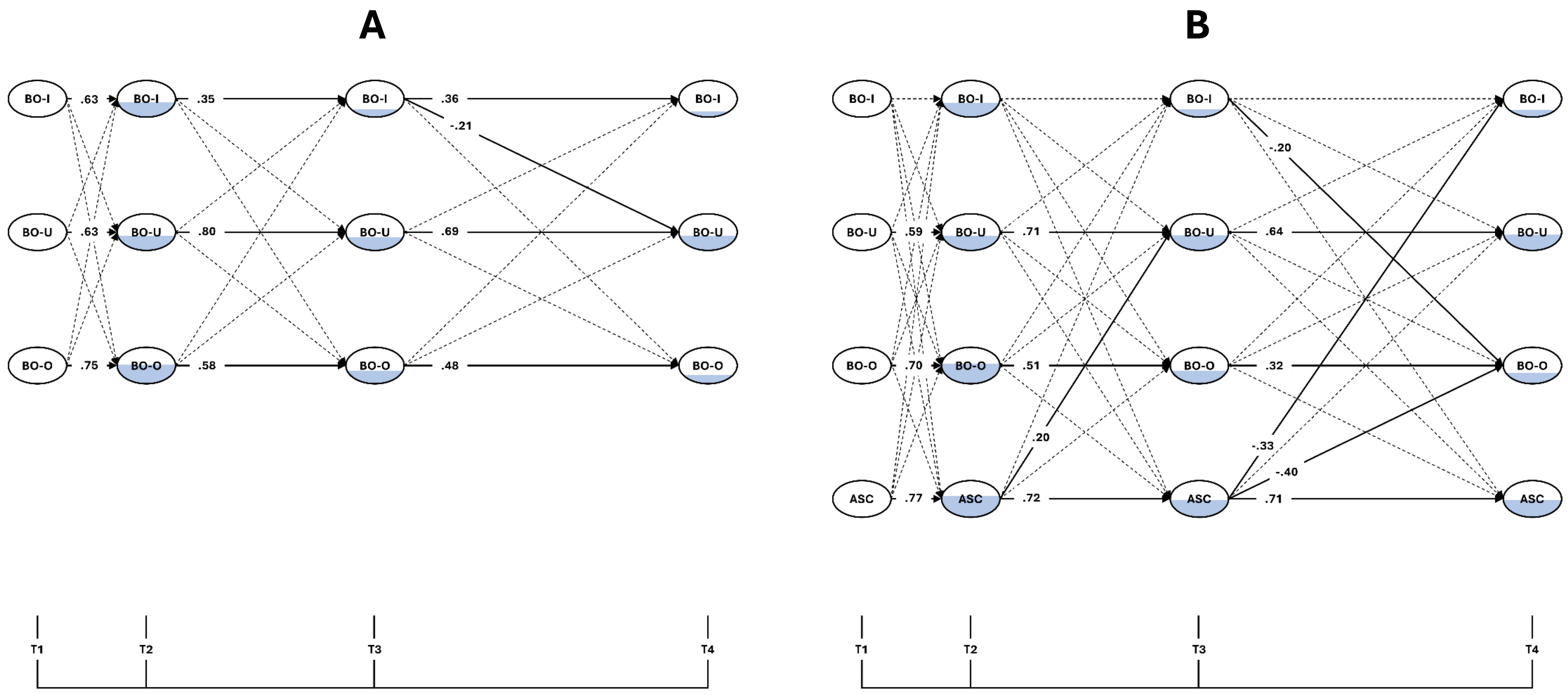
Mathematics Results

We found significant autoregressive paths within the mathematics 3-Variable-REM for all forms of academic boredom. BO-I in grade 6 negatively predicted BO-U in grade 8 ($\beta = -.21$, moderate effect). No other cross-lagged paths for academic boredom reached statistical significance.

Including ASC in 4-Variable REMs resulted in insignificant autoregressive paths for BO-I. Regarding cross-lagged paths, previous ASC predicted BO-U in grade 6 ($\beta = .20$), as well as BO-I ($\beta = -.33$, large effect) and BO-O in grade 8 ($\beta = -.40$). Additionally, BO-I in grade 6 no longer significantly predicted BO-U in grade 8. Instead, BO-I in grade 6 negatively predicted BO-O ($\beta = -.20$) in grade 8. No other cross-lagged paths reached statistical significance (see Figure 2.1).

Figure 2.1

Significant Standardized Path Coefficients from 3- (A) and 4-Variable-REMs (B) in Mathematics



Note. Circle fillings indicate explained variance in endogenous variables (R^2). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

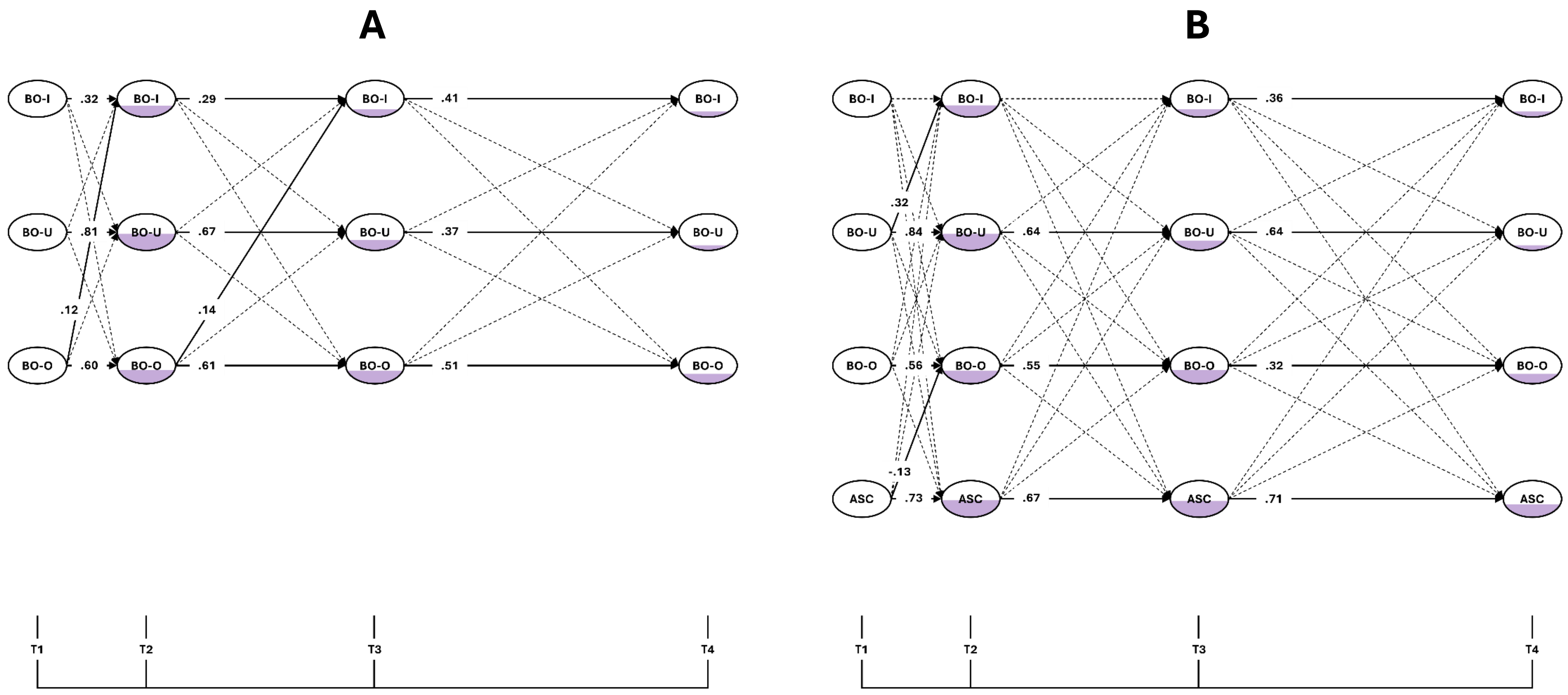
German Results

We found significant autoregressive paths within the German 3-Variable-REM for all forms of academic boredom. BO-O in grade 5 positively predicted later BO-I in both grade 5 ($\beta = .12$, moderate effect) and 6 ($\beta = .14$). No other cross-lagged paths for academic boredom reached statistical significance.

Including ASC in 4-Variable-REMs returned all-significant autoregressive paths, except for BO-I from grades 5 to 6. In addition, two cross-lagged paths reached statistical significance. BO-U positively predicted BO-I at mid-term evaluations ($\beta = .32$, large effect) in grade 5. Covering the same time span, ASC negatively predicted later BO-O ($\beta = -.13$). BO-O no longer predicted subsequent BO-I at early waves of measurement (see Figure 2.2).

Figure 2.2

Significant Standardized Path Coefficients from 3- (A) and 4-Variable-REMs (B) in German



Note. Circle fillings indicate explained variance in endogenous variables (R^2). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

For a complete list of correlation and path coefficient estimates from 3- and 4-Variable REMs, including standard errors, see Table A2.2 of the Appendix.

Discussion

The present study provided results strengthening a three-factor-structure for academic boredom, separating intensity from under- and overchallenge aspects. We also explored cross-sectional relations between boredom and academic self-concept, and their common development over the first half of secondary education in German high-track schools. We found limited evidence for reciprocal relations over time but clear indications of developmental changes in construct interplay bearing subject-specific tendencies.

Boredom as a Many-faceted Emotion

In the one vs. many boredom(s)-debate, results from this study add to a many boredoms-perspective. Our 3F-model showed superior factorial validity to a 1F-model across our investigation. Findings generalized over subject domains. Our results showed that different forms of academic boredom can be separated by CFA methodology. Functional accounts argue in favor of unitary conceptualizations of boredom (e.g., Elpidorou, 2021). This view is limited to boredom’s motivational component (i.e., motivating change in activity by signaling situational discomfort) and neglects important variation in cognitive components, such as appraisals or attributions (Scherer, 2009). In challenge-driven academic contexts, reasons for experiencing boredom form important, distinguishable latent factors (Acee et al., 2010; Daschmann et al., 2011).

Moreover, we provided correlational evidence on the 3F-model’s congruent and discriminant validity regarding academic self-concept. Boredom intensity and ASC were negatively correlated in both subjects (Clem et al., 2021; Krannich et al., 2019). Boredom due to underchallenge correlated positively with ASC in mathematics (Goetz & Frenzel, 2010; Preckel et al., 2010) but negatively with boredom due to underchallenge in German in grade

5. Boredom due to overchallenge and ASC were consistently and negatively correlated in both subjects. Interestingly, all boredom-ASC-correlations increased in absolute size over the course of our study, indicating that self-concept appraisals move closer to boredom constructs in a nomological sense as secondary school progresses. Future work should explore these correlational tendencies further, including older age groups.

Considering the interrelations of different forms of boredom, we found subject specificities in our data. The most pressing changes in correlational patterns considered boredom due to under- and overchallenge. For mathematics, an expected negative correlation was only evident in grade 8 (Daschmann et al., 2011; Goetz & Frenzel, 2010). In grade 5, BO-U and BO-O were positively correlated. We argue that this is due to students continuously acquiring meta-emotional competencies (Frenzel & Stephens, 2013) in early secondary school. Consequently, they can accurately distinguish between under- and overchallenging situations by grade 8. For German, this proposed process appears to move at a slower pace as BO-U-BO-O-correlations were consistently positive throughout our investigation. However, they decreased in size, with T4 only showing an insignificant association. For both subjects, BO-U-BO-O-correlations did not differ in 3- and 4-Variable-REMs (cf. Table A2.2 of the Appendix), indicating that under- and overchallenge differentiation occurs independently of ASC development. Taken together, older students were more likely to report being bored due to either under- or overchallenge, while younger students reported being bored due to both forms of suboptimal challenge.

No Reciprocities Between Boredom and Academic Self-concepts

Although we found predictive paths between boredom and academic self-concept constructs within REMs for both domains, no form of academic boredom showed reciprocal effects with ASC. For mathematics, we found cross-lagged effects for ASC predicting different forms of boredom between grades 6 and 8. For German, ASC only predicted later boredom due to overchallenge from early to mid-5th grade. Taken together, reciprocal relations only showed sporadic patterns of statistical significance, limiting interpretations. Interestingly, the inclusion of ASC-predictions in REMs led to decreased stability in boredom intensity, particularly during periods, in which ASC predicted boredom. This could speak to boredom intensity development being overpowered by self-concept predictions, stressing ASC's role as protective factor to boredom experiences (Pekrun et al., 2010). Conversely, and despite high cross-sectional correlations, ASC appears to be unaffected by boredom. This finding could speak to a diminished severity of boredom regarding self-evaluative aspects of subjective control. Comparable to our results, previous studies found one-sided ASC-on-boredom-effects over time (Clem et al., 2021; Forsblom et al., 2021). Clem et al. (2021), for instance, suggested disengagement from scholar learning in response to boredom to hinder change in ASC due to a lack of self-evaluation. Overall, our results oppose the notion of reciprocal relations between boredom and ASC, as also proposed by CVT's feedback loops (Pekrun, 2019).

Limitations

Besides our large student sample (N = 1,432) and long investigation period (grades 5-8), generalizability of our results is limited. Participants only represent a select portion of German secondary education, namely, high-track schools from two out of 16 federal states.

From a methodological standpoint, using 2-item short scales for different forms of academic boredom is not ideal. Despite this limitation, the 3F-model of boredom showed strict measurement invariance over time and returned good model fits for REMs. In addition, we measured boredom subject-specifically to provide cross-domain comparisons that are often neglected in educational research.

In addition, this study focused on control appraisals included in CVT and did not include value appraisals. This limits theoretical implications regarding antecedents to boredom as proposed by CVT (Pekrun, 2006).

Regarding results for the mathematics domain, interpretations are limited regarding T4 because of nonoptimal model fit for the 3F-model in CFAs. Still, model fit for the 3F-model far surpassed that of the competing 1F-model. Furthermore, all longitudinal analyses using the 3F-model returned good model fits for both subject domains.

Future Directions

Results presented here were derived from analyses with underlying linearity assumptions for investigated correlational variable interplay. In various theoretical accounts, however, curvilinear relations of subjective control and boredom have been proposed (Csikszentmihalyi, 1975; Pekrun et al., 2010; Raffaelli et al., 2018). Subjective control and boredom can hence be assumed to show a *U*-shaped relation with lowest experiences of boredom for tasks that are experienced as neither too easy nor too difficult (i.e., *U*'s bottom stretch). Under- and overchallenge would accordingly be associated with high boredom (i.e., *U*'s right or left "legs"). Westgate and Wilson (2018) provide evidence for a *U*-shaped relation between boredom and perceived task difficulty. However, they did not use domain-specific assessments of boredom, and their study was not conducted in an educational setting. Available findings from educational settings point to linear relations between boredom and control (Pekrun et al., 2010). A possible reason for this is that classrooms constitute constrained environments in which few students perceive themselves as being highly in control. Future studies should explore linearity in boredom-control-relations (e.g., by locally weighted scatter plot smoothers, see Cleveland et al., 1992) and employ nonlinear analytical methodology (e.g., by exponential transformation of monotonous curvilinear correlations, see Tukey, 1977), if required.

This work showed the separability of different forms of academic boredom by factor analytical methodology. In addition to such factorial distinctions, future research should also address profile distinctions. Goetz et al. (2014), for instance, found five different types of boredom (indifferent boredom, calibrating boredom, searching boredom, reactant boredom, and apathetic boredom) using latent profile analysis on experience sampling data. Nett et al. (2011) found different boredom coping profiles (reappraisers vs. criticizers vs. evaders), focusing on response mechanisms to boredom experiences. These accounts are promising alternatives to CFA-based distinctions that merit further investigation and could provide important insights in search of adequate academic boredom interventions.

Implications for Research and Practice

The present study's results show that different forms of academic boredom should be considered in future theorizing. In CVT's latest iteration (Pekrun et al., 2023), activity emotions, such as boredom, are suggested to also have prospective and retrospective equivalents. Along this line, our study proposes boredom due to under- and overchallenge as relevant retrospective forms of activity boredom. Besides, subject specificities need further consideration regarding boredom development. Furthermore, developmental stage needs consideration, as differentiation between different forms of boredom increased with age.

As a practical implication, results from educational studies should, in general, be made available to educational practitioners. More specifically, subject-specific emotional development should be taught at school so that both learners and teachers can build on shared knowledge on psychological dynamics within their classroom. Teachers, for example, should be aware that attributions of one's own boredom to differential subjective control or challenge are not adequately developed in students in early secondary school. As students get older, however, classrooms may profit from repeated control ratings by students to approximate an optimal difficulty that keeps experiences of under- and overchallenge in balance. In giftedness research, for instance, underchallenge has often been identified as a source of boredom and underachievement in high-ability students (Feldhusen & Kroll, 1991; Obergriesser & Stoeger, 2015). It is important to note, however, that ability grouping as a classroom composition intervention, has not been found to be effective in reducing boredom due to underchallenge over time (Feuchter & Preckel, 2022b). Instead, nonoptimal individual challenge or control perceptions could be tackled by promoting intraindividual performance feedback and offering difficulty choices inside the same classroom. “Leveling up” through various difficulty stages will, in our opinion, concurrently promote individual ability self-concepts, enhance flow experiences, boost enjoyment, and reduce boredom at school.

Chapter 3 – Study 2[4]

Reducing Boredom in Gifted Education – Evaluating the Effects of Full-time Ability Grouping

Markus D. Feuchter and Franzis Preckel

Department of Psychology, University of Trier

Abstract

Ability grouping provides an advanced learning environment for gifted students, possibly buffering them from common long-term increases in academic boredom. We present a 3.5-year longitudinal investigation, spanning four waves of measurement (T1-T4), featuring secondary school students (grades 5 through 8) from five different German schools with full-time ability grouping (*N* = 1,861, 55.4% male). We used propensity score matching and latent growth curve modeling to determine the effects of class type on three types of boredom (intensity of boredom, boredom due to underchallenge, boredom due to overchallenge) in two subject domains (mathematics and German). We separated the effects of intervention effectiveness and efficacy, analysing full and matched sample data. All types of boredom increased over time in both subjects. Ability grouping significantly reduced the intensity of boredom in mathematics in special classes for the gifted ($\beta_{effectiveness} = -.158$, $\beta_{efficacy} = -.206$) but had no further effects on the development of subject-specific academic boredom.

[4] Chapter 3 is the republished accepted version of an article published in the Journal of Educational Psychology. Permission by the American Psychological Association was obtained on 08/17/2023 (see Appendix 5).

Keywords: ability grouping, boredom, giftedness, longitudinal, propensity score matching

Educational Impact and Implications Statement

This 3.5-year study examines the development of boredom in mathematics and German in Grade 5 to 8 in secondary school students. Students either attended regular classes or special classes for the gifted (i.e., full-time ability grouping). Comparing boredom development across class types, we found only limited evidence for benefits of special classes for the gifted regarding the development of boredom. Rather, boredom increased in both class types over time. Despite other favorable effects of special classes for the gifted, tackling boredom does not seem to be one of them. Direct boredom prevention deserves increased attention throughout secondary school independent of class-type.

Reducing Boredom in Gifted Education – Evaluating the Effects of Full-time Ability Grouping

Grouping high ability students into special classrooms has been discussed as a means to prevent underchallenge and boredom in intellectually gifted populations (Bar-On, 2007; Plucker et al., 2004). Intellectual giftedness is often assumed to be associated with increased boredom in regular classrooms (*boredom hypothesis*, see Feldhusen & Kroll, 1991). As the boredom hypothesis is highly accepted in practice, ability grouping into special classrooms is often advocated as one solution to prevent boredom because it leads to a more adequate learning environment for high ability students (e.g., Little, 2012; Rogers, 2007). However, very few studies have put this assumption under scrutiny, and those that did returned mixed results (e.g., Gjesme, 1977; Larson & Richards, 1991; Preckel et al., 2010).

Available findings on the boredom hypothesis and the effects of high ability grouping are scarce and point to limited effects of ability-grouped classes (Hornstra et al., 2017; Preckel et al., 2010). Past results remain inconclusive, however, as no study we know of provided a systematic and convincing framework to test the boredom hypothesis and its implications for successful boredom prevention. Requirements of such a study include: (a) a longitudinal design spanning several years that would allow the investigation of long-term effects of ability grouping, (b) an adequate control group with comparable demographic, cognitive, and affective-motivational properties in settings without high ability grouping, (c) a confirmatory methodological approach (i.e., latent modeling) to measure boredom and its development, and (d) consecutive investigations of different measures of boredom in different subject domains (e.g., numeric and verbal subjects) as research points to different types and to the domain specificity of boredom.

Therefore, the present study aims to close the remaining research gaps and to provide conclusive evaluative results on ability grouping's potential to reduce boredom in gifted populations.

In the following, we provide an evaluation study of full-time ability grouping, conducting a longitudinal comparison of boredom development of intellectually gifted students attending either ability-grouped or regular classes. Regarding boredom experiences, we focused on two subject domains, namely mathematics and German, and differentiated between three types of boredom (i.e., *intensity of boredom*, *boredom due to underchallenge*, and *boredom due to overchallenge*). We evaluated full-time ability grouping's potential to be both effective (i.e., beneficial in naturalistic settings) and efficacious (i.e., beneficial in a controlled setting; Ernst & Pittler, 2006) at preventing gifted students' boredom at school. The latter was achieved by using propensity score matching methodology (e.g., Rosenbaum & Rubin, 1983).

Academic Boredom and the Role of Perceived Challenge

Boredom can be characterized as "an aversive state of wanting, but being unable, to engage in satisfying activity" (Eastwood et al., 2012, p. 482). Experiences of boredom are accompanied by task-unrelated thoughts, perceived time expansion, and reduced agency (Raffaelli et al., 2018). Students report feeling bored up to 58% of the time during class (Larson & Richards, 1991; Nett et al., 2011), making it an emotion of exceptional importance within the academic realm. As such, *academic boredom* is part of Pekrun's (2006) taxonomy of achievement emotions. Tied directly to the classroom, experiences of academic boredom are domain-specific and therefore vary between subjects taught at school (Goetz et al., 2007). Pekrun's *control value theory* (CVT; 2006; see also Pekrun et al., 2010) provides a framework for the antecedents of experiences of academic boredom. According to CVT, academic boredom is the result of a lack of subjective value (i.e., succeeding is perceived as irrelevant) in combination with either low or high subjective control (i.e., succeeding is perceived as either too easy or too difficult; Pekrun, 2006). Other situational antecedents of academic boredom include characteristics of the task (e.g., monotonous, highly repetitive, understimulating) and classroom instruction (e.g., lack of clarity and structuring; Goetz et al., 2019).

The emergence of boredom in learning and achievement contexts is closely tied to the subjective amount of challenge that is present in a given situation. *Flow theory* (Csikszentmihalyi, 1975, 1990), for instance, builds upon *U*-shaped relations between stimulation and learning capacity. According to flow theory, differential subjective task difficulty results in different emotional experiences: too easy tasks elicit boredom, just optimally difficult tasks elicit joy (or flow), and too hard tasks elicit anxiety. In recent work, however, the relation between task difficulty and boredom has been shown to be more complex. Both under- and overchallenging tasks, for instance, have been associated with increased boredom, replicating the aforementioned *U*-shaped trend (Westgate & Wilson, 2018, Study 3) in line with CVT's control- based assumptions. There appears to be only a small window for optimal experience. If task difficulty surpasses individual capabilities, boredom results from being overchallenged, whereas boredom due to being underchallenged results from capabilities surpassing task difficulty (Daschmann et al., 2011; Krannich et al., 2019; Raffaelli et al., 2018). However, experiences of under- and overchallenge are not two endpoints of a one-dimensional continuum but rather empirically separable constructs (Acee et al., 2010). They are differentially correlated, for instance, with lower self-efficacy for self-regulated learning (boredom due to overchallenge) and lower school achievement, as measured by grades (boredom due to underchallenge; Tze et al., 2014). Taken together, subjective challenge is at the center of subjective experiences of boredom in academic contexts. Therefore, both, boredom due to under- and overchallenge should be featured in investigations of academic boredom, especially in gifted research.

Assumptions on Gifted Students' Academic Boredom

In their synthesis of the psychological science behind talent, giftedness, and expertise, Subotnik, Olszewski-Kubilius and Worrell (2011) define giftedness as the manifestation of potential and performance that is clearly at the upper end of a distribution in a talent domain. Accordingly, gifted individuals are those who demonstrate outstanding levels of aptitude (i.e., exceptional ability to reason and learn) or competence (i.e., performance or achievement in top 10% or rarer) in one or more achievement domains. An achievement domain includes any structured area of activity with its own symbol system (e.g., mathematics, language) and/or a set of sensorimotor skills (e.g., painting, dance; National Association for Gifted Children, n.d., para. 5). The manifestation of potential and performance at the upper end of the distribution in the academic-intellectual achievement domain is named intellectual giftedness. Within this work, we use the term "gifted" or "giftedness" to refer solely to its intellectual component.

Aside from having higher cognitive abilities than their nongifted peers (Terman, 1922), gifted students have also been found to be more motivated at school (Wirthwein et al., 2019), to have higher levels of performance-related self-concepts (Košir et al., 2016; Litster & Roberts, 2011), and to be more persistent and ready to work to exhaustion (Winner, 1998). These learning characteristics, albeit favorable, foster a commonly accepted belief that gifted students are underchallenged in regular classrooms, which, in turn, leads to increased boredom (Feldhusen & Kroll, 1991). This *boredom hypothesis*, although well aligned with findings on the role of perceived challenge in academic boredom (see above), has rarely been challenged scientifically, using a relevant sample. Nevertheless, it forms the basis of many theoretical assumptions on gifted students' boredom. For instance, in line with the boredom hypothesis, Little (2012) points to a possible *U*-shaped relation between cognitive ability and boredom – similar to *U*-shaped relations between subjective challenge and boredom, discussed above – potentially making gifted students more vulnerable to experiences of boredom compared to average-ability peers. According to Pekrun and colleagues (2010), on the other hand, gifted students' high competence may protect them against boredom. While some studies have found higher boredom in gifted students in elementary school (e.g., grades 4-5, Stambaugh, 2017), as well as secondary school (e.g., grades 5-9, Larson & Richards, 1991), others found no support of the boredom hypothesis for similar age groups (e.g., K-6 students; Feldhusen & Kroll, 1991; 9th grade students, Preckel et al., 2010).

In spite of mixed empirical results, acceptance of the boredom hypothesis in practice has led researchers to promote full-time ability grouping in special classrooms as an educational intervention for preventing gifted students' boredom (Bar-On, 2007; Plucker et al., 2004). With respect to other academically relevant outcomes, meta-analyses support the effectiveness of gifted ability grouping with regard to academic achievement (Steenbergen-Hu et al., 2016), or social facilitation (Rogers, 2007). Opposed to these beneficial effects, a trend of reduced academic self-concept (ASC) is discussed in the literature as the big-fish-little-pond effect (BFLPE; Marsh & Parker, 1984). The BFLPE states that negative contrast effects of class- average ability will result in reduced ASC in special ability-grouped classes for the gifted. While this holds on a meta-analytic level (Fang et al., 2018), newer studies found no significant contrast effects in special ability-grouped classes for gifted secondary school students (for the verbal domain, see Herrmann et al., 2016; for the mathematics domain, see Preckel et al., 2019). Also, existing contrast effects can be compensated for by positive assimilation effects due to enhanced status of ability-grouped classes (e.g., for 5th graders, Preckel & Brüll, 2010). In sum, whereas ability grouping's effects on achievement and self-concept have been thoroughly studied, its potential to reduce boredom has largely been neglected in educational research, especially regarding its development over time.

Correlates and Consequences of Academic Boredom and its Development Over Time

Generally, boredom has been shown to be negatively associated with several academic outcomes (e.g., elaboration or exam grades), with correlation effect sizes of $r \approx -.30$. (Goetz et al., 2019. In a recent meta-analysis (Tze et al., 2016), for instance, boredom had a significant negative overall effect on motivational and learning outcomes ($\bar{r}$= -.24). This effect was slightly higher for secondary students ($\bar{r}$= -.26) as compared to tertiary school students ($\bar{r}$= -.23). On a longitudinal level, academic boredom and achievement have reciprocal effects on each other, such that increased boredom is associated with a decrease in later performance and vice versa, both for secondary school students (Pekrun et al., 2017) and college students (Pekrun et al., 2014). Besides achievement and learning-related outcomes, boredom also impacts individuals' psychological well-being and health. For instance, increased boredom is associated with decreased satisfaction with life (Todman, 2013) and deteriorating interpersonal relationships in adults (Watt & Vodanovich, 1999). Furthermore, boredom is associated with increased alcohol abuse, as well as higher rates of depression and anxiety (LePera, 2011). Specific to gifted populations, a growing body of recent dissertational work has shown boredom to be associated with self-harm in high school (Williams, 2015), experiences of disengagement and underachievement in middle school (Baker, 2016), as well as increased rates of depression, anxiety, and disruptive behavior in elementary school (Stambaugh, 2017).

Longitudinal studies indicate that academic boredom generally increases over time, a tendency that can be observed in primary education (grades 4-6, Hornstra et al., 2017; grades 2- 5, Vierhaus et al., 2016), secondary education (grades 4-7, Vierhaus et al., 2016; grades 5-9,Pekrun et al., 2017; throughout first half of 9th grade, Preckel et al., 2010; grades 8-12, Weybright et al., 2019), and tertiary education (college level, Pekrun et al., 2014). Furthermore, a recent meta-analysis of longitudinal studies by Scherrer and Preckel

(2019) revealed that intrinsic value and domain-specific academic self-concepts decrease over the course of the school career (Glass's Δ = -.11 over 1.65 years on average). According to the assumptions of CVT, this related developmental trend might concurrently boost experiences of boredom through reduced intrinsic control and/or value (Pekrun et al., 2010; see also Goetz et al., 2012; Putwain et al., 2018; Ruthig et al., 2008).

Preckel and colleagues (2010) found little change in boredom trajectories in mathematics for 9^{th} grade students in regular classes over the course of one school semester. In 9^{th} grade ability-grouped classes, however, boredom due to under- and overchallenge did change over time, with decreasing boredom due to underchallenge alongside increasing boredom due to overchallenge. This observed change pattern supports the notion that ability grouping can be effective in providing an appropriate level of challenge for gifted students. However, the study of Preckel et al. (2010) only spanned half a year, did not use confirmatory testing, and had no suitable control group (i.e., comparable students in both class types). Hornstra and colleagues (2017) applied propensity score matching and investigated boredom development over a two year time span in primary school (grades 4-6) with a latent modelling approach and a suitable control group. They found lower levels of boredom for gifted students in part-time ability- grouped classes when compared to both regular and full-time ability-grouped classes (β = -.50), revealing considerable potential for grouping programs with respect to boredom prevention. Boredom development did not differ between the three conditions. However, the authors used only one boredom measure, and did not include boredom due to under- or overchallenge as outcomes.

Isolating Treatment Effects Using Propensity Score Matching (PSM)

One of our aims for this study was to evaluate the effects of full-time ability grouping as an educational intervention in terms of its *effectiveness* and *efficacy* (for definitions of effectiveness and efficacy, see Ernst & Pittler, 2006; Flay, 1986; Flay et al., 2005; Kellam & Langevin, 2003). Effectiveness-studies examine intervention effectiveness based on authentic field sample data. They can commonly be achieved in cooperation with schools. Efficacy-studies examine intervention effectiveness in a controlled setting. They are optimally realized by conducting randomized controlled trials (RCTs) with a randomized assignment for all participants to either the treatment or control condition (see Baber, 1994; Schulz et al., 2010). RCTs cannot always be realized within an active educational system, due to ethical reasons and self-selection for special educational programs, like special classes for the gifted. An educational intervention's efficacy, though not attainable through RCTs, can still be estimated, nevertheless. Propensity score matching (Rosenbaum & Rubin, 1983, 1984, 1985) provides a suitable methodology for this purpose and has grown increasingly popular in educational research in recent years (for examples of PSM studies in educational contexts see Becker et al., 2014; Hornstra et al., 2017; Preckel et al., 2019; Wirthwein et al., 2019). Generally speaking, a propensity score (PS) represents "the probability that a particular case would be assigned or exposed to a treatment condition" (Ridgeway et al., 2014, p. 1) – in our case, the probability of either attending a regular class or a special class for the gifted. This probability is estimated for every individual of a given sample as a function of a chosen set of matching variables and their interaction terms. These probabilities can help reduce bias due to confounding variables in field data, if a large number of potentially relevant covariates (or *matching variables*) are used for PS estimation. Matching participants from inherently different comparison groups (e.g., regular vs. special classes for the gifted) based on their PSs, can hence provide post-hoc statistical control by balancing out potentially influencing variables. Therefore, field data can be retrospectively adapted to resemble a quasi-experiment with balanced treatment and control groups via PSM (Fan & Nowell, 2011), providing best possible isolation of treatment effects (i.e. efficacy). Prerequisites for optimal functioning are a large enough original sample and a sufficient number of potentially influential matching variables.

Up until now, Hornstra and colleagues (2017) conducted the only study we know of that applied PSM for the investigation of the effects of gifted ability grouping in grades 4-6. As we alluded to above, study findings did not support full-time ability grouping as effective to reduce boredom as compared to regular classes. Despite this singular result, research on the boredom hypothesis and the adequacy of ability-grouped classrooms as a suitable prevention strategy for boredom in academia, remains inconclusive, especially considering secondary education.

The Present Study

The first part of this study is a contribution to the small pool of longitudinal studies on academic boredom development. Existing works indicate that boredom at school increases over time (e.g., Hornstra et al., 2017; Pekrun et al., 2014; Vierhaus et al., 2016; Weybright et al., 2020), whereas overall motivation, including intrinsic value and subjective control, decreases over the course of one's school career (Scherrer & Preckel, 2019), possibly further facilitating experiences of boredom. Therefore, our first research question and pertaining hypothesis were:

RQ1: How do experiences of subject-specific academic boredom develop over time?

H1: Experiences of subject-specific academic boredom increase over time.

Second, the evaluative part of this study was achieved by conducting analyses using two different data sets: (a) a sample acquired in a longitudinal field study (also forming the data base for RQ1/H1), and (b) a matched subsample of this full sample derived via PSM methodology. Intervention effectiveness was inferred by using data from (a), whereas intervention efficacy was inferred by using data from (b). Research questions 2 and 3, alongside pertaining hypotheses, represented our evaluation interests, encompassing the notion that full-time ability grouping is an adequate anti-boredom intervention (Bar-On, 2007; Little, 2012; Plucker et al., 2004; Rogers, 2007):

RQ2: Does full-time ability grouping of students into either regular classes or special classes for the gifted predict the developmental trajectories of subject-specific academic boredom in the field (i.e., is full-time ability grouping in special classes for the gifted effective)?

H2: Ability grouping of students into either regular classes or special classes for the gifted predicts the developmental trajectories of subject-specific academic boredom, such that boredom is either buffered (i.e., lower or no increase) or even reversed (i.e., decrease) in special classes for the gifted.

RQ3: Does full-time ability grouping of students into either regular classes or special classes for the gifted predict the developmental trajectories of subject-specific academic boredom when controlling for potentially influencing factors via PSM (i.e., is full-time ability grouping in special classes for the gifted efficacious)?

H3: Ability grouping of students into either regular classes or special classes for the gifted predicts the developmental trajectories of subject-specific academic boredom, such that boredom is either buffered (i.e., lower or no increase) or even reversed (i.e., decrease) in special classes for the gifted.

Method

Procedure

Data used in this work were collected as part of a longitudinal study (AVG-project), examining the motivational, social, and affective development of German students over the course of secondary school and beyond in the country’s top academic track, called “Gymnasium”. The AVG-project was approved by the Supervision and Services Directorate of Rhineland-Palatinate (Aufsichts- und Dienstleistungsdirektion) (protocol number:32-03 405/29/05). An important part of this project was the investigation of the effects of full-time ability grouping on various educational outcomes after an installment at the beginning of secondary education. Target subject domains were mathematics and German. Both are highly important main subjects taught throughout primary and secondary school in the German educational system. They represent different achievement domains (i.e., numerical vs. verbal) and are often associated with differential interests and academic self-concepts, even though performance measures are usually correlated (Möller & Marsh, 2013). In total, five successive cohorts from five different German schools participated in this 15-year-long investigation from 2005 to 2020. All schools offered regular as well as special classes for the gifted from Grade 5 to Grade 10 (afterwards students selected courses that were open to all students). Those students who entered the special classes applied for these classes, usually based on their parents' initiative. Due to this self-selection, not all gifted students attended special classes. The schools employed similar multistage selection procedures for the special classes including prior academic achievement, intelligence test results (IQ ≥ 120), teacher observations of students' behavior during one day of probationary class, and interviews with parents. Applicants were then selected in a conference of teachers, school psychologists, and school board members based on a partly compensatory strategy (i.e., high achievement could partly compensate for an IQ of slightly below 120 and vice versa). Special classes programs featured in this study were all conceptualized as a combination of acceleration (e.g., skipping 9th grade in unison) and enrichment methods and differed from regular schooling in several ways. For instance, both English and Latin were taught as second languages from 5th grade on, with French added in 6th grade. From 7th grade on, geography and history class were taught in English. Furthermore, practice-based enriched mathematics and science education was a main priority (e.g., via additional experimental and IT-classes from 5th grade on).

Our investigation period spanned the first half of German secondary education (grades 5 to 8). Before entering Grade 5, all students had attended a regular primary school. Thus, the study started at the beginning of secondary school. There were four waves of measurement (T1- T4) with self-report questionnaires (T1: two weeks into 5^{th} grade; T2/3/4: directly after mid-term evaluations in 5^{th}/6^{th}/8^{th} grade). In addition, after three months in 5^{th} grade, students were tested with an intelligence test. Data collection took place in class in group sessions. Students participated voluntarily. Parental consent was provided for all student participants. Student data across measurement waves were matched by using a pseudo-anonymized identification code (i.e., self-generated alphanumeric code according to a predefined procedure).

Participants

Full Sample

The original field sample comprised a total of 1,861 students (55.4% male, 44.1% female, 0.5% without entry); 1,471 of them attended regular classes during the investigation period (53.4% male, 46.2% female, 0.4% without entry). The remaining 390 students attended special classes for the gifted (63.1% male, 36.2% female, 0.8% without entry). In total, students attended 58 different classes. Mean age for all participants was 10.14 years (*SD* = .61) at T1, 10.47 years (*SD* = .63) at T2, 11.49 years (*SD* = .63) at T3, and 13.48 years (*SD*= .63) at T4. Students predominantly grew up speaking German, with 9.6% of the sample stating a different first language. For these students, the mean time speaking German was 7.47 years (*SD* = 2.43) at the study’s start.

Matched Sample

The matched sample comprised 546 students (61.2% male, 37.9% female, 0.9% without entry) from the full sample, with 273 attending regular classes (61.9% male, 37.4% female, 0.7% without entry) and 273 attending special classes for the gifted (60.4% male, 38.5% female, 1.1% without entry). Students in the matched samples stemmed from all 58 original classes. Mean age for all matched students was 9.92 years (*SD* = .67) at T1, 10.32 years (*SD* = .70) at T2, 11.35 years (*SD* = .70) at T3, and 13.33 years (*SD* = .71) at T4. The predominant language background was German. For the 7.3% of the matched sample who did not state German as their first language, the average time speaking German was 6.97 years (*SD* = .30) at the study's start.

Variables and PSM

Subject-specific Academic Boredom

We assessed subject-specific boredom using self- report measures of three different types of boredom that were treated as separate latent variables in further analyses (see below). Due to aforementioned project design, we examined *mathematics* and *German* as subject domains. For all boredom measures, participants responded on 5-point Likert-type rating scales, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). By replacing the word "mathematics" with "German", German-specific item framing was provided. This procedure is common practice in cross-domain classroom-based self-report studies (e.g., Goetz et al., 2012). All three types of boredom were assessed with 2-item short scales because of limited questionnaire space due to the assessment of multiple constructs in the project. Short scales have been shown to be a reliable alternative to full scales in the affective-motivational realm (Gogol et al., 2014). Subsequent item examples are translations from the original German wording.

Intensity of Boredom. We used two items from the Achievement Emotions Questionnaire – Mathematics (AEQ-M; Pekrun et al., 2005), originally developed in the PALMA-project (see Pekrun et al., 2003). Items were: "I find [mathematics / German] to be boring" and "I find it hard to stay awake during [mathematics / German] class out of sheer boredom". As the scales contained two items (Eisinga et al., 2013), the Spearman-Brown coefficient (ρ) was used as the appropriate measure of scale reliability (ρ's for math-specific boredom: ρ_{T1} = .71 / n_{T1} = 1,270, ρ_{T2} = .74 / n_{T2} = 1,286, ρ_{T3} = .79 / n_{T3} = 1,264, ρ_{T4} = .75 / n_{T4} = 1,235; ρ's for German-specific boredom: ρ_{T1} = .79 / n_{T1} = 1,274, ρ_{T2} = .78 / n_{T2} = 1,277, ρ_{T3} = .82 / n_{T3} = 1,257, ρ_{T4} = .82 / n_{T4} = 1,231). However, boredom assessed via the AEQ-M does not differentiate between boredom due to under- vs. overchallenge.

Boredom Due to Under- and Overchallenge. We used four additional items, originally developed in the PALMA project (see Pekrun et al., 2003, 2007), to measure *boredom due to underchallenge* (2 items) and *boredom due to overchallenge* (2 items). Thereby, we expanded general boredom assessment (i.e. intensity) by adding boredom measures based on the degree of individually perceived mismatch between class difficulty and one's own ability. For boredom due to underchallenge, items were: "My experience is that I'm bored in [mathematics / German] class because the subject matter is so easy" and "My experience is that I'm bored in [mathematics / German] class because the teacher goes on about trivial points" (ρ's for math- specific boredom: ρ_{T1} = .71 / n_{T1} = 1,265, ρ_{T2} = .70 / n_{T2} = 1,288, ρ_{T3} = .74 / n_{T3} = 1,265, ρ_{T4} = .75 / n_{T4} = 1,236; ρ's for German-specific boredom: ρ_{T1} = .76 / n_{T1} = 1,270, ρ_{T2} = .72 / n_{T2} = 1,279, ρ_{T3} = .69 / n_{T3} = 1,244, ρ_{T4} = .76 / n_{T4} = 1,225). The two items measuring boredom due tooverchallenge were: "When I'm bored in [mathematics / German] class, this is because I can't follow the teacher" and "When I'm bored in [mathematics / German] class, this is because the [mathematics / German] subject matter is too difficult for me" (ρ's for math-specific boredom: ρ_{T1} = .76 / n_{T1} = 1,274, ρ_{T2} = .76 / n_{T2} = 1,278, ρ_{T3} = .78 / n_{T3} = 1,264, ρ_{T4} = .84 / n_{T4} = 1,237; ρ's for German-specific boredom: ρ_{T1} = .80 / n_{T1} = 1,275, ρ_{T2} = .82 / n_{T2} = 1,278, ρ_{T3} = .77 / n_{T3} = 1,253, ρ_{T4} = .75 / n_{T4} = 1,231).

Other Matching Variables

To attain the matched sample presented above, we estimated individual propensity scores (PS) based on T1-data of a total of 32 matching variables, including demographics, cognitive ability, motivational variables, and social variables. Demographic information covered participants' age (in months), gender, main language (*German* vs. *other*), socioeconomic background (i.e. mother's and father's highest educational degree and profession), as well as school and cohort attended. Matching on cognitive ability was realized using the composite IQ score (α = .93 / n = 919)[5] from the Cognitive Ability Test for Grades 4-12 (KFT 4-12+R; Heller & Perleth, 2000), which was administered via group testing in the classroom throughout the first semester of 5th grade. Scale data used for matching on motivational and social variables included self-esteem, academic interest in mathematics and German, social self-concepts (of acceptance / of assertion), academic self-concepts (general, mathematics, verbal), and achievement goals (performance approach / avoidance, mastery; in mathematics and German). All three types of subject-specific academic boredom were included as matching variables. For details (e.g., instruments, descriptives, reliabilities) on all matching variable scales, see Table A3.1.

PSM Procedure

Our PSM procedure consisted of two separate steps. In Step 1, we estimated generalized boosted regression PS in R Statistics version 3.6.1 for every student, using the Toolkit for Weighting and Analysis of Nonequivalent Groups ("twang"; Ridgeway et al., 2015; for details on boosted regression see McCaffrey et al., 2004). Implementing PS via "twang" allows users to include matching variables with different scales of measurement at the same time, while also treating missing values as an additional source of information. Step 2 was the actual matching of students from the different class types based on the PS generated via "twang". We used the "MatchIt" package for R Statistics (Ho et al., 2011), employing 1:1 nearest neighbor matching. We integrated exact matching on school attended

[5] Sample size and Cronbach's α reported here are based on the 90-minute short version. Additionally, 21 students completed the full-version of the KFT 4-12+R (α = .86). Cronbach's α of this subsample, however, is not entirely trustworthy because it was based on a reduced item pool (i = 116 instead of 195), with several item variances yielding a value of zero.

into this matching scheme, in order to eliminate possible context effects. That way, only students attending the same school were matched with each other. To assess balance after the matching process, we calculated standardized differences for all matching variables between regular and special classes. We also compared subsamples from the full and matched samples based on standardized differences, in order to assess appropriate representation after PSM. Values of standardized differences not surpassing the upper threshold of a small effect size (e.g., $d \leq .20$; Cohen, 1992; see also Sedlmeier & Renkewitz, 2013) were considered indicative of acceptable balance between groups. For balance checks of all 32 matching variables, see Table 3.1. For further details on the PSM process, please refer to Appendix 3.

Data Analyses

Analyses for investigating our research questions were embedded in a structural equation modeling (SEM) approach and executed in Mplus version 8.4 (Muthén & Muthén, 1998 - 2019). We adapted the analytical strategy of Preckel and colleagues (2019) to measure the effects of class type on long-term motivational and achievement development to experiences of boredom in mathematics and German. This resulted in a three-step process. First, we tested the measurement invariance of our boredom scales over time and across class types. Second, given at least a level of partially scalar measurement invariance over time (Kleinke et al., 2017), we test H1, considering boredom development within a latent growth curve modeling (LGCM) approach. Third, given a level of at least partially scalar measurement invariance across class types, we address H2 and H3, considering the effects of ability grouping in regular vs. special classes on boredom trajectories, operationalizing class type as a manifest predictor to previously established latent growth factors. For all analyses, full information maximum likelihood estimation (MLR) was employed. Missing value pattern analyses, spanning T1-T4, showed a range of covariance coverage of 49% –

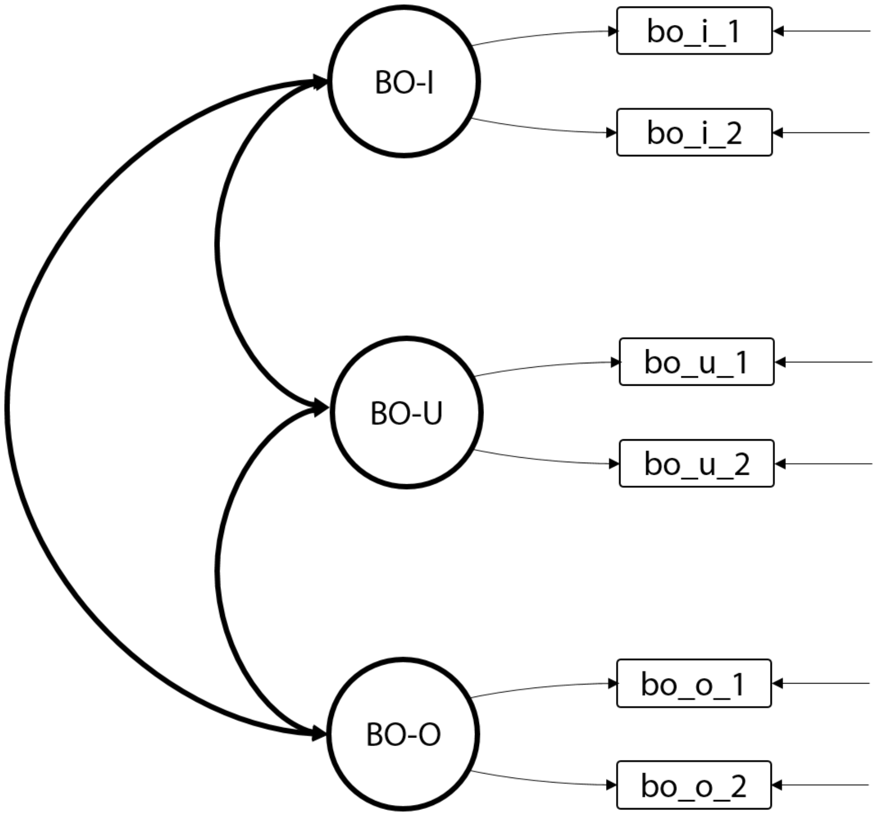
72% for the full sample (matched sample: 48% – 71%) for both subject domains. To cope with these missing values, we used the Mplus FIML procedure in model estimation.

Measurement Invariance Testing

Measurement Invariance Over Time. A prerequisite for LGCM is a level of at least partially scalar measurement invariance over time, allowing for interpretations of latent means (Kleinke et al., 2017). Therefore, we performed longitudinal invariance tests on a first-order factor model (Brunner et al., 2012), including intensity of boredom, boredom due to underchallenge, and boredom due to overchallenge, for the full sample as well as for class type subsamples (see Figure 3.1; for respective fit indexes for T1-T4, including McDonald's ω's for factor reliabilities, see Table A3.3). We used a step-up approach (Brown, 2015), testing longitudinal confirmatory factor analyses (CFA) with correlated uniqueness between repeated item measures (Geiser, 2011) for (a) configural, (b) metric, and (c) scalar invariance. For model identification, we used T. D. Little et al.'s (2006) effects coding method in order to obtain comparable metrics for latent and manifest variables. Furthermore, we used Mplus's "type is complex" option to correct for biased standard error (*SE*) estimation due to hierarchical data structure, with individual students nested in 58 classrooms in total. Results were evaluated based on Chen's (2007) propositions on GFI differences (metric invariance holds, if ΔCFI < -.01, ΔRMSEA < .015, and ΔSRMR < .03; scalar invariance holds, if ΔCFI < -.01, ΔRMSEA < .015, and ΔSRMR < .01).

Figure 3.1

Schematic First-order Factor Model of Boredom



Note. BO-I = intensity of boredom. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. bo_i_1, bo_i_2, bo_u_1, bo_u_2, bo_o_1, and bo_o_2 represent item indicators to latent boredom factors (see method section for details on variable measurement). Latent boredom factors were allowed to correlate.

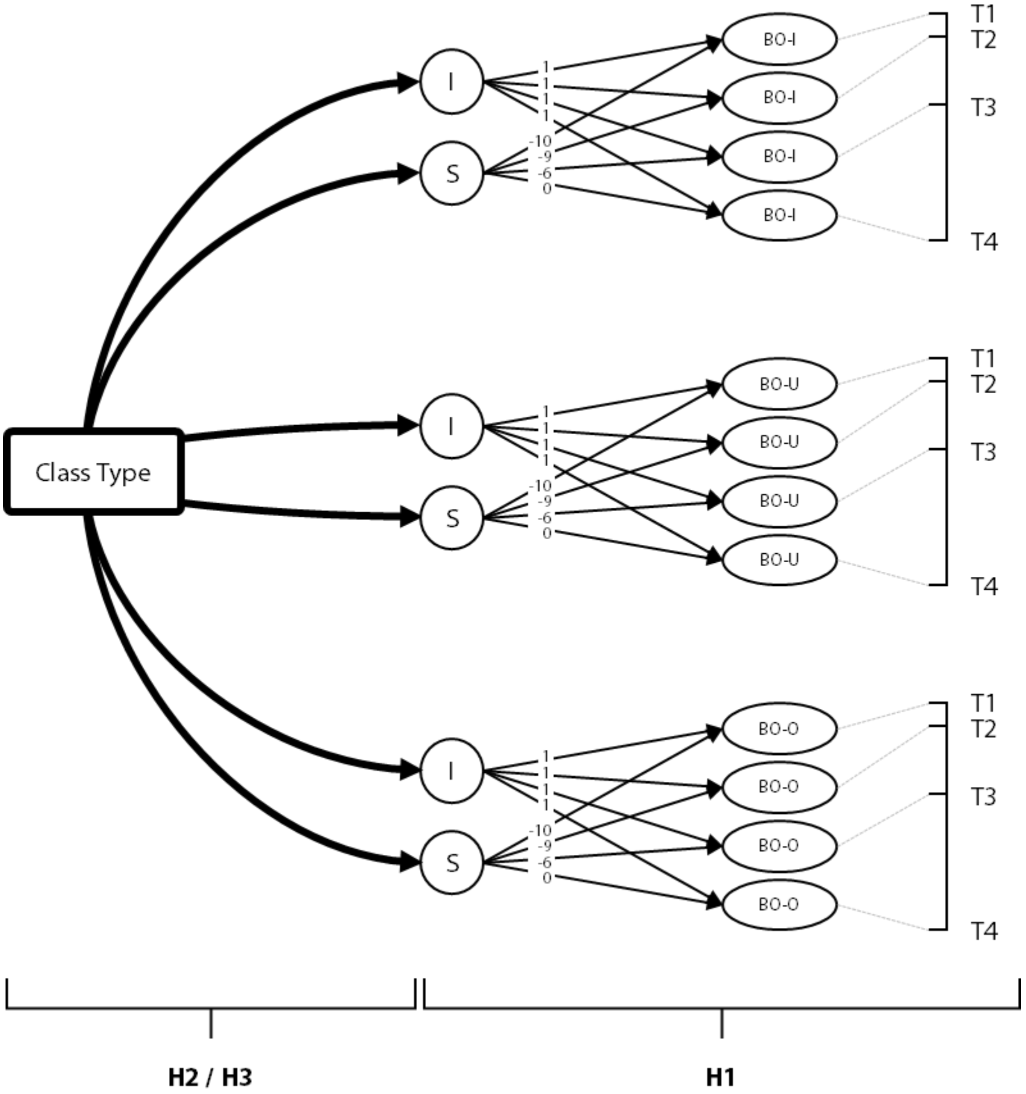
Measurement Invariance Across Class Types. When subsuming different groups, in our case regular and special classes for the gifted, into a categorical predictor variable, scale invariance across those groups is required to ensure comparability of measurement results (Byrne, 2012; Jöreskog, 1971). Measurement invariance models were based on the first-order factor model introduced earlier (see Figure 3.1), for either mathematics or German in the full sample. We estimated multiple group comparison models (Kleinke et al., 2017; see also Byrne, 2012; Wang & Wang, 2012) in Mplus version 8.4 (Muthén & Muthén, 1998 - 2019) to test for invariance across class types at T1, T2, T3, and T4, employing a step-up approach (Brown, 2015), up to a level of scalar invariance. Results were again evaluated according to Chen (2007).

Latent Growth Curve Modeling

To test for an increase in different types of boredom (H1), we included latent intercept and slope factors in the previously established first-order factor model (see Figure 3.2 for a schematic depiction). Latent intercept factor loadings were held constant at $\lambda_{intercept,T1} = \lambda_{intercept,T2} = \lambda_{intercept,T3} = \lambda_{intercept,T4} = 1$, while latent slope factor loadings were based on measurement waves within study design ($\lambda_{slope,T1} = -10$, $\lambda_{slope,T2} = -9$, $\lambda_{slope,T3} = -6$, $\lambda_{slope,T4} = 0$). We placed the origin of time at T4 because we were primarily interested in the boredom constellations at the end of the growth period (Biesanz et al., 2004).

Figure 3.2

Schematic Latent Growth Curve Model of Boredom (Representing H1), Including Predictions by Class Type (Representing H2 / H3)



Note. I = latent intercept factor. S = latent slope factor. BO-I = intensity of boredom. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. T1-4 represent respective waves of measurement. Latent slope factor loadings represent time codes for T1-T4. Item indicators, correlated uniqueness, correlations between latent boredom factors at each wave of measurement, correlations between latent growth factors, and residual variances are omitted.

Examining intervention effectiveness (H2) and efficacy (H3) for special classes for the gifted, we added class type (0 = regular classes, 1 = special classes) as an exogenous, manifest predictor, regressing on all latent intercept and slope factors (see Figure 3.2). Testing the model in the full sample provided information on the intervention effectiveness; testing it in the matched sample provided information on the intervention efficacy. The same specifications used for longitudinal measurement invariance testing (i.e., correlated uniqueness, effects coding for model identification, and “type is complex”) were applied to all latent growth curve models. To evaluate model fit, we turned to Hu and Bentler’s (1999) recommendations on GFI values to provide an acceptable balance between Type-I and Type-II error (CFI ≈ .95, RMSEA ≈ .06, and SRMR ≈ .08).

Results

First, we briefly evaluate the PSM solution and we summarize the results of all auxiliary analyses (i.e. measurement invariance testing). Details are given in Appendix 3. Second, we present results of latent growth curve analyses in more detail, since they are the main focus of the present study and directly relate to our three hypotheses.

Propensity Score Matching and Subsample Comparisons

Table 3.1 shows central tendency and dispersion for all 32 matching variables, including subsample comparisons using appropriate standardized difference values according to respective scale of measurement (Sedlmeier & Renkewitz, 2013).

Table 3.1

Descriptives and Subsample Comparisons for 32 Matching Variables at T1

		Central tendency (Dispersion)				Standardized effect size			
		Full sample		Matched sample		Subsample comparison			
Variables (level)	Rel. Inf. in %	RC (n = 1,471)	GC (n = 390)	RC (n = 273)	GC (n = 273)	GC_{full} vs. RC_{full}	$GC_{matched}$ vs. $RC_{matched}$	GC_{full} vs. $GC_{matched}$	RC_{full} vs. $RC_{matched}$
Propensity score (metric)	-	.13	.52	.27	.38	**1.38**	**.51**	**.48**	**-1.23**
Demographics									
Age, in months (metric)	30.14	128.23 (5.53)	123.28 (7.79)	125.68 (6.90)	124.22 (7.86)	**-.64**	-.19	-.12	**.46**
Gender[a] (metric)	1.61	1.46 (.50)	1.36 (.48)	1.38 (.49)	1.39 (.49)	**-.21**	.03	-.05	.18
Main language background[a] (metric)	.02	1.14 (.34)	1.09 (.29)	1.10 (.30)	1.11 (.31)	-.15	.03	-.05	.11
Mother's highest educational degree[b], child report (ordinal)	1.26	4 (2)	5 (1)	5 (2)	5 (2)	**4.27*****	1.67	.65	-.68
Mother's highest educational degree[b], self-report (ordinal)	11.87	4 (2)	5 (1)	4 (2)	4 (2)	**5.47*****	1.04	.68	-1.04
Mother's profession[c], child report (nominal)	.58	5	6	5	6	**.18**	**.17**	.08	.09
Father's highest educational degree[b], child report (ordinal)	.00	4 (2)	5 (2)	5 (1)	5 (1)	**4.18*****	1.19	.94	-1.17
Father's highest educational degree[b], self-report (ordinal)	1.06	4 (2)	5 (1)	5 (2)	5 (2)	**5.85*****	1.76	.55	-.53
Father's profession[c], child report (nominal)	.38	5	7	9	7	**.13**	**.22**	.07	.06
School (nominal)	8.84	-	-	-	-	**.21**	0[f]	**.19**	**.13**
Cohort[d] (nominal)	.61	3	3	3	3	.05	.05	**.11**	.09
Cognitive Ability									
Composite IQ[e] (metric)	31.25	107.83 (11.26)	120.88 (11.06)	116.07 (11.70)	117.60 (10.74)	**1.18**	.14	**.30**	**-.73**
Motivational and Social Variables (all metric)									
General									
Academic self-concept	.86	4.03 (.68)	4.23 (.68)	4.20 (.63)	4.23 (.70)	**.30**	.04	.01	**-.25**
Social self-concept of acceptance	.00	1.62 (.80)	1.59 (.81)	1.67 (.83)	1.66 (.90)	-.04	-.01	-.09	-.06
Social self-concept of assertiveness	1.44	2.31 (.99)	2.24 (.99)	2.30 (.98)	2.33 (1.06)	-.07	.03	-.09	.01
Self-esteem	.68	4.08 (.66)	4.21 (.67)	4.18 (.58)	4.18 (.69)	.20	-.01	.05	-.16
Math-specific									
Academic interest	.54	3.47 (.93)	3.73 (.92)	3.52 (.99)	3.72 (.90)	**.29**	**.23**	.01	-.05
Academic self-concept	.77	4.05 (.79)	4.36 (.71)	4.24 (.80)	4.33 (.72)	**.44**	.12	.05	**-.24**
Approach achievement motivation	.01	3.02 (1.03)	2.83 (1.12)	2.94 (1.09)	2.85 (1.14)	-.17	-.07	-.02	.08
Avoidance achievement motivation	1.10	3.53 (1.03)	3.19 (1.24)	3.36 (1.17)	3.19 (1.24)	**-.27**	-.13	-.01	.16
Competence motivation	.71	3.87 (1.01)	3.91 (1.04)	3.85 (1.10)	3.90 (1.03)	.05	.04	.01	.01
Intensity of boredom	.00	1.75 (.92)	1.66 (.92)	1.81 (.99)	1.70 (.96)	-.10	-.11	-.05	-.07

Boredom due to being underchallenged	.55	2.02 (0.98)	2.03 (1.08)	2.15 (1.11)	2.07 (1.09)	.01	-.07	-.04	-.13
Boredom due to being overchallenged	.00	1.65 (0.88)	1.51 (.81)	1.48 (.82)	1.55 (.85)	-.18	.08	-.05	.20
German-specific									
Academic interest	1.03	3.40 (.90)	3.60 (.91)	3.48 (.90)	3.60 (.91)	**.29**	.13	.00	-.09
Academic self-concept	.51	3.91 (.80)	4.14 (.73)	4.10 (.76)	4.15 (.74)	**.32**	.07	-.01	**-.24**
Approach achievement motivation	.60	3.02 (1.03)	2.95 (1.13)	2.97 (1.14)	3.01 (1.14)	-.06	.04	-.06	.04
Avoidance achievement motivation	.20	3.52 (1.09)	3.25 (1.25)	3.44 (1.19)	3.29 (1.25)	**-.21**	-.13	-.03	.07
Competence motivation	2.08	3.75 (1.04)	3.71 (1.15)	3.78 (1.08)	3.69 (1.15)	-.03	-.08	.02	-.03
Intensity of boredom	.20	1.80 (.94)	1.83 (1.06)	1.73 (.99)	1.86 (1.10)	.03	.12	-.03	.08
Boredom due to underchallenge	.39	2.00 (.98)	2.02 (1.10)	1.92 (1.02)	2.10 (1.14)	.02	.16	-.07	.08
Boredom due to overchallenge	.00	1.61 (.86)	1.40 (.76)	1.45 (.79)	1.45 (.84)	**-.28**	.00	-.08	.18

Note. Rel. Inf. = relative influence on propensity score estimation. RC = regular classes. GC = special classes for the gifted.

Descriptive information on school attended is omitted for reasons of anonymization.

Appropriate values for central tendency, dispersion, and standardized effect sizes are presented according to respective scale of measurement (Sedlmeier & Renkewitz, 2013):

- for metric variables, *M*, *SD*, and Cohen's *d*'s (calculated as $\frac{M_{\mathrm{GC}} - M_{\mathrm{RC}}}{SD_{\mathrm{GC}}}$ or $\frac{M_{\mathrm{GC,\ full}} - M_{\mathrm{GC,\ matched}}}{SD_{\mathrm{GC,\ full}}}$) are presented
- for ordinal variables, *Mdn*, interquartile range, and *z*-values (derived from Mann-Whitney *U*-tests) are presented are presented
- for nominal variables, Mode and *w*-values (derived from χ^2-tests; calculated as $\sqrt{\frac{\chi^2}{N}}$) are presented.

d's > .20, significant *z*-values (*** $p < .001$), and *w*'s > .10 are highlighted in bold face.

[a] Gender (1 = „male“, 2 = „female“) and main language background (1 = „German“, 2 = „other“) were treated as metric variables, since they both took only two different values (Sedlmeier & Renkwitz, 2013).

[b] Mother's and father's highest educational degree were assessed two ways, a) retrospectively, via child report at the end of secondary school (grade 12/13), and b) via separate parent self-report questionnaires. These variables took values from 1-6, with 1 = „no degree", 2 = „primary- or middle school graduate", 3 = „high-school graduate", 4 = „Abitur" (high-track secondary school graduate), 5 = „university degree", 6 = „PhD".

[c] Mother's and father's profession took values from 1-11, with 1 = "student (school)", 2 = "student (university)", 3 = "trainee", 4 = "worker", 5 = "employee", 6 = "senior employee", 7 = "highly qualified employee", 8 = "employee in leading position", 9 = "self-employed", 10 = "stay-at-home", 11 = "pensioner".

[d] Cohort took values from 1-5, representing individual 5th grade starting terms, with 1 = „2005/2006", 2 = „2006/2007", 3 = "2007/2008", 4 = "2008/2009", 5 = "2009/2010".

[e] Composite IQ as measured by the Cognitive Ability Test for Grades 4-12 (Heller & Perleth, 2000).

[f] Only students from the same school were matched via exact matching.

In the original full sample, students attending special classes for the gifted were younger (*d* = -.64, moderate effect, Cohen, 1992) and had a higher IQ (*d* = 1.18, large effect) compared to students from regular classes. Also, the gender distribution in special versus regular classes showed an effect > |.20| in favor of boys (*d* = -.21). Regarding boredom, the original regular and special classes only differed with respect to boredom due to overchallenge in German, with regular classes showing higher scale means (*d* = -.28). Post-matching, only mother's (*w* = .17) and father's profession (*w* = .22), and academic interest in mathematics (*d* = .23) showed differences surpassing the upper bound of a small effect size in favor of special classes (for *w*, values ≤ .10 are considered small, Cohen, 1992). Overall, appropriate balance was achieved by PSM. When comparing special classes from the full and matched samples, a moderate difference in IQ was evident (*d* = .30), with a higher mean level in the original subsample. The opposite pattern emerged when comparing the original and matched regular classes, yielding a large difference in IQ (*d* = -.73), with higher levels in the matched regular classes, which also were moderately younger than in the original sample (*d* = .46). Additionally, regular class students had higher general (*d* = -.25) and subject-specific academic self-concept (*d* mathematics: *d* = -.24; German: *d* = -.24) in the matched sample, compared to the original sample.

Measurement Invariance Testing

Regarding measurement invariance over time, boredom scales in mathematics and German showed scalar measurement invariance for both subject domains, fulfilling the prerequisite to LGCM and the interpretation of latent means (Kleinke et al., 2017). Assumptions held for both the full samples (including class type subsamples) and the matched samples (see Table A3.4, for details). Regarding measurement invariance across class types, findings indicated (partially) scalar measurement invariance for both subject domains, legitimizing class type to be used as an exogenous predictor of latent growth

(Jöreskog, 1971). For mathematics, in the full sample, scalar invariance held for T1-T3, with partially scalar invariance for T4. The same invariance pattern was evident in the matched sample. For German, in the full sample, scalar invariance held at every wave of measurement. In the matched sample, scalar invariance held at T2 and T3. For T1 and T4, partially scalar invariance was achieved (see Tables A3.5 and A3.6, for details).

Latent Growth in Subject-specific Boredom

Mean factor scores and standard deviations (*SD*) for intensity of boredom, boredom due to underchallenge, and boredom due to overchallenge taken from models of scalar measurement invariance over time are included in Table 3.2 (for class type subsample statistics see also Table A3.7). Given the observed factor scores, a linear increasing trend in all three types of boredom seems plausible.

Table 3.2

Mean Factor Scores and Standard Deviations for Latent Variables

	Mathematics (*n* = 1,802)			
Boredom	T1	T2	T3	T4
BO-I	1.76 (.62)	1.81 (.65)	2.08 (.82)	2.48 (.74)
BO-U	2.05 (.67)	2.06 (.66)	2.12 (.70)	2.25 (.71)
BO-O	1.64 (.60)	1.75 (.64)	1.94 (.71)	2.27 (.87)
	German (*n* = 1,797)			
BO-I	1.82 (.69)	1.96 (.73)	2.24 (.83)	2.45 (.81)
BO-U	2.01 (.70)	2.05 (.66)	2.25 (.66)	2.38 (.66)
BO-O	1.58 (.60)	1.70 (.69)	1.75 (.63)	1.76 (.59)

Note. BO-I = Intensity of Boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge. Factor scores had the same metric as item indicators due to effects coding (Little et al., 2006), spanning a possible range of 1 – 5.

To test the observed developmental pattern, we estimated linear latent growth curve models, spanning T1-T4, including all three boredom factors. Linear latent growth curve models approximated the full sample data well in both subject domains (mathematics: *n* = 1,802 / $\chi^2(207) = 521.91$ / CFI = .963 / RMSEA = .029, 90% CI [.026, .032] / SRMR = .048; German: *n* = 1,797 / $\chi^2(207) = 314.26$ / CFI = .987 / RMSEA = .017, 90% CI [.013, .021] / SRMR = .032).

Table 3.3

Unstandardized Parameter Estimates for Latent Growth Factors in Linear LGC Models

	Intercept		Slope	
Boredom	Mean (*SE*)	Variance (*SE*)	Mean (*SE*)	Variance (*SE*)
Mathematics (*n* = 1,802)				
BO-I	2.497 (.057)***	.734 (.123)***	.075 (.007)***	.008 (.002)***
BO-U	2.244 (.052)***	.877 (.110)***	.020 (.007)**	.010 (.002)***
BO-O	2.290 (.042)***	.705 (.095)***	.063 (.005)***	.006 (.001)***
German (*n* = 1,797)				
BO-I	2.498 (.066)***	.679 (.109)***	.062 (.007)***	.007 (.001)***
BO-U	2.395 (.049)***	.414 (.091)***	.036 (.006)***	.006 (.001)***
BO-O	1.792 (.034)***	.414 (.066)***	.016 (.004)***	.003 (.001)***

Note. BO-I = Intensity of Boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge. ** $p < .01$. *** $p < .001$.

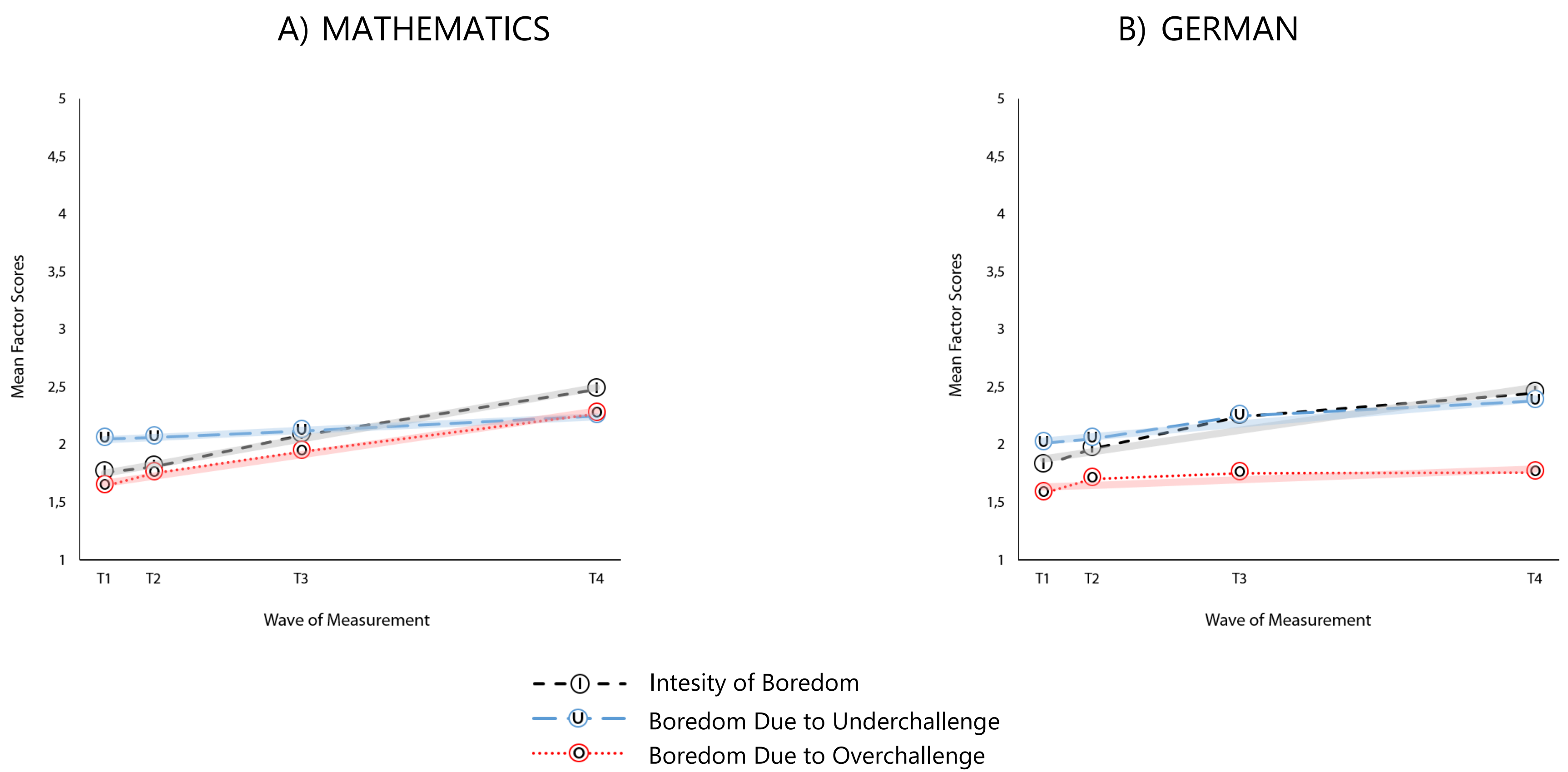
For the domain of mathematics, all estimated growth factor means and variances were significant at an α-level ≤ .01 (see Table 3.3). Mean growth rates of boredom were highest for intensity (*M* = .075, *SE* = .007), followed by boredom due to overchallenge (*M* =.063, *SE* = .005), and boredom due to underchallenge (*M* = .020, *SE* = .007), indicating increases of less than 10% of an original unit per wave of measurement for all three types of boredom. Individual growth rates for all types of boredom, albeit statistically significant, showed little variation (all variances ≤ .01, all *SE* < .002).

For German, mean growth rates of boredom were largest for intensity (*M* = .062, *SE*

= .007), followed by boredom due to underchallenge (*M* = .036, *SE* = .006) and boredom due to overchallenge (*M* = .016, *SE* = .004). Compared to mathematics, mean growth rates were smaller for intensity of boredom as well as boredom due to overchallenge. Boredom due to underchallenge, however, showed a higher mean growth rate compared to the mathematics model. Similar to mathematics, there was little individual variability in the growth rates for all three types of boredom in German (all variances < .01, all *SE* = .001). Estimated trajectories for all three latent factors are depicted in Figure 3.3 alongside mean factor scores from models of scalar measurement invariance over time.

Figure 3.3

Mean Factor Scores for Three Types of Boredom in Mathematics (A, n = 1,802) and German (B, n = 1,797)



Note. Factor scores had the same metric as item indicators due to effects coding (Little et al., 2006), spanning a possible range of 1 – 5. Dashed lines indicate implied trajectories of mean factor scores taken from models of scalar measurement invariance over time. Broad transparent lines indicate respective estimated linear latent growth curves.

Predicting Latent Growth by Class Type

Intervention Effectiveness (Full Sample)

Adding class type as predictor to the latent growth curves resulted in a good approximation of the data (mathematics: n = 1,802 / $\chi^2(225)$ = 552.75 / CFI = .962 / RMSEA = .028, 90% CI [.025, .031], SRMR = .047; German: n = 1,797 / $\chi^2(225)$ = 328.43 / CFI = .987 / RMSEA = .016, 90% CI [.012, .020], SRMR = .031; see Table 3.4 for parameter estimates).

Table 3.4

Standardized Effects of Class Type on Subject-Specific Boredom Trajectories

Predictive Paths	Mathematics		German	
	$\beta_{\text{Effectiveness}}$ (*SE*) (*n* = 1,802)	β_{Efficacy} (*SE*) (*n* = 528)	$\beta_{\text{Effectiveness}}$ (*SE*) (*n* = 1,797)	β_{Efficacy} (*SE*) (*n* = 525)
BO-I				
Class Type→I	-.16* (.07)	-.21* (.10)	-.05 (.07)	.07 (.09)
Class Type→S	-.12 (.06)	-.18 (.10)	-.03 (.07)	.03 (.10)
BO-U				
Class Type→I	.02 (.06)	-.08 (.08)	-.04 (.08)	-.08 (.15)
Class Type→S	.04 (.05)	-.03 (.08)	-.02 (.07)	-.15 (.16)
BO-O				
Class Type→I	-.18*** (.04)	-.03 (.07)	-.07 (.06)	.13 (.09)
Class Type→S	-.11 (.06)	-.06 (.19)	.06 (.08)	.14 (.11)

Note. BO-I = Intensity of Boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge. I = latent intercept factor. S = latent slope factor. Class Type was coded as 0 = regular classes, 1 = special classes for the gifted, with positive β's indicating higher values for gifted compared to regular classes and vice versa.

* $p < .05$. *** $p < .001$.

For mathematics, class type significantly predicted the mean level of intensity of boredom at T4 (β = -.16) as well as the mean level of boredom due to overchallenge at T4 (β = -.18). According to Keith (2006), both of these path coefficients represent moderate effects. Predictions of mean growth rates for both intensity of boredom (β = -.12) and boredom due to overchallenge (β = -.11) by class type also showed moderate effects (Keith, 2006) but were not statistically significant. For boredom due to underchallenge, neither mean levels (β = .02) nor mean growth rates (β = .04) could be significantly predicted by class type. All correlations between latent intercept and slope factors were significant, yielding large effect sizes (*r*'s ≥ .70 , Cohen, 1992).

For German, class type did not significantly predict any of the latent growth factors. Absolute β's ranged from .02 to .07, representing small effects (Keith, 2006). Intercept-slope correlations showed large effects (*r*'s ≥ .55), with the highest associations between mean levels and mean growth rates for intensity of boredom (*r* = .71).

Intervention Efficacy (Matched Sample)

Both LGC models showed good fit to the data (mathematics: n = 528 / $\chi^2(225)$ = 348.028 / CFI = .956 / RMSEA = .032, 90% CI [.025, .039], SRMR = .060; German: n = 525 / $\chi^2(225)$ = 325.478 / CFI = .967 / RMSEA = .029, 90% CI [.022, .036], SRMR = .050; see Table 3.4 for parameter estimates).

For mathematics, class type significantly predicted mean intensity of boredom at T4 (β = -.21) but did not significantly predict intensity's mean growth rate (β = -.18). Both of these effects were of moderate size (Keith, 2006). For boredom due to underchallenge as well as boredom due to overchallenge, coefficients yielded small to moderate nonsignificant effect sizes (range of absolute β's = .03 – .08) for mean levels as well as mean growth rates of all types of boredom. Intercepts and slopes were positively correlated (*r*'s ≥ .78, *p*'s < .001).

For German, class type did not significantly predict any of the latent growth factors. Moderate predictive effects were found, however, for mean boredom due to overchallenge at T4 (β = .13), as well as mean growth rates for boredom due to underchallenge (β = -.15) and boredom due to overchallenge ($\beta = .14$). Mean levels of both intensity of boredom at T4 ($\beta = .07$) and boredom due to overchallenge at T4 ($\beta = .13$) were higher in special classes for the gifted, with small to moderate effect sizes (Keith, 2006). The same applied to mean growth rates (range of β's = .03 – .15).

Discussion

We investigated the development of academic boredom in the domains of mathematics and German over a three-and-a-half-year period with special attention to the effects of full-time ability grouping in regular high-track classes vs. special classes for the gifted. We assessed three types of boredom in a large student sample in four waves of measurement. We expected an increase in boredom over time (H1) and investigated ability grouping's intervention effectiveness (H2) and efficacy (H3) regarding boredom development. Overall, the assumed increase in boredom over time was evident in our data, whereas support for ability grouping's effectiveness and efficacy to reduce boredom was limited.

In support of H1, we found an increase in all types of boredom in both mathematics and German over time. This finding is in line with previous work showing increasing general levels of boredom in academia (e.g., Hornstra et al., 2017; Pekrun et al., 2014; Vierhaus et al., 2016; Weybright et al., 2020). However, the present study went a step further, empirically separating three different types of boredom (i.e., intensity of boredom, boredom due to underchallenge, and boredom due to overchallenge).

For both subjects, intensity of boredom showed the highest mean levels at the last wave of measurement and the highest mean growth rates compared to the other two types of boredom. While intensity of boredom showed a very similar trajectory for both subject domains, boredom due to under- and overchallenge showed different trajectories for mathematics and German. In mathematics, boredom due to underchallenge stayed rather stable whereas boredom due to overchallenge increased at a similar pace as intensity. In German, boredom due to overchallenge grew rather slowly. Boredom due to underchallenge in German did not show a slope as steep as that of intensity but mean levels were higher compared to those of intensity for T1 – T3. These findings indicate that perceived underchallenge and overchallenge develop differently, depending on the subject. Mathematics appear to be increasingly overchallenging over the course of secondary school, possibly due to enhanced perceived or actual difficulty of the subject matter. German, on the other hand, provides a rather low and constant level of overchallenge. Underchallenge in German starts out at a slightly higher level compared to mathematics and it increases more rapidly.

An important side note to our findings is the fact that analyses on boredom trajectories were based on a combined sample of students from special and regular classes. In our supplementary subsample analyses for special classes for the gifted (see Table A3.7), we found lower levels of boredom due to underchallenge in both subjects, as well as lower levels of intensity of boredom in mathematics at T2, compared to T1. This trend is indicative of ability grouping's early effectiveness in avoiding boredom due to underchallenge during the first semester of secondary school, and hence in support of the boredom hypothesis. However, this observed early decrease in boredom levels quickly regressed towards the generally increasing trajectory that we also found in regular classes from T2 onwards.

Regarding H2 and H3, we found small to moderate effects of gifted ability grouping on academic boredom development regarding both efficacy and effectiveness, most of which did not reach statistical significance. Therefore, evidence in favor of the boredom hypothesis is limited at best and points to only small effects, which is in line with previous work (Hornstra et al., 2017; Preckel et al., 2010).

In the domain of mathematics, both in the full and in the matched samples, special classes for the gifted showed significantly lower levels of boredom intensity compared to regular ones, indicating the desired effect of ability grouping with regard to boredom intensity. Additionally, in the analysis of intervention effectiveness in the full sample, we found higher mean levels of boredom due to overchallenge in regular classes. Boredom due to overchallenge showed a smaller growth rate for special classes compared to regular classes (small effect). In the domain of German, none of the effects of ability grouping were significant and most effect sizes were small, indicating no meaningful impact of full-time ability grouping in special classes for the gifted on any of the three types of boredom.

Strengths, Limitations, and Future Research

First, our sample consisted exclusively of students attending grades five through eight (i.e., the first half of secondary education) in the highest of three secondary school tracks of the German educational system (“Gymnasium”). Secondary education tracking as it applies in Germany, can be regarded as a superordinate form of full-time ability grouping in and of itself (Becker et al., 2014). However, we were interested in the effects of full-time ability grouping into special classes for the gifted. Intellectually gifted students predominantly attend “Gymnasium”, which is also the track in which gifted ability grouping is usually carried out in practice. Thus, comparing special classes with regular classes in the highest track reveals a rather conservative test of the boredom hypothesis. Furthermore, the five schools providing data were located in only two out of the sixteen German federal states and all of them featured gifted programming in the form of ability-grouped classes. This is worth mentioning because the German educational system is structured and regulated on a federal state level. However, we achieved a large total sample (N = 1,861), allowing for complex longitudinal analyses using confirmatory latent models.

Second, our boredom measurement methodology showed certain limitations. We used only two items per factor within our SEM approach. This is due to the longitudinal design of the project the data is from, which investigated multiple constructs so that space for assessing a single construct was limited. In addition, all boredom measures were self-reports, and therefore subject to faking and social desirability. Unlike most boredom studies however, we separated three types of boredom (instead of general boredom only) and two subject domains (instead of only focusing on the prominent mathematics domain), obtaining differential and subject-specific practical implications, which past studies did not.

Third, our PSM procedure returned small differences (d's > .20 / w's > .10) between regular and special classes for three matching variables (for details see Table 3.1). For the three boredom-related scales, mean differences increased post-matching, however still yielding d's < .20. Considering the high number of matching variables (32), our PSM approach was able to produce two comparable samples of matched high ability students, enabling fairly precise separation of intervention effectiveness and efficacy while controlling for sample-related differences in demographics, cognitive ability, motivational and social parameters.

Beyond the present investigation, future studies should include other school tracks and preferably a longer investigation period, for example, spanning all of secondary education (i.e., grades 5 through 12). The inclusion of schools with and without ability grouping plans across different federal states or countries could help widen the scope of future results as well. Academic boredom should also be explored in less researched subject domains (e.g., physics, chemistry, sports, second and third language classes). Moreover, future studies should feature systematic cross-domain comparisons in order to further clarify the subject-specific effects of ability grouping. Measurement models of academic boredom should continue to include different types of subject-specific boredom simultaneously, while using a greater number of item indicators for enhanced scale reliability and the possibility of testing for cross-sectional scale invariance. As calls for more objective measures of boredom become more and more frequent (e.g., Goetz et al., 2019), physiological measures should also be included to measure boredom and compare results with classic self-report scales. Finally, future ability grouping studies should continue to apply PSM procedures in order to gain empirically relevant control groups. While academic boredom should stay relevant as a primary outcome in those studies, it would be important to include additional matching variables, such as instruction quality (Goetz, 2004), learning preferences (Obergriesser & Stoeger, 2016), or academic performance.

Practical Implications

Taking given limitations into consideration, the present study yields several practical implications for educational systems struggling with boredom among students. Increasing boredom over the course of secondary education is an alarming tendency, requiring ongoing, direct boredom prevention over the course of the school career, especially when transitioning from primary to secondary school. Clarity and structure of classroom instruction (Goetz, 2004), as well as an understandable and supportive presentation style at the hand of the teacher (Goetz et al., 2013b), for example, can help prevent boredom. Furthermore, tasks should be designed in a way that features an individually optimized level of stimulation (Westgate & Wilson, 2018), which requires profound knowledge of students' abilities and learning preferences. From CVT's (Pekrun, 2006) perspective, the real-life value of mathematics and German as subjects should be systematically communicated in order to increase students' subjective value. This may be accomplished, for instance, by incorporating more life-like task framing at school, such as by determining the shape of a spreading virus curve (in math class) or by writing a CV (in German class). Boredom due to overchallenge needs to be specifically tackled in math class (e.g., by revisiting curricula and common instruction forms for math class in "Gymnasium").

Our evaluation of full-time ability grouping of the gifted as a classroom-based intervention can be described as effective in reducing math-specific boredom intensity and math- specific boredom due to overchallenge, while it cannot buffer increasing German-specific boredom trajectories effectively. There are alternative explanations for these subject-specific findings. The selection processes for gifted class programs in our sample included intelligence tests, which correlate highly with math abilities (Roth et al., 2015). A selection approach favoring verbal over math abilities, for example, might have produced a different result pattern regarding effects of ability grouping on subject-specific boredom development. Moreover, mathematics might have a higher value than German in the special classes for both students and teachers, because of higher perceived predictive value for cognitive ability and, hence, intellectual giftedness (Hannover & Kessels, 2004). Therefore, it is plausible to assume that both teachers and students in the special classes invested more in mathematics as compared to German, which might have reduced feelings of boredom. In addition, dimensional comparisons between both domains (Möller & Marsh, 2013; Wigfield et al., 2020) that are based on higher math abilities and a higher value of mathematics in the special classes might lead to a more unfavorable development in German. Of course, these explanations are rather speculative and require validation by further research.

Theoretical Implications

From a theoretical standpoint, our study indicates that boredom as a psychological construct, needs to be researched with differentiation and specificity, meaning that subject-specific framing and simultaneous consideration of multiple types of boredom should be preferred over general frameworks. Moreover, theoretical frameworks of academic boredom could be refined by including more specific predictions for the role of challenge. The first-order factor model we used in our study, systematically captures the interplay of boredom intensity, boredom due to underchallenge, and boredom due to overchallenge, providing an empirically sound, economically implementable base that could be expanded on in future theory building.

Presently, there is a plurality of boredom measurement approaches with different underlying theoretical and factorial frameworks (Vodanovich & Watt, 2016), even when limiting the search to purely academic conceptualizations of boredom (e.g., Acee et al., 2010; Daschmann et al., 2011; Pekrun et al., 2011). While we do realize that our first-order factor model and its pertaining measurement approach add to this already vast field, it might be useful for reducing heterogeneity in existing boredom conceptualizations at the same time. By integrating implications from flow theory (Csikszentmihalyi, 1975, 1990), findings from empirical studies of different boredom types (Acee et al., 2010; Daschmann et al., 2011; Westgate & Wilson, 2018), and CVT (Pekrun, 2006), our three-type boredom framework might contribute to the development of a unified, comprehensive model of academic boredom.

Conclusion

In sum, we found no general support for the boredom hypothesis (Feldhusen & Kroll, 1991) or for the capacity of ability-grouped classrooms to reduce boredom (e.g., Bar-On, 2007; Little, 2012; Plucker et al., 2004; Rogers, 2007). However, ability grouping has been shown to be effective in several other ways, for example by increasing academic performance (Steenbergen- Hu et al., 2016), or providing social facilitation for gifted students (Rogers, 2007). Academic self-concept, on the other hand, has been shown to suffer from grouping processes (Fang et al., 2018), although some newer studies have not found evidence for impaired academic self- concepts in high ability grouping settings (Herrmann et al., 2016; Preckel & Brüll, 2010; Preckel et al., 2019). Ultimately, convincing pro-grouping arguments other than boredom reduction remain intact. Still, careful consideration of several psychologically relevant developmental processes (e.g., cognitive, affective-motivational, social, health-related) is advised when contemplating an installment of ability-grouped classrooms in a given school environment. An increase in PSM studies is needed, however, to properly foster the next best thing to systematic effectiveness or efficacy studies, which, for obvious ethical reasons, cannot be applied with appropriate methodological precision (i.e., randomization) in an active educational system.

Chapter 4 – Study 3

Gender Differences in Academic Boredom and Its Development in Secondary School

Markus D. Feuchter and Franzis Preckel

Department of Psychology, University of Trier

Abstract

This longitudinal study examines gender differences in subject-specific academic boredom and its development (i.e., in mathematics and German) in an adolescent sample (N = 1,428; 53.3% male). Students attended the regular German high-track secondary education (Gymnasium) and were assessed four times in Grades 5-8. We estimated linear latent growth curves for three forms of academic boredom, namely *boredom intensity* (BO-I), *boredom due to underchallenge* (BO-U), and *boredom due to overchallenge* (BO-O) over four waves of measurement (T1-T4). We predicted latent intercept and slope factors by gender. Subject-specific appraisals (i.e., academic self-concept; academic interest) and performance (i.e., previous grade in the target subject at the end of primary education) were included as mediators. Results indicated enhanced boredom levels for male students in both subjects at the end of our investigation period ($\beta_{\text{BO-U in mathematics}} = -.15^{**}$, $\beta_{\text{BO-I in German}} = -.11^{*}$, $\beta_{\text{BO-O in German}} = -.22^{***}$), as well as an enhanced boredom growth rate for boredom due to overchallenge in German ($\beta = -.19^{**}$). Implications for educational theorizing and practice are discussed.

Keywords: gender differences, academic boredom, latent growth curve modeling, mathematics, language

Gender Differences in Academic Boredom and Its Development in Secondary School

Boredom is “an affective indicator of unsuccessful attentional engagement in valued goal-congruent activities” (Westgate, 2020, p. 34; see also Westgate & Wilson, 2018). Boredom in school (i.e., academic boredom) constitutes a major challenge in education due to its high prevalence and detrimental relations with negative outcomes in- and outside the classroom. It is associated with lower academic achievement (Tze et al., 2016), lower physical and psychological well-being (Schwartze et al., 2021), depressive affect and anxiety (Freund et al., 2021; LePera, 2011), as well as unhealthy behaviors, like substance abuse or problem gambling (for an overview, see Bench & Lench, 2013). Boredom increases in adolescence and over the course of secondary education (Meyer & Schlesier, 2021; Spaeth et al., 2015; Weybright et al., 2020), further exacerbating the academic boredom problematic.

While boredom at school affects all students, there are gender differences in girls vs. boys’ boredom levels, especially when considering subject matter (e.g., Pekrun et al., 2017; Zaccoletti et al., 2020). To understand these level differences and their emergence over time, longitudinal findings are needed. Especially adolescence seems to be a time with significant increases in students’ boredom (e.g., Vierhaus et al., 2016). However, not much evidence exists on potential gender differences in boredom trajectories over larger time periods (Spaeth et al., 2015; Weybright et al., 2020). The core purpose of the present study is to investigate gender differences in subject-specific academic boredom longitudinally in adolescents over a three-and-a-half-year period. Thereby, we care to learn whether girls and boys are at differential risk for boredom, and hence, its negative consequences. To our knowledge, no studies have yet tested for gender differences in subject-specific academic boredom development in secondary school.

Introduction

Academic Boredom

The academic realm is characterized by achievement situations. Boredom in these settings can be conceptually defined as one of 19 achievement emotions (Pekrun, 2006, 2017) that is experienced on-task (*activity emotion*; e.g., joy) as opposed to prospectively or retrospectively (*outcome emotions*; e.g., pride or shame). Boredom is “the most prevalent emotion experienced in a classroom” (Goetz et al., 2020, p. 4), with prevalence rates of more than 50% (during mathematics class; Nett et al., 2011). Experiences of boredom at school are domain-specific (Goetz et al., 2006, 2010) and can be the result of either under- or overchallenge (cf. Raffaelli et al., 2018). Only a few modelling approaches distinguish between boredom due to under- and overchallenge, however (cf. Acee et al., 2010; Daschmann et al., 2011). One of these, is the 3F-model of boredom (Feuchter & Preckel, 2022b), which comprises three separate but correlated factors of boredom, namely *boredom intensity* (BO-I), *boredom due to underchallenge* (BO-U) and *boredom due to overchallenge* (BO-O). This model is also featured in this work. Regarding learning and achievement, meta-analyses indicate substantial negative correlations between academic boredom and performance (σ = -.25, Camacho-Morles et al., 2021; $\bar{r}$ = -.16, Tze et al., 2016), as well as motivation ($\bar{r}$ = -.40, Tze et al., 2016). Furthermore, boredom and academic achievement have been shown to have negative reciprocal effects on each other at every stage of the educational system (Lichtenfeld et al., 2022; Pekrun et al., 2014, 2017).

Development of Academic Boredom in Secondary School

A multitude of studies point to increasing boredom over the course of secondary school. For example, Spaeth et al. (2015) found modest increases in leisure boredom for 10- to 14-year-olds in Germany. More recently, Weybright et al. (2020) showed a rise in boredom from 2008-2017 in the US, using representative samples of 8th, 10th, and 12th Grade students.

Boredom increases were also apparent within Grade levels. Regarding academic boredom, a longitudinal study of two German samples (Grade 2-5 and 4-7) both showed increases in general boredom, with substantial changes between Grade 5 and 7 (Vierhaus et al., 2016, p. 12). Focusing on the transition from primary to secondary school, Meyer and Schlesier (2021) found similar increases in general boredom over time during the first two years of secondary school in Germany. Preckel and colleagues (2010) found different trajectories of three different forms of mathematics-specific boredom in an Austrian student sample during the first half of Grade 9, with initial increases in the frequency of boredom and in boredom due to overchallenge that later stabilized. Boredom due to underchallenge, on the other hand, remained stable after an initial decrease. Longitudinal studies separating different forms of academic boredom are scarce, however (cf. Feuchter & Preckel, 2022b). Likewise, subject domains are seldomly compared, with mathematics as the predominant research domain.

Gender Differences in Academic Boredom

Academic boredom is one of the most prevalent emotions experienced in modern-day classrooms, regardless of gender (Goetz et al., 2020). Nonetheless, there are gender differences in school-related boredom levels. For mathematics, Pekrun et al. (2017) found lower levels of boredom for girls vs. boys. Regarding challenge-related forms of mathematics boredom, Goetz and Frenzel (2010) found boredom due to underchallenge to be more pronounced in boys, whereas boredom due to overchallenge was more pronounced in girls. Comparable results were obtained by Daschmann et al. (2011). For verbal domains and language instruction, Zaccoletti et al. (2020) found gender differences indicating higher boredom for boys vs. girls when engaged in reading. Similarly, Raccanello et al. (2019) found increased native language boredom in elementary school boys vs. girls. Challenge-related boredom research is lacking for verbal domains. Considering gender differences in verbal achievement, research consistently points towards higher verbal achievement in girls vs. boys (Logan & Johnston, 2009; Mostafa & Schwabe, 2019; Petersen, 2018; Reilly et al., 2019; Voyer & Voyer, 2014). This could imply differential challenge perception, with girls potentially experiencing more boredom due to underchallenge and less boredom due to overchallenge in language classes because of their higher verbal achievement. However, more research is needed here.

Even less is known about gender differences in academic boredom trajectories. Regarding nonacademic boredom, Spaeth et al. (2015) found no gender differences in boredom levels or growth rates in their sample. Weybright et al. (2020), on the other hand, found girls to experience steeper growth rates than boys from 2008 to 2017. However, this result represents historic rather than longitudinal effects. The few existing longitudinal studies of academic boredom have mostly included gender as a control variable, putting less emphasis on its potential impact on boredom slopes or boredom levels at later stages (e.g., Forsblom et al., 2021; Meyer & Schlesier, 2021; Pekrun et al., 2017). No studies we know of systematically examined gender differences in domain-specific academic boredom trajectories. We aim to close this research gap with the present work, acknowledging different forms of academic boredom in the process.

Explaining Gender Differences in Academic Boredom

It is important to note that statistical gender differences do not necessarily reflect actual differences that can be traced back to gender identities. Observed gender differences are likely to emerge due to otherwise relevant variables that are correlated with gender instead. As for gender differences in academic emotions, control-value theory (CVT; Pekrun, 2006, 2019) posits that appraisal antecedents to emotional experience, namely subjective control and value appraisals, potentially mediate distal effects of individual or contextual antecedents such as gender. In addition, subjective challenge perceptions that are related to academic performance indicators, such as previously received school grades, may also mediate gender differences in academic boredom – especially considering boredom due to under- and overchallenge as separate, challenge-related forms (cf. Feuchter & Preckel, 2022b).

For both, appraisal antecedents and performance indicators, gender differences are much more clearly researched in academic contexts, compared to academic emotions (e.g., Keller et al., 2021; Voyer & Voyer, 2014). Results mostly point in the same direction: appraisals, like subjective control and value, as well as performance, are enhanced for boys (vs. girls) in the mathematics domain, whereas the opposite pattern of enhanced appraisals and performance can be observed for girls (vs. boys) in the verbal domain (e.g., Frenzel et al., 2007; Goetz et al., 2008, 2013a; Goetz & Frenzel, 2010; Keller et al., 2021). Mathematics performance has a special role to play, nevertheless. As reported by international student assessment studies (cf. Mostafa & Schwabe, 2019; OECD, 2012), gender gaps in mathematics achievement have been closing in recent years. Moreover, girls often receive better mathematics grades than boys at school (Pennington et al., 2021; Voyer & Voyer, 2014). In conclusion, if students' appraisals and performance are controlled by including them as mediators (cf. CVT), remaining gender differences in academic boredom cannot be traced back to these plausible and well-documented alternative explanations.

The Present Study

We investigated boredom in mathematics and the verbal domain of German and used the 3F-model of boredom, which distinguishes boredom intensity (BO-I), boredom due to underchallenge (BO-U), and boredom due to overchallenge (BO-O) (Feuchter & Preckel, 2022b). Gender differences throughout Grades 5-8 of secondary education were examined longitudinally, regarding boredom *levels* at Grade 8, as well as boredom *trajectories* between Grade 5 and Grade 8. In previous work with partly overlapping data but a larger composite sample, we found linearly increasing trajectories for these three forms of academic boredom in mathematics and German, using latent growth curves (cf. Feuchter & Preckel, 2022b).

These previous analyses are repeated for the present study's sample and amended by a gender difference perspective. Additionally, appraisal antecedents and previous academic performance were included as potential mediators to statistical effects of gender. The following sections contain our research questions (RQ) and pertaining hypotheses (H).

Gender Differences in Academic Boredom Levels

Drawing from previous results from studies examining subject-specific gender differences in academic boredom (Daschmann et al., 2011; Goetz & Frenzel, 2010; Pekrun et al., 2017; Raccanello et al., 2019; Zaccoletti et al., 2020), we expect boys to show higher levels of boredom intensity in mathematics and German, as well as higher levels boredom due to underchallenge in mathematics. We expect girls to show higher levels of boredom due to overchallenge in mathematics.

Drawing from evidence on girls' higher verbal performance (Logan & Johnston, 2009; Mostafa & Schwabe, 2019; Petersen, 2018; Reilly et al., 2019; Voyer & Voyer, 2014), we expect girls to have higher levels of boredom due to underchallenge in German. We expect boys to show higher levels of boredom due to overchallenge in German.

RQ 1: Are there gender differences in *levels* of subject-specific academic boredom at Grade 8?

H1a) Boys show higher mean levels of boredom intensity in mathematics and German, boredom due to underchallenge in mathematics, and boredom due to overchallenge in German than girls.

H1b) Girls show higher mean levels of boredom due to overchallenge in mathematics and higher mean levels of boredom due to underchallenge in German than boys.

Gender Differences in Academic Boredom Trajectories

In line with previous works on boredom development in secondary school (Feuchter & Preckel, 2022b; Meyer & Schlesier, 2021; Spaeth et al., 2015; Vierhaus et al., 2016;

Weybright et al., 2020), we expect linearly increasing boredom trajectories over time. As no theory of boredom formulates assumptions on gender differences regarding boredom development and respective findings are missing, we investigate gender differences in boredom *trajectories* over time as open research question.

H2: Boredom intensity (BO-I), boredom due to underchallenge (BO-U), and boredom due to overchallenge increase over time in mathematics and the verbal domain of German.

RQ 2: Are there gender differences in developmental *trajectories* of subject-specific academic boredom from Grades 5-8?

Material and Methods

Participants and Procedure

Data of this study were collected as part of a longitudinal project (AVG-project), examining the motivational, social, and affective development of German students over the course of secondary school in the top track ("Gymnasium") of the German three-track secondary school system. The AVG-project was approved by the Supervision and Services Directorate of Rhineland-Palatinate (Aufsichts- und Dienstleistungsdirektion) (protocol number:32-03 405/29/05). Target subject domains were mathematics and German. Both are highly important subjects taught throughout primary and secondary school in Germany. Our study sample comprised N = 1,428 secondary school students (53.3% identified as male, n = 761; 46.7% identified as female, n = 667), attending Grades 5-8 at five German high-track schools located in the federal states of Rhineland-Palatinate (4 schools) and Bavaria (1 school). We investigated five successive cohorts from the school term of 2005/2006 onwards. Students completed questionnaires over four waves of measurement (T1: early 5th Grade, T2-T4: after mid-term evaluations in 5th Grade, 6th Grade, and 8th Grade)[6]. Previous teacher-assigned

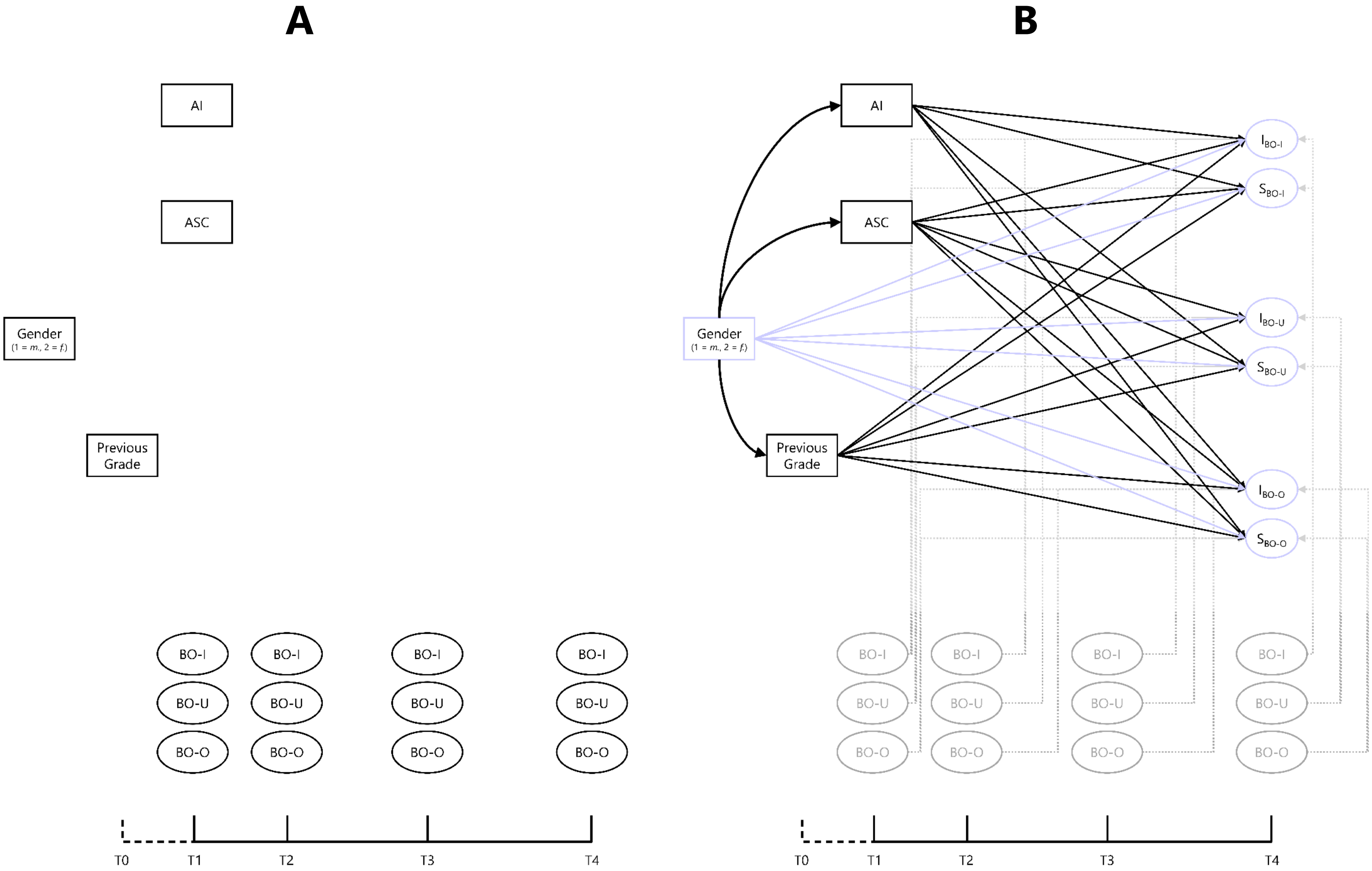
[6] Precise waves of measurement differed slightly for each school under investigation. On average, students attended 5th Grade for 46.01 days (*SD* = 17.41) at T1. On average, periods in between successive waves of measurement spanned 122.95 days (*SD* = 32.56) for T1 and T2, 369.82 days (*SD* = 31.89) for T2 and T3, and 721.06 days (*SD* = 27.23) for T3 and T4.

school grades coming out of primary school were acquired beforehand (T0: at the beginning of 5th Grade). See Figure 4.1A for a visualization of our study design.

Respective mean ages (in years) for students at each wave of measurement were M_{T1} = 10.24 (SD = 0.56), M_{T2} = 10.54 (SD = 0.57), M_{T3} = 11.56 (SD = 0.57), and M_{T4} = 13.55 (SD = 0.57). We assessed socio-economic background by student reports of highest educational degrees for both parents (for mothers: 12.4% university graduates, 10.3% high school graduates, 6.2% high-track secondary school graduates, 2.5% PhDs, 2.4% primary or middle school graduates, 0.1% without degree, 66% without entry; for fathers: 12.3% university graduates, 7% high school graduates, 5.8% high-track secondary school graduates, 5% PhDs, 3.3% primary or middle school graduates, 0.1% without degree, 66.5% without entry). The study was conducted entirely in German. Most students stated German as their main language (66.1%; 23.5% missing). Only 10.4% had a different first language and reported speaking German for 7.63 years on average (SD = 2.31). Hence, we expected no language-related comprehension problems during our investigation.

Figure 4.1

Visualization of Study Design (A) and Full MIMIC Model (B)



Note. Gender was codes as 1 – *male*, 2 – *female*. AI = academic interest. ASC = academic-self-concept. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. I = latent intercept factor. S = latent slope factor. T0 – T4 = Waves of measurement.

Latent slope factor loadings were: $\lambda_{slope,\ T1} = -10$, $\lambda_{slope,\ T2} = -8.377$, $\lambda_{slope,\ T3} = -4.979$, and $\lambda_{slope,\ T4} = 0$ for T1-T4 respectively, with identical latent intercept factor loadings ($\lambda_{intercept} = 1$). Correlations of study variables were allowed at each individual wave of measurement, as well as for all latent intercept and slope factors (omitted here). Colored variables and paths in B form the intermediate Step 2-MIMIC model which was extended in Step 3, by including mediation paths. See method section for further details.

Variables and Measures

Gender

Focusing on gender differences between adolescent girls and boys, we assessed students' gender as a binary variable (1 – *male*, 2 – *female*) in demographic questionnaires (*n* = 9 students had been excluded beforehand due to missing entries on gender).

Boredom Intensity

Assessed at T1-T4, we used a two-item scale drawn from the Achievement Emotions Questionnaire – Mathematics (AEQ-M; Pekrun et al., 2005) to measure subject-specific boredom intensity. Item responses were given on a 5-point rating scale (1 – *strongly disagree* to 5 – *strongly agree*). For item wording, see Table A4.1 of the Appendix. Using brief measures can be necessary in long-term assessment projects due to limited questionnaire space. In Gogol et al. (2014), for instance, brief scales produced an adequate correlational representation of motivational affective constructs, comparable to long scales. We calculated Spearman-Brown ρs as a reliability measure for two-item scales (Eisinga et al., 2013), for the mathematics (.70 / .73 / .79 / .76 for T1-T4) and German scales (.77 / .76 / .82 /. 82 for T1-T4).

Challenge-related Forms of Boredom

Assessed at T1-T4, we used two-item scales adapted from the PALMA project (Pekrun et al., 2007) to measure subject-specific boredom due to underchallenge and boredom due to overchallenge. Item responses were given on a 5-point rating scale (1 – *strongly disagree* to 5 – *strongly agree*). For item wording, see Table A4.1 of the Appendix. Spearman-Brown ρs for the mathematics scale of boredom due to underchallenge were .70 / .69 / .72 / .74 for T1-T4; for the German scale, ρs were .74 / .72 / .70 / .75 for T1-T4. Spearman-Brown ρs for the mathematics scale of boredom due to overchallenge were .77 / .74 / .79 / .82 for T1-T4; for the German scale, ρs were .79 / .84 / .76 / .73 for T1-T4.

Mediator Variables

Appraisals. Assessed at T1, we used four items adapted from the PALMA project (Pekrun et al., 2007) to assess subject-specific academic interest (AI). Item responses were given on a 5-point rating scale (1 – *strongly disagree* to 5 – *strongly agree*). For item wording, see Table A4.1 of the Appendix. Subjective value as an appraisal antecedent to achievement emotions, is often operationalized using measures of intrinsic, extrinsic, or utility value (Pekrun, 2019; cf. Putwain et al., 2018). AI, in our case, is most representative of intrinsic value. Cronbach's α was .87 for the mathematics scale and .85 for the German scale.

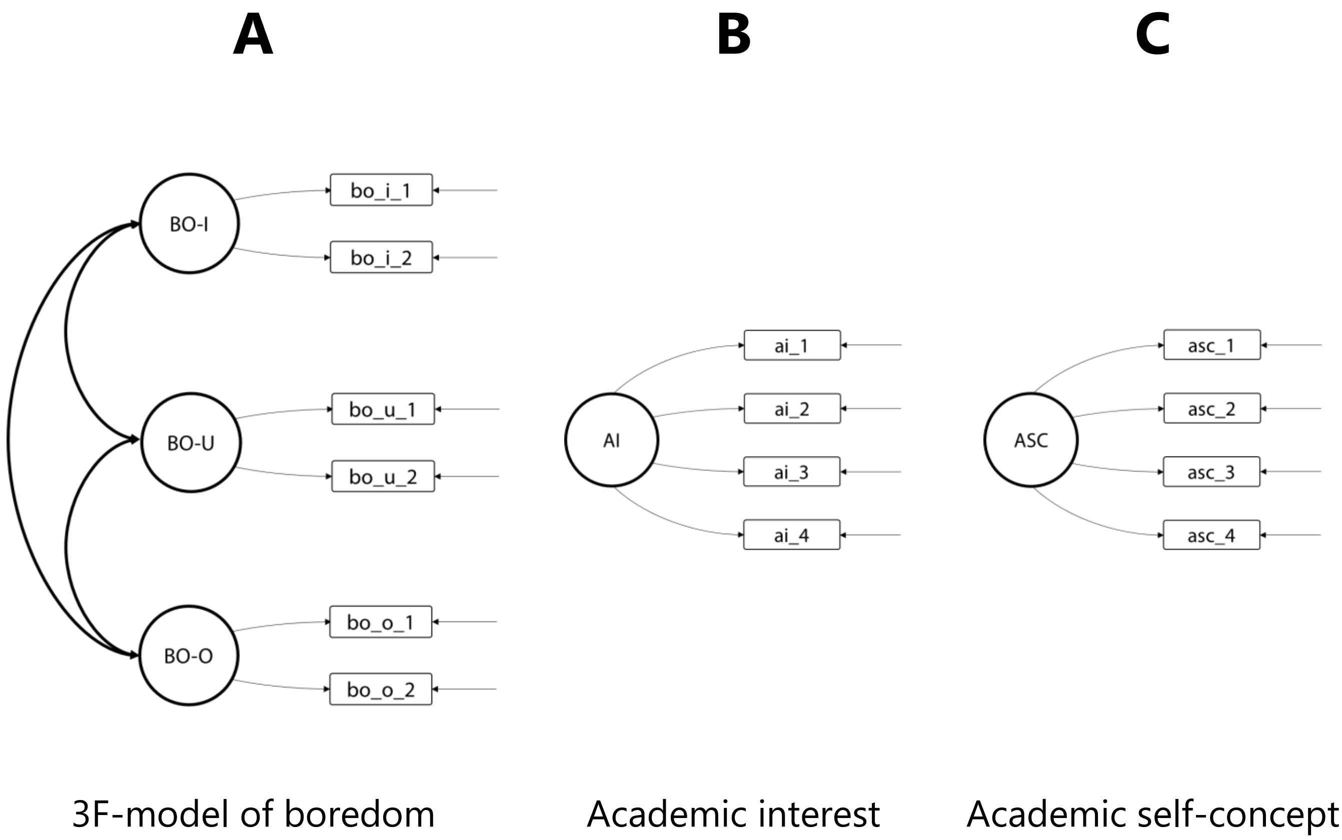
Also assessed at T1, we used four items from the short version of the Self-Description Questionnaire (SDQ-II; Marsh, 2007, 1990) to measure subject-specific academic self-concept (ASC). Item responses were given on a 5-point rating scale (1 – *strongly disagree* to 5 – *strongly agree*). For item wording, see Table A4.1 of the Appendix. Besides expectancies and attributions, ability self-concepts are often used indicators of subjective control as an appraisal antecedent to achievement emotions (Pekrun, 2019). Cronbach's α was .88 for the mathematics ASC scale and .88 for the German ASC scale.

Previous School Grades. Assessed at T0, we included previous teacher-assigned school grades received at the end of German primary school (i.e., Grade 4) as indicator of academic performance. Even though school grades are lacking in psychological measurement quality, as compared to standardized achievement tests, for example, they do possess high ecological validity (Pekrun et al., 2017). Therefore, school grades may represent individual challenge, as well as academic achievement and can predict subsequent achievement emotions via feedback processes (Pekrun, 2019). We asked students for their final grades in mathematics and German, coming out of primary school. In Germany, scholar grades take values from 1-6 (1 – *very good,* 2 – *good,* 3 – *satisfying,* 4 – *sufficient,* 5 – *deficient*, 6 –

insufficient), with higher values indicating poorer performance. We hence coded previous grade inversely for more intuitive interpretation of associated effect sizes in later analyses.

Figure 4.2

Measurement Models for Academic Boredom (A), Academic Interest (B), and Academic Self-concept (C)



Note. Models were estimated separately for mathematics and German domains and served as base for further model expansion. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept. AI = academic interest.

Data Analyses

Our main analyses concerned gender differences in academic boredom examined longitudinally by latent growth curve modeling methodology (LGCM; Bollen, 2007; Meredith & Tisak, 1990). Appraisals and previous grade were controlled for, by including mediation paths, connecting exogenous gender to boredom. That way, we could eliminate gender differences in these antecedents as alternative explanations.

All analyses were embedded in a structural equation modeling approach. For our main analyses, we estimated multiple-indicator multiple-cause models (MIMIC; Muthén, 1989; see also Müller & Schäfer, 2017; Thompson & Green, 2013), predicting latent growth factors in boredom (i.e., latent intercept and slope factors) by gender, appraisal antecedents, and previous grade (for a visualization, see Figure 4.1B). All analyses were based on the 3F-model of boredom (cf. Feuchter & Preckel, 2022b), with separate growth processes for three correlated boredom factors, representing BO-I, BO-U, and BO-O. Both domains of mathematics and German were analysed separately.

Analyses were conducted using Mplus version 8.4 (Muthén & Muthén, 1998-2019; for Mplus-specific methods literature see Byrne, 2012; Kleinke et al., 2017). Models were identified using effects coding (Little et al., 2006) to obtain comparable metrics for latent variables and manifest item indicators. We used the MLR estimator to account for distributional anomalies and the FIML procedure to cope with missing values.

Auxiliary tests of measurement models for latent variables (i.e., 3F-model of boredom, academic interest, and academic self-concept, see Figure 4.2 for a graphic depiction), via confirmatory factor analyses (CFAs), are reported in the Appendix (Tables A4.2 & A4.3).

Measurement Invariance Testing

As analytical prerequisite to using gender as an exogenous predictor of latent boredom, we tested the 3F-model of boredom (T1, T2, T3, T4), as well as ASC and AI models (T1) for (cross-sectional) measurement invariance across genders. We employed a step-up approach, up to a level of scalar invariance (Brown, 2015; Byrne, 2012; Kleinke et al., 2017). We therefore computed multi-group CFAs with invariance constraints across gender subsamples (French & Finch, 2008; see also Byrne, 2012; Kleinke et al., 2017), using Mplus's default "type is meanstructure" option. We followed Chen's (2007) criterion of ΔCFI < -.01 between adjacent models to evaluate measurement invariance across genders.

As analytical prerequisite to interpreting latent mean development over time, we also tested the 3F-model of boredom for (longitudinal) measurement invariance over time. We again employed a step-up approach, up to a level of scalar invariance (Brown, 2015; Kleinke et al., 2017). We conducted correlated uniqueness longitudinal CFAs with invariance constraints for repeated item indicators (Geiser, 2011; Kleinke et al., 2017), using Mplus's "type is complex" option to account for our nested data structure (i.e., students were nested in 43 different classrooms). To assess measurement invariance, we again turned to Chen's (2007) criterion of ΔCFI < -.01 between adjacent models.

Main Analyses

Given intended levels of measurement invariance for all relevant constructs, we would construe MIMIC models in a three-step process. In Step 1, we would estimate academic boredom development using linear latent growth curve (LGC) models based on the 3F-model of boredom. In Step 2, we would add gender as an exogenous predictor to latent growth factors of previous LGC models. In Step 3, we would include appraisal and performance mediator variables, arriving at our main analytical models. Model fits from all steps are evaluated using the GFI criteria proposed by Hu and Bentler (1999), with CFI ≈ .95, RMSEA ≈ .06, and SRMR ≈ .08, corresponding to a good fit.

Linear LGC Models of Academic Boredom (Step 1). Carrying over measurement invariant constraints over time, we estimated linear latent growth curve models based on the

3F-model of boredom. These linear trajectories were established in our previous work, using a larger sample that included this study's participants (Feuchter & Preckel, 2022b). Latent slope factor loadings represented coding of time, with mean periods in between waves of measurement as the criterion (Biesanz et al., 2004). The origin of time was placed at T4. Latent slope factor loadings for each form of boredom within the 3F-model were: $\lambda_{slope, T1} = -10$, $\lambda_{slope, T2} = -8.377$, $\lambda_{slope, T3} = -4.979$, $\lambda_{slope, T4} = 0$. Latent intercept factor loadings were identical at $\lambda_{intercept} = 1$ for each form of boredom and each wave of measurement.

Adding Gender as Exogenous Predictor (Step 2). To assess gender differences in boredom development, we regressed previously established latent growth factors on gender. Due to the coding of gender (1 – *male,* 2 – *female*), negative predictive path coefficients indicate higher values for boys, whereas positive coefficients reflect higher values for girls.

Adding Appraisal and Performance Mediator Variables (Step 3). In addition to modeling gender effects on latent growth in boredom, we also regressed latent growth factors of boredom on proximal appraisal antecedents (i.e., ASC and AI, assessed at T1) and previous performance (i.e., previous grade in mathematics or German, assessed at T0). Appraisal antecedents and performance were also regressed on gender, making them mediators to gender-boredom-relations. To reduce model complexity while also acknowledging measurement error, we would use ASC- and AI-factor scores stored from previous CFAs, given measurement invariance across genders and adequate factor determinacies (i.e., > .90, Mulaik, 2009; for details, see Table A4.3 & A4.5 of the Appendix).

Transparency and Openness

We report how we determined our sample size, all data exclusions, and all measures in this manuscript (see above). Materials and analysis code are available by emailing the corresponding author. Raw data cannot be made available to the broader public (except for reviewer requests) due to contractual obligations considering the [associate institution]. Data were analysed using Mplus, version 8.4 (Muthén & Muthén, 1998-2019). We used IBM SPSS Statistics, version 29 (IBM Corporation, 2022) for computing descriptive statistics and manifest correlations for all study variables. This study's design and its analyses were not pre-registered.

Results

In addition to results listed here, the Appendix reports the results on CFAs for latent variables (Tables A4.2 & A4.3). Descriptive statistics (Table 4.1), as well as manifest and latent correlations of all study variables (Tables 4.2 & 4.3) are displayed in Tables 4.1-4.3.

Table 4.1

Manifest Descriptive Statistics for Entire Sample and by Gender

	Entire sample									Males only									Fenales only								
	n	*M*	SE_{Mean}	*SD*	*Var*	Skewness	$SE_{Skewness}$	Kurtosis	$SE_{Kurtosis}$	*n*	*M*	SE_{Mean}	*SD*	*Var*	Skewness	$SE_{Skewness}$	Kurtosis	$SE_{Kurtosis}$	*n*	*M*	SE_{Mean}	*SD*	*Var*	Skewness	$SE_{Skewness}$	Kurtosis	$SE_{Kurtosis}$
Mathematics																											
Previous Grade$_{T0}$	1,070	5.14	.02	.62	.38	-.21	.08	.04	.15	552	5.20	.03	.60	.36	-.31	.10	.50	.21	518	5.07	.03	.63	.40	-.10	.11	-.31	.21
AI$_{T1}$	1,025	3.50	.03	1.06	1.13	-.35	.08	-.74	.15	526	3.67	.04	1.02	1.03	-.48	.11	-.53	.21	499	3.31	.05	1.08	1.17	-.20	.11	-.87	.22
ASC$_{T1}$	1,025	3.93	.03	.88	.77	-.64	.08	-.07	.15	526	4.14	.04	.83	.69	-.94	.11	.64	.21	499	3.72	.04	.88	.77	-.40	.11	-.30	.22
BO-I$_{T1}$	1,022	1.75	.03	.93	.86	1.40	.08	1.58	.15	526	1.74	.04	.91	.83	1.46	.11	2.03	.21	496	1.77	.04	.94	.88	1.33	.11	1.18	.22
BO-I$_{T2}$	1,028	1.83	.03	.98	.95	1.32	.08	1.29	.15	522	1.89	.05	1.04	1.08	1.24	.11	.88	.21	506	1.76	.04	.90	.82	1.38	.11	1.73	.22
BO-I$_{T3}$	998	2.08	.04	1.13	1.27	.98	.08	.10	.16	510	2.14	.05	1.19	1.43	.92	.11	-.13	.22	488	2.03	.05	1.05	1.10	1.01	.11	.31	.22
BO-I$_{T4}$	986	2.55	.04	1.13	1.27	.48	.08	-.54	.16	484	2.50	.05	1.15	1.32	.57	.11	-.41	.22	502	2.59	.05	1.11	1.22	.40	.11	-.66	.22
BO-U$_{T1}$	1,025	2.02	.03	.99	.98	.90	.08	.26	.15	526	2.13	.04	1.03	1.06	.72	.11	-.19	.21	499	1.91	.04	.93	.87	1.11	.11	.99	.22
BO-U$_{T2}$	1,030	2.08	.03	.98	.97	.91	.08	.39	.15	523	2.25	.05	1.05	1.10	.68	.11	-.18	.21	507	1.90	.04	.88	.78	1.16	.11	1.39	.22
BO-U$_{T3}$	995	2.08	.03	.99	.99	.98	.08	.64	.16	508	2.28	.05	1.04	1.09	.84	.11	.30	.22	487	1.87	.04	.90	.81	1.12	.11	1.05	.22
BO-U$_{T4}$	975	2.25	.03	1.05	1.10	.80	.08	.08	.16	475	2.44	.05	1.04	1.08	.66	.11	-.02	.22	500	2.08	.05	1.02	1.05	1.00	.11	.42	.22
BO-O$_{T1}$	1,020	1.66	.03	.89	.79	1.56	.08	2.36	.15	524	1.66	.04	.93	.86	1.53	.11	1.95	.21	496	1.65	.04	.84	.71	1.60	.11	2.91	.22
BO-O$_{T2}$	1,026	1.78	.03	.93	.86	1.18	.08	.79	.15	522	1.77	.04	.93	.87	1.13	.11	.46	.21	504	1.79	.04	.93	.86	1.24	.11	1.17	.22
BO-O$_{T3}$	994	1.96	.03	1.02	1.03	.96	.08	.16	.16	506	1.94	.05	1.03	1.06	1.05	.11	.43	.22	488	1.98	.05	1.01	1.01	.87	.11	-.10	.22
BO-O$_{T4}$	979	2.32	.04	1.18	1.39	.65	.08	-.54	.16	478	2.24	.05	1.15	1.32	.64	.11	-.55	.22	501	2.39	.05	1.21	1.45	.65	.11	-.58	.22
German																											
Previous Grade$_{T0}$	1,050	5.11	.02	.57	.33	-.21	.08	.88	.15	546	4.98	.02	.56	.31	-.40	.11	1.61	.21	504	5.25	.03	.56	.31	-.05	.11	-.01	.22
AI$_{T1}$	1,023	3.41	.03	1.00	1.00	-.19	.08	-.67	.15	524	3.29	.04	1.01	1.01	-.13	.11	-.72	.21	499	3.54	.04	.97	.95	-.25	.11	-.61	.22
ASC$_{T1}$	1,021	3.77	.03	.88	.78	-.38	.08	-.40	.15	523	3.65	.04	.89	.80	-.36	.11	-.23	.21	498	3.90	.04	.85	.73	-.39	.11	-.67	.22
BO-I$_{T1}$	1,019	1.81	.03	.94	.89	1.29	.08	1.33	.15	521	1.94	.04	1.02	1.04	1.09	.11	.62	.21	498	1.67	.04	.84	.70	1.50	.11	2.36	.22
BO-I$_{T2}$	1,026	1.96	.03	1.02	1.05	1.13	.08	.79	.15	520	2.14	.05	1.08	1.16	.92	.11	.26	.21	506	1.78	.04	.93	.86	1.38	.11	1.67	.22
BO-I$_{T3}$	992	2.26	.04	1.12	1.25	.76	.08	-.22	.16	503	2.37	.05	1.14	1.31	.69	.11	-.37	.22	489	2.13	.05	1.08	1.16	.83	.11	-.05	.22
BO-I$_{T4}$	976	2.45	.04	1.14	1.29	.54	.08	-.48	.16	476	2.57	.05	1.17	1.37	.47	.11	-.57	.22	500	2.34	.05	1.09	1.19	.59	.11	-.40	.22
BO-U$_{T1}$	1,019	2.00	.03	.98	.97	.85	.08	.14	.15	521	2.08	.04	1.01	1.03	.75	.11	-.10	.21	498	1.90	.04	.94	.89	.96	.11	.46	.22
BO-U$_{T2}$	1,023	2.05	.03	.99	.98	.86	.08	.30	.15	518	2.16	.04	1.01	1.03	.76	.11	.14	.21	505	1.93	.04	.95	.90	.97	.11	.54	.22
BO-U$_{T3}$	991	2.24	.03	1.01	1.01	.62	.08	-.15	.16	505	2.32	.05	1.02	1.05	.59	.11	-.24	.22	486	2.15	.04	.98	.97	.66	.11	-.05	.22
BO-U$_{T4}$	967	2.38	.03	.99	.98	.52	.08	-.23	.16	467	2.41	.05	.98	.96	.47	.11	-.30	.23	500	2.35	.04	1.00	1.00	.58	.11	-.16	.22
BO-O$_{T1}$	1,018	1.61	.03	.86	.73	1.69	.08	2.91	.15	520	1.68	.04	.89	.79	1.48	.11	2.13	.21	498	1.53	.04	.81	.66	1.95	.11	4.12	.22
BO-O$_{T2}$	1,021	1.71	.03	.93	.87	1.35	.08	1.36	.15	517	1.82	.04	.99	.99	1.16	.11	.74	.21	504	1.60	.04	.85	.73	1.58	.11	2.26	.22
BO-O$_{T3}$	986	1.77	.03	.89	.79	1.16	.08	.97	.16	503	1.93	.04	.92	.85	.82	.11	.14	.22	483	1.60	.04	.83	.69	1.66	.11	2.80	.22
BO-O$_{T4}$	965	1.76	.03	.87	.75	1.28	.08	1.61	.16	466	1.91	.04	.89	.79	.98	.11	.90	.23	499	1.62	.04	.82	.68	1.65	.11	2.96	.22

Note. Descriptive statistics taken from SPSS-outputs (listwise deletion for missing values). AI = academic interest. ASC = academic-self-concept. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. T0 – T4 = Waves of measurement. Previous grade was inversely coded regarding original numeric values (see method section).

Table 4.2

Manifest (Upper Array) and Latent Correlations (Lower Array) of Study Constructs in Mathematics

	$BO\text{-}I_{T1}$	$BO\text{-}U_{T1}$	$BO\text{-}O_{T1}$	$BO\text{-}I_{T2}$	$BO\text{-}U_{T2}$	$BO\text{-}O_{T2}$	$BO\text{-}I_{T3}$	$BO\text{-}U_{T3}$	$BO\text{-}O_{T3}$	$BO\text{-}I_{T4}$	$BO\text{-}U_{T4}$	$BO\text{-}O_{T4}$	$I_{BO\text{-}I}$	$S_{BO\text{-}I}$	$I_{BO\text{-}U}$	$S_{BO\text{-}U}$	$I_{BO\text{-}O}$	$S_{BO\text{-}O}$	ASC_{T1}	AI_{T1}	Previous $Grade_{T0}$	Gender
$BO\text{-}I_{T1}$		**.547**	**.264**	**.463**	**.281**	**.166**	**.308**	**.171**	**.171**	**.185**	.008	**.088**							**-.266**	**-.518**	**.074**	.013
$BO\text{-}U_{T1}$	**.794**		**.131**	**.315**	**.446**	**.096**	**.222**	**.341**	**.082**	**.083**	**.141**	-.018							**.070**	**-.162**	-.058	**-.108**
$BO\text{-}O_{T1}$	**.375**	**.192**		**.197**	.031	**.502**	**.106**	-.025	**.335**	**.116**	-.030	**.167**							**-.348**	**-.225**	**.146**	-.009
$BO\text{-}I_{T2}$	**.606**	**.466**	**.263**		**.514**	**.331**	**.342**	**.160**	**.177**	**.258**	.043	**.140**							**-.242**	**-.399**	.042	**-.069**
$BO\text{-}U_{T2}$	**.448**	**.599**	**.089**	**.695**		**.145**	**.238**	**.398**	.042	**.163**	**.269**	-.012							**.113**	-.039	-.062	**-.175**
$BO\text{-}O_{T2}$	**.256**	**.111**	**.689**	**.432**	**.182**		**.186**	-.019	**.440**	**.153**	**-.107**	**.324**							**-.307**	**-.225**	**.155**	.012
$BO\text{-}I_{T3}$	**.381**	**.303**	**.197**	**.360**	**.248**	**.200**		**.344**	**.352**	**.323**	.045	**.109**							**-.129**	**-.276**	.030	-.049
$BO\text{-}U_{T3}$	**.271**	**.450**	.023	**.234**	**.485**	-.015	**.474**		-.033	**.096**	**.452**	**-.151**							**.232**	**.087**	**-.161**	**-.205**
$BO\text{-}O_{T3}$	**.200**	**.079**	**.488**	**.207**	.023	**.506**	**.450**	-.040		**.166**	**-.161**	**.352**							**-.268**	**-.187**	**.199**	.022
$BO\text{-}I_{T4}$	**.221**	**.194**	**.175**	**.283**	**.170**	**.196**	**.397**	**.159**	**.212**		**.071**	**.462**							**-.145**	**-.266**	**.079**	.037
$BO\text{-}U_{T4}$	.019	**.225**	-.065	.036	**.349**	**-.124**	.071	**.577**	**-.217**	.086		**-.275**							**.260**	**.152**	**-.229**	**-.173**
$BO\text{-}O_{T4}$	**.144**	.045	**.278**	**.161**	-.025	**.359**	**.170**	**-.156**	**.464**	**.594**	**-.346**								**-.266**	**-.209**	**.241**	.062
$I_{BO\text{-}I}$	**.273**	**.240**	**.217**	**.350**	**.210**	**.242**	**.491**	**.196**	**.262**	**.809**	.167	**.287**										
$S_{BO\text{-}I}$	**-.386**	**-.263**	-.058	**-.195**	**-.191**	-.019	**.133**	-.049	.054	**.572**	.143	.134	**.707**									
$I_{BO\text{-}U}$	.021	**.245**	-.071	.039	**.380**	**-.135**	.077	**.628**	**-.236**	.147	**.918**	**-.348**	.181	.156								
$S_{BO\text{-}U}$	**-.469**	**-.385**	**-.178**	**-.366**	**-.197**	**-.226**	**-.186**	**.169**	**-.290**	-.023	**.643**	**-.362**	-.028	**.384**	**.701**							
$I_{BO\text{-}O}$	**.188**	.059	**.365**	**.211**	-.033	**.469**	**.223**	**-.204**	**.607**	**.304**	**-.418**	**.764**	**.375**	.175	**-.455**	**-.473**						
$S_{BO\text{-}O}$	-.069	-.057	**-.359**	**-.039**	-.118	**-.186**	.036	**-.227**	**.144**	.137	**-.357**	**.501**	.170	**.231**	**-.389**	**-.306**	**.656**					
ASC_{T1}	**-.306**	**.076**	**-.385**	**-.297**	**.132**	**-.377**	**-.195**	**.215**	**-.322**	**-.129**	**.311**	**-.266**	**-.159**	**.162**	**.339**	**.234**	**-.349**	.017		**.653**	**-.382**	**-.240**
AI_{T1}	**-.639**	**-.212**	**-.269**	**-.511**	-.080	**-.268**	**-.357**	.038	**-.238**	**-.275**	**.191**	**-.208**	**-.340**	**.204**	**.208**	**.321**	**-.272**	-.016	**.671**		**-.186**	**-.170**
Previous $Grade_{T0}$	**-.067**	.050	**-.169**	**-.066**	**.088**	**-.195**	**-.054**	**.160**	**-.220**	-.056	**.246**	**-.251**	-.070	-.003	**.268**	**.203**	**-.328**	**-.169**	**.373**	**.182**		**.102**
Gender	-.050	**-.173**	-.023	-.040	**-.188**	-.007	-.017	**-.201**	.023	.005	**-.207**	.055	.006	.052	**-.225**	-.046	.072	.094	**-.233**	**-.154**	**-.095**	

Note. Significant correlations ($p < .05$) highlighted in boldface. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. I = latent intercept factor. S = latent slope factor. ASC = academic self-concept. AI = academic interest. T1-T4 represent waves of measurement.

Gender (1 – *male*, 2 – *female*) was treated as a metric variable (Sedlmeier & Renkewitz, 2013). Latent correlations taken from the MIMIC model outputs.

Underlying sample sizes for manifest vs. latent correlation may differ due to differential treatment of missing data (pairwise deletion for manifest correlations;

FIML for latent correlations).

Table 4.3

Manifest (Upper Array) and Latent Correlations (Lower Array) of Study Constructs in German

	BO-I$_{T1}$	BO-U$_{T1}$	BO-O$_{T1}$	BO-I$_{T2}$	BO-U$_{T2}$	BO-O$_{T2}$	BO-I$_{T3}$	BO-U$_{T3}$	BO-O$_{T3}$	BO-I$_{T4}$	BO-U$_{T4}$	BO-O$_{T4}$	I$_{BO-I}$	S$_{BO-I}$	I$_{BO-U}$	S$_{BO-U}$	I$_{BO-O}$	S$_{BO-O}$	ASC$_{T1}$	AI$_{T1}$	Previous Grade$_{T0}$	Gender
BO-I$_{T1}$		**.591**	**.331**	**.444**	**.344**	**.174**	**.270**	**.203**	**.196**	**.172**	.015	**.107**							**-.339**	**-.534**	**.128**	**-.144**
BO-U$_{T1}$	**.805**		**.324**	**.390**	**.471**	**.181**	**.265**	**.329**	**.171**	**.131**	**.147**	**.123**							**-.084**	**-.284**	**.083**	**-.093**
BO-O$_{T1}$	**.429**	**.419**		**.258**	**.197**	**.442**	**.209**	**.109**	**.381**	**.146**	**.005**	**.262**							**-.304**	**-.217**	**.225**	**-.086**
BO-I$_{T2}$	**.525**	**.491**	**.297**		**.585**	**.380**	**.362**	**.231**	**.215**	**.178**	**.100**	**.133**							**-.206**	**-.373**	**.069**	**-.177**
BO-U$_{T2}$	**.493**	**.628**	**.248**	**.794**		**.298**	**.267**	**.347**	**.134**	**.126**	**.155**	**.078**							-.055	**-.191**	.042	**-.115**
BO-O$_{T2}$	**.295**	**.247**	**.545**	**.490**	**.384**		**.228**	**.112**	**.427**	**.082**	-.007	**.277**							**-.238**	**-.148**	**.213**	**-.120**
BO-I$_{T3}$	**.364**	**.336**	**.232**	**.331**	**.301**	**.197**		**.511**	**.289**	**.317**	**.161**	**.145**							**-.142**	**-.273**	.023	**-.107**
BO-U$_{T3}$	**.318**	**.453**	**.159**	**.295**	**.429**	**.123**	**.689**		**.109**	**.159**	**.287**	.008							.064	**-.118**	**-.073**	**-.086**
BO-O$_{T3}$	**.259**	**.234**	**.519**	**.220**	**.187**	**.468**	**.344**	**.158**		**.104**	-.068	**.354**							**-.226**	**-.171**	**.201**	**-.183**
BO-I$_{T4}$	**.221**	**.196**	**.188**	**.259**	**.206**	**.167**	**.364**	**.249**	**.177**		**.339**	**.366**							**-.170**	**-.260**	.046	**-.099**
BO-U$_{T4}$	.064	**.194**	.029	**.119**	**.240**	.002	**.189**	**.344**	-.054	**.455**		.061							**.116**	.009	-.056	-.030
BO-O$_{T4}$	**.166**	**.181**	**.406**	**.149**	**.130**	**.403**	**.148**	.019	**.528**	**.484**	.090								**-.232**	**-.136**	**.133**	**-.164**
I$_{BO-I}$	**.298**	**.264**	**.253**	**.350**	**.278**	**.225**	**.490**	**.335**	**.238**	**.742**	**.412**	**.229**										
S$_{BO-I}$	**-.344**	**-.338**	-.097	**-.163**	**-.224**	-.063	.129	-.002	-.015	**.511**	**.317**	.068	**.688**									
I$_{BO-U}$	.091	**.278**	.042	**.170**	**.344**	.003	**.270**	**.493**	-.078	**.437**	**.699**	-.200	**.590**	**.453**								
S$_{BO-U}$	**-.563**	**-.566**	**-.288**	**-.387**	**-.409**	**-.269**	-.142	-.100	**-.319**	.141	**.342**	**-.359**	.190	**.722**	**.490**							
I$_{BO-O}$	**.203**	**.222**	**.498**	**.183**	**.159**	**.495**	**.182**	.023	**.648**	**.209**	**-.172**	**.815**	**.282**	.083	-.245	**-.441**						
S$_{BO-O}$	**-.279**	-.163	**-.347**	**-.213**	**-.171**	**-.210**	-.122	**-.200**	.005	-.028	**-.239**	**.357**	-.038	.234	-.342	-.097	**.438**					
ASC$_{T1}$	**-.364**	**-.099**	**-.326**	**-.231**	-.056	**-.281**	**-.181**	.035	**-.280**	**-.146**	**.164**	**-.241**	**-.197**	.076	**.234**	**.295**	**-.296**	.125		**.603**	**-.323**	**.140**
AI$_{T1}$	**-.618**	**-.367**	**-.247**	**-.443**	**-.269**	**-.207**	**-.334**	**-.154**	**-.192**	**-.252**	.012	**-.141**	**-.339**	**.190**	.017	**.371**	**-.173**	**.154**	**.617**		**-.133**	**.127**
Previous Grade$_{T0}$	**-.114**	**-.077**	**-.271**	**-.101**	-.054	**-.232**	**-.073**	-.004	**-.227**	-.050	.066	**-.188**	-.067	.056	.095	**.160**	**-.231**	**.121**	**.324**	**.132**		**-.240**
Gender	**-.195**	**-.142**	**-.131**	**-.178**	**-.129**	**-.131**	**-.138**	**-.099**	**-.174**	**-.110**	-.055	**-.222**	**-.149**	.061	-.079	.095	**-.273**	-.127	**.143**	**.135**	**.245**	

Note. Significant correlations ($p < .05$) highlighted in boldface. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. I = latent intercept factor. S = latent slope factor. ASC = academic self-concept. AI = academic interest. T1-T4 represent waves of measurement.

Gender (1 – *male*, 2 – *female*) was treated as a metric variable (Sedlmeier & Renkewitz, 2013). Latent correlations taken from the MIMIC model outputs.

Underlying sample sizes for manifest vs. latent correlation may differ due to differential treatment of missing data (pairwise deletion for manifest correlations; FIML for latent correlations).

Measurement Invariance Testing

Details and model fits for measurement invariance testing are reported in the Appendix (Tables A4.4 – A4.6). To summarize, the 3F-model of boredom showed scalar measurement invariance across genders for German, whereas the mathematics model showed partially scalar measurement invariance across genders for T2 and T3. Hence, requirements were met for using gender as an exogenous predictor of latent boredom development. Both ASC and AI showed scalar measurement invariance across genders for German, and partially scalar measurement invariance across genders for mathematics at T1, allowing the regression of appraisals on gender. Scalar measurement invariance over time held for the 3F-model of boredom in both subject domains, fulfilling the prerequisite for LGCM. In sum, all three of the planned steps could be executed and interpreted in accordance with conventional modeling demands (Byrne, 2012; Jöreskog, 1971; Kleinke et al., 2017).

Main Analyses

Table 4.4 shows fit indexes for LGC models (Step 1) and MIMIC models (Steps 2 and 3) of academic boredom development over time. Overall, analytical models approximated the underlying data well in both subjects ($.95 \leq$ CFI $\leq .99$, $.013 \leq$ RMSEA $\leq .032$, $.032 \leq$ SRMR $\leq .048$, see Hu & Bentler, 1999). For linear LGC models (Step 1), unstandardized parameter estimates, i.e., mean values and variances alongside standard errors, are listed in Table 4.5. Standardized predictive effects (βs) for our MIMIC models (Steps 2 and 3), including mediations by appraisals and performance indicators, are depicted in Figure 4.3. An overview of standardized indirect effects (βs) of gender, including variance accounted for in boredom growth factors (R^2) is presented in Table 4.6.

Table 4.4

Fit Indexes for Linear Latent Growth Curve (LGC) and MIMIC Models

	Sample Size and Missingness		Model Fit					
Model	*n*	Range of Covariance Coverage	χ^2	*df*	SCF	CFI	RMSEA [90%-CI]	SRMR
			Mathematics					
LGC (Step 1)	1,418	.488-.723	438.755***	207	1.2452	.965	.028 [.024-.032]	.046
MIMIC (Step 2)[a]	1,418	.488-.723	508.178***	225	1.2250	.958	.030 [.026-.033]	.047
MIMIC (Step 3)[a,b]	1,418	.488-.755	682.507***	275	1.1889	.951	.032 [.029-.035]	.048
			German					
LGC (Step 1)	1,414	.483-.721	260.452**	207	1.3167	.991	.014 [.007-.018]	.033
MIMIC (Step 2)[a]	1,414	.483-.721	275.823*	225	1.2962	.992	.013 [.006-.017]	.032
MIMIC (Step 3)[a,b]	1,415	.483-.742	403.889***	275	1.2684	.983	.018 [.014-.022]	.034

Note. * $p < .05$. ** $p < .01$. *** $p < .001$. SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019). Longitudinal invariance constraints were carried over from scalar invariance models (see Table A4.6 of the Appendix). Latent slope factor loadings were $\lambda_{T1} = -10$ / $\lambda_{T2} = -8.377$ / $\lambda_{T3} = -4.979$ / $\lambda_{T4} = 0$ (increasing linear trajectory).

[a] Gender added as exogenous predictor to latent growth factors.

[b] Appraisals and previous grade added as mediators to gender-boredom-relations.

Table 4.5

Unstandardized Parameter Estimates for Latent Growth Factors in Linear LGC Models

	Intercept		Slope	
Academic Boredom	*M* (*SE*)	*Var* (*SE*)	*M* (*SE*)	*Var* (*SE*)
Mathematics (n = 1,418)				
BO-I	2.54 (.06)***	.61 (.12)***	.08 (.01)***	.01 (.002)***
BO-U	2.23 (.06)***	.69 (.10)***	.02 (.01)*	.01 (.002)***
BO-O	2.34 (.04)***	.67 (.08)***	.07 (.01)***	.01 (.001)***
German (n = 1,414)				
BO-I	2.52 (.07)***	.52 (.11)***	.07 (.01)***	.01 (.001)***
BO-U	2.41 (.05)***	.36 (.08)***	.04 (.01)***	.01 (.001)***
BO-O	1.82 (.03)***	.39 (.06)***	.02 (.004)***	.003 (.001)***

Note. BO-I = Boredom intensity. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge. ** $p < .01$. *** $p < .001$.

Primary Results on Gender Differences

Effects of Gender on Latent Boredom Growth. For mathematics, gender significantly predicted BO-U-levels at T4 (i.e., the latent intercept factor), with moderately higher levels for boys vs. girls ($\beta = -.22$, $p < .001$; moderate effect size for $.10 \leq \beta \leq .25$, see Keith, 2006). For German, gender significantly predicted BO-I-levels ($\beta = -.15$, $p < .01$) and BO-O-levels at T4 ($\beta = -.23$, $p < .001$), indicating moderately to highly enhanced levels for boys vs. girls (large effect size for $\beta > .25$, see Keith, 2006). None of the other latent growth factors were significantly predicted by gender ($.02 \leq |\beta| \leq .13$, respective *p*s > .05; small effect size for $\beta < .10$, see Keith, 2006).

Variance in latent boredom growth accounted for by gender predictions, was altogether small ($.00 \leq R^2 \leq .08$; small effect size for $.02 \leq R^2 < .13$, see Cohen, 1988). Only BO-U-levels at T4 in mathematics ($R^2 = .05$), and BO-O-levels at T4 in German ($R^2 = .08$) showed significant levels of explained variance.

Mediated Effects of Gender on Latent Boredom Growth. When controlling for appraisals (i.e., ASC and AI) and prior school achievement in mathematics, gender significantly predicted BO-U-levels at T4, with moderately higher levels for boys vs. girls (β = -.15, $p < .01$). Besides, gender significantly predicted AI (β = -.15, $p < .001$), ASC (β = -.23, $p < .001$), and previous grade (β = -.10, $p < .01$), all with moderately higher values for boys vs. girls. For German, gender significantly predicted BO-I-levels (β = -.11, $p < .05$) and BO-O-levels at T4 (β = -.22, $p < .001$), with moderately higher levels for boys vs. girls. Gender also significantly predicted latent linear BO-O-trajectories (i.e., the latent slope factor, β = -.19, $p < .01$), with moderately steeper increases for boys vs. girls over the course of our investigation period (T1-T4). Besides, gender predicted AI (β = .14, $p < .001$), ASC (β = .14, $p < .001$), and previous grade (β = .25, $p < .001$) in German, all with moderately higher values for girls vs. boys. None of the other latent growth factors were significantly predicted by gender ($.01 \leq |\beta| \leq .11$, respective ps > .05; small effect size for $\beta < .10$, see Keith, 2006).

Total indirect effects of gender via appraisal and performance antecedents from Step 3-MIMIC models were altogether of small effect size ($.00 \leq |\beta| \leq .08$). For mathematics, gender had significant total indirect effects on all three forms of boredom levels at T4 ($\beta_{\text{indirect} \rightarrow \text{BO-I}}$ = .04; $\beta_{\text{indirect} \rightarrow \text{BO-U}}$ = -.07; $\beta_{\text{indirect} \rightarrow \text{BO-O}}$ = .08). Additionally, gender had significant total indirect effects on BO-I- and BO-U-trajectories ($\beta_{\text{indirect} \rightarrow \text{BO-I}}$ = -.04; $\beta_{\text{indirect} \rightarrow \text{BO-U}}$ = -.06). For German, gender had significant total indirect effects on BO-I- ($\beta_{\text{indirect} \rightarrow \text{BO-I}}$ = -.04) and BO-O-levels at T4 ($\beta_{\text{indirect} \rightarrow \text{BO-O}}$ = -.06), as well as on BO-U- ($\beta_{\text{indirect} \rightarrow \text{BO-U}}$ = .08) and BO-O-trajectories ($\beta_{\text{indirect} \rightarrow \text{BO-O}}$ = .06).

By including appraisal and performance mediation paths (Step 3-MIMIC model), variance in latent boredom growth accounted for by all predictors increased ($.04 \leq R^2 \leq .17$). For mathematics, BO-I- (R^2 = .13; moderate effect size for $R^2 \geq .13$, see Cohen, 1988), BO-U- (R^2 = .16), and BO-O-levels at T4 (R^2 = .17), as well as the BO-U-trajectory (R^2 = .13) all showed significant, moderate amounts of variance explained. For German, BO-I- (R^2 = .13), BO-U- (R^2 = .10), and BO-O-levels at T4 (R^2 = .15), as well as BO-U- (R^2 = .15) and BO-O-trajectories ($R^2 = .07$), all showed significant, small to moderate levels of explained variance. For further details on mediation analyses results, please refer to Table 4.6.

Secondary Results by Mediator Variable

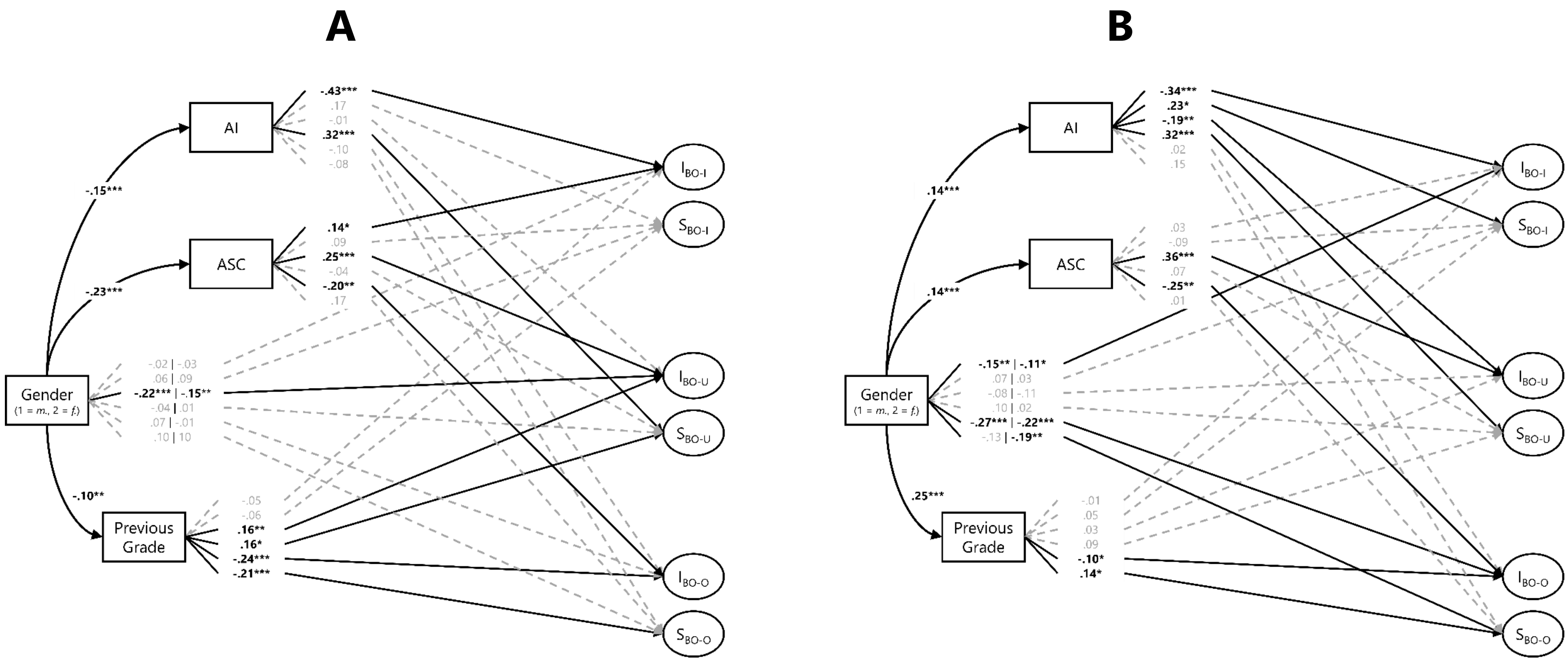
Academic Interest. For mathematics, AI negatively predicted BO-I-levels at T4 ($\beta = -.43$, $p < .001$, large effect size for $\beta > .25$, see Keith, 2006) and positively predicted BO-O-levels at T4 ($\beta = .32$, $p < .001$). For German, AI negatively predicted BO-I-levels ($\beta = -.34$, $p < .001$) and BO-U-levels at T4 ($\beta = -.19$, $p < .01$). AI also positively predicted linear BO-I-trajectories ($\beta = .23$, $p < .05$) and BO-U-trajectories ($\beta = .32$, $p < .001$). Other predictions by AI did not reach statistical significance ($.02 \leq |\beta| \leq .17$, respective *p*s > .05).

Academic Self-concept. For mathematics, ASC positively predicted BO-I-levels ($\beta = .14$, $p < .05$) and BO-U-levels at T4 ($\beta = .25$, $p < .001$). ASC also negatively predicted BO-O-levels at T4 ($\beta = -.20$, $p < .01$). For German, ASC positively predicted BO-U-levels at T4 ($\beta = .36$, $p < .001$) and negatively predicted BO-O-levels at T4 ($\beta = -.25$, $p < .01$). Other predictions by ASC did not reach statistical significance ($.01 \leq |\beta| \leq .17$, respective *p*s > .05).

Previous Grade. For mathematics, previous grade positively predicted both BO-U-levels at T4 ($\beta = .16$, $p < .01$) and the BO-U-trajectory ($\beta = .16$, $p < .05$). Previous grade also negatively predicted both, BO-O-levels at T4 ($\beta = -.18$, $p < .01$) and the BO-O-trajectory ($\beta = -.21$, $p < .01$). For German, previous grade negatively predicted BO-O-levels at T4 ($\beta = -.10$, $p < .05$), while positively predicting the BO-O-trajectory across time ($\beta = .14$, $p < .05$). Other predictions by previous grade did not reach statistical significance ($.00 \leq |\beta| \leq .09$, respective *p*s > .05).

Figure 4.3

Results for Standardized Predictive Paths (βs) of MIMIC Models for Mathematics (A) and German (B)



Note. Negative paths indicate higher values for boys, positive paths indicate higher values for girls. Gender effect β's include results for Step 2- (first coefficient) and Step 3-MIMIC models (second coefficient; after including mediation paths). Significant predictions ($p < .05$) were highlighted. Nonsignificant predictions are indicated by dashed lines. * $p < .05$. ** $p < .01$. *** $p < .001$. Gender was codes as 1 – *male*, 2 – *female*. AI = academic interest. ASC = academic-self-concept. I = latent intercept factor. S = latent slope factor. Latent boredom factors, as well as latent intercept- and slope factor loadings were included but are omitted in depiction. Correlations of study variables were allowed at each individual wave of measurement, as well as for all latent intercept and slope factors (omitted here). See results section for further details.

Table 4.6

Summary of Standardized Gender Effects on Latent Growth in Academic Boredom

Gender effects on	Mathematics							German						
	Direct	Indirect via ASC	Indirect via AI	Indirect via Previous Grade	Total indirect	Total	R^2 for growth factor	Direct	Indirect via ASC	Indirect via AI	Indirect via Previous Grade	Total indirect	Total	R^2 for growth factor
MIMIC (Step 2)														
BO-I														
→ Intercept	-.02 (.95)						.00 (.00)	**-.15 (.05)****						.02 (.02)
→ Slope	.06 (.05)						.004 (.01)	.07 (.06)						.01 (.01)
BO-U														
→ Intercept	**-.22 (.05)*****						**.05 (.02)***	-.08 (.06)						.01 (.01)
→ Slope	-.04 (.06)						.00 (.01)	.10 (.07)						.01 (.01)
BO-O														
→ Intercept	.07 (.05)						.01 (.01)	**-.27 (.05)*****						**.08 (.03)****
→ Slope	.10 (.06)						.01 (.01)	-.13 (.07)						.02 (.02)
MIMIC (Step 3)														
BO-I														
→ Intercept	-.03 (.02)	-.03 (.05)	**.07 (.02)*****	.00 (.00)	**.04 (.02)***	.01 (.05)	**.13 (.04)****	**-.11 (.05)***	.01 (.01)	**-.05 (.01)****	-.00 (.01)	**-.04 (.01)****	**-.15 (.05)****	**.13 (.04)****
→ Slope	.09 (.06)	-.02 (.02)	-.03 (.02)	.01 (.01)	**-.04 (.02)***	.05 (.05)	.05 (.04)	.03 (.05)	-.01 (.02)	.03 (.02)	.01 (.02)	.03 (.02)	.06 (.06)	.04 (.03)
BO-U														
→ Intercept	**-.15 (.05)****	**-.06 (.02)*****	.00 (.01)	**-.02 (.01)***	**-.07 (.01)*****	**-.23 (.05)*****	**.16 (.04)*****	-.11 (.07)	**.05 (.02)****	**-.03 (.01)***	.01 (.02)	.03 (.02)	-.08 (.06)	**.10 (.04)***
→ Slope	.01 (.06)	.01 (.02)	**-.05 (.02)****	**-.02 (.01)***	**-.06 (.02)****	-.05 (.06)	**.13 (.05)****	.02 (.06)	.01 (.01)	**.04 (.02)****	.02 (.02)	**.08 (.02)*****	.10 (.07)	**.15 (.05)****
BO-O														
→ Intercept	-.01 (.06)	**.05 (.02)***	.02 (.01)	**.02 (.01)****	**.08 (.02)*****	.07 (.05)	**.17 (.04)*****	**-.22 (.05)*****	**-.04 (.02)***	.00 (.01)	**-.03 (.01)***	**-.06 (.02)*****	**-.27 (.05)*****	**.15 (.04)****
→ Slope	.10 (.07)	-.04 (.02)	.01 (.01)	**.02 (.01)***	-.01 (.02)	.10 (.06)	.05 (.03)	**-.19 (.06)****	.00 (.02)	.02 (.02)	**.04 (.02)***	**.06 (.02)****	-.13 (.07)	**.07 (.03)***

Note. Gender was coded as 1 = male, 2 = female. ASC = academic self-concept. AI = academic interest. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. Standard errors (*SE*) in parentheses. Significant parameter estimates in boldface. * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This longitudinal study presents a multi-domain examination of gender differences in academic boredom and its development in secondary education. We included appraisal and performance differences as mediators to gender-boredom-relations in our analyses. Therefore, observed gender differences in boredom levels and trajectories cannot be traced back to appraisal or performance differences between genders. Findings for gender differences in boredom were rather consistent, with and without controlling for appraisal and performance differences. As expected, boys experienced higher levels of boredom due to underchallenge in mathematics (cf. Daschmann et al., 2011; Goetz & Frenzel, 2010), higher levels of boredom intensity in German (cf. Raccanello et al., 2019; Zaccoletti et al., 2020), and higher levels of boredom due to overchallenge in German at Grade 8 (H1a). However, boys and girls reported comparable levels of boredom intensity in mathematics at Grade 8 (H1a). Contrary to our hypothesis, girls did not experience higher levels of boredom due to overchallenge in mathematics, nor did they show higher levels of boredom due to underchallenge in German at Grade 8 (comparable levels for both genders; H1b). As expected, boredom intensity, boredom due to underchallenge, and boredom due to overchallenge increased over time in both domains (H2). Developmental slopes for boredom intensity (mathematics and German), boredom due to underchallenge (mathematics and German), and boredom due to overchallenge in mathematics were not related to gender (RQ2, cf. Spaeth et al., 2015). However, in German boys experienced steeper increases than girls in boredom due to overchallenge over the first half of secondary school (RQ2). Before discussing these findings, we point out the limitations of our study that must be considered when interpreting our findings.

Limitations

Conceptually, we adopted a developmental perspective, modeling latent growth in three forms of subject-specific boredom. In doing so, we were able to estimate individual and mean latent growth curves. While this approach provides a detailed understanding of individual differences in developmental trajectories of the given constructs over time, it neglects their interplay, i.e., potential cross-lagged relations. Several studies examining longitudinal boredom used reciprocal effects models instead (cf. Forsblom et al., 2021; Pekrun et al., 2014, 2017). We decided against this approach because we were explicitly interested in gender differences in boredom *trajectories* over time. Besides, by using the MIMIC modelling approach, we were able to provide predictive effect sizes associated with gender that can readily be compared to predictive effects of other metric predictors across studies. Methodologically, our study benefits from the large sample and the longitudinal data structure, spanning a considerable portion of secondary school (Grades 5-8). However, project constraints limited us to two-item indicators for each boredom construct. Consequently, (partially) scalar measurement invariance models of boredom across genders showed only acceptable fit in some cases (e.g., CFI .923 and .947; Hu & Bentler, 1999) in the mathematics domain from T2 onwards. Invariance constraints across genders held, however, at an at least partially scalar level (see results section and Table A4.4 & A4.5 of the Appendix). We analysed mathematics and German separately to account for domain-specific factors in academic emotions (Goetz et al., 2007). This allowed us to compare results for both domains side-by-side but not in a comprehensive fashion. One could paint a more complete picture of boredom at school by including cross-domain relations, often referred to as dimensional comparisons (Möller & Marsh, 2013) at the cost of greater model complexity. Besides, the present study only assessed gender as a binary variable (male vs. female). “Male” and “female” are the two academically stereotyped genders (Cvencek et al., 2011;

Eagly et al., 2020; Vuletich et al., 2020) and, hence, have been included in various studies of academic gender differences (e.g., Petersen, 2018; Keller et al., 2021; Goetz & Frenzel, 2010). Gravely however, nonbinary students are not represented in our assessment. Future gender assessments should use additional categories and/or an open text response to capture student gender identities more accurately. Active sampling for nonbinary students is furthermore required to improve gender research and enhance its methodological and analytical complexity in future investigations. Lastly, study participants attended Grades 5-8 of the German Gymnasium. Study results can therefore not be generalized to other segments of German secondary education or other school tracks.

Discussion of Findings and Future Directions

Main Findings on Gender Differences

As expected, in mathematics we found no gender differences in boredom intensity levels and higher levels of boredom due to underchallenge in boys than girls. However, contrary to our expectations and previous findings (Daschmann et al., 2011; Goetz & Frenzel, 2010), girls did not show higher levels of boredom due to overchallenge in mathematics in our sample. This is especially surprising, as mathematics appraisals and mathematics performance were both lower for girls vs. boys, reflecting past research (Frenzel et al., 2007; Keller et al., 2021; OECD, 2012), and supposedly leading to increased boredom (Pekrun, 2006). We can only speculate about possible explanations to this finding. Possibly, this reverts to gender differences in emotional response patterns regarding perceived overchallenge. Boys could, for instance, respond to overchallenge by becoming bored with a higher probability than girls, whereas girls may respond differently (e.g., by becoming more eager to compensate for increasing difficulty, or by feeling frustrated rather than bored, etc.). For further clarification, future work should address gender differences in immediate responses to experiences of under- or overchallenge.

In line with our expectations for the domain of German, boys reported higher levels of boredom intensity and boredom due to overchallenge than girls. Increased levels of first language boredom in boys are well aligned with previous studies (Raccanello et al., 2019; Zaccoletti et al., 2020). Unlike for our mathematics results, boys' higher levels of boredom due to overchallenge in German may be explained by their observed lower appraisals and performance (see also Keller et al., 2021; Mostafa & Schwabe, 2019; Petersen, 2018). Despite their higher performance, girls unexpectedly did not experience higher levels of boredom due to underchallenge in German. Again, on a speculative note, this finding could be linked to differential emotional responses in girls vs. boys regarding perceived challenge (cf. comparable levels of boredom due to overchallenge in mathematics). Girls may hence respond differently than boys to perceived challenge mismatches in class, altogether. Regarding boredom development over time, increasing trajectories for all forms of academic boredom in both subjects reflect broad motivational declines in adolescence (for a meta-analysis, see Scherrer & Preckel, 2019), and are well aligned with past results (Feuchter & Preckel, 2022b; Spaeth et al., 2015; Vierhaus et al., 2016; Weybright et al., 2020). Increasing boredom over the course of adolescence is also reflective of assumptions made in stage-environment fit theory (SEFT; Eccles et al., 1993). According to SEFT, a mismatch between adolescent learners' developing needs and the opportunities provided by the secondary school environment may result in increasing levels of boredom (and decreasing levels of overall motivation) in adolescence. Concurrently, academic achievement is devalued among adolescent peers (Wigfield & Wagner, 2005), potentially making boredom a socially desirable emotional response to academic tasks and challenges. No explicit assumptions are made in SEFT, however, on potential gender differences in assumed increasing patterns of longitudinal academic boredom. In this work, growth rates for boredom did not differ between genders except for boredom due to overchallenge in German (higher growth rates for boys vs. girls), and only when including appraisal and performance antecedents as mediators. Steeper increases in boys' boredom due to overchallenge in German could be associated with unique performance-boredom relations in German. Counterintuitively, better previous grades in German were linked to steeper increases in boredom due to overchallenge (– we discuss this secondary finding further down). According to our results, boys suffer more from these potential changes than girls. Future studies should aim to learn more about subject-specific qualitative task differences and associations with perceived challenge.

Secondary Findings

Regarding all mediator variables, we found similarities as well as differences in associations with boredom for the domains of mathematics and German. Increased academic interest formed a protective factor to experiences of boredom intensity after three-and-a-half years but also accelerated increases in boredom due to underchallenge. The latter might be due to interest-driven investment in learning activities which could in turn lead to actual performance improvement. Additionally for German, this same pattern was also observed regarding levels of boredom due to underchallenge and the boredom intensity trajectory. This indicates an extended significance for academic interest in German boredom development, albeit by similar mechanisms when compared to the domain of mathematics. Experiences of overchallenge do not revert to academic interest however, further separating boredom due to under- and overchallenge as conceptually distinct forms (cf. Acee et al., 2010; Daschmann et al., 2011).

Academic self-concept showed associations with later levels of challenge-related forms of boredom, i.e., negative predictions of later boredom due to overchallenge and positive prediction of later boredom due to underchallenge. This contrasting pattern of self-concept predicting challenge-related boredom is highly plausible and well aligned with previous work (Goetz & Frenzel, 2010). Only for mathematics, academic self-concept also positively predicted later levels of boredom intensity. Like academic interest for German, academic self-concept seems to have a larger significance regarding boredom development for the domain of mathematics. Furthermore, boredom’s developmental trajectories are unimpacted by academic self-concept. This underlines the fact that commonly increasing boredom over the course of secondary school (cf. Feuchter & Preckel, 2022b; Meyer & Schlesier, 2021; Spaeth et al., 2015; Vierhaus et al., 2016; Weybright et al., 2020) should be tackled by classroom-based rather than ability self-concept interventions.

Previous grade as a marker of initial academic performance had a protective effect regarding later levels of boredom due to overchallenge. Besides, previous grade showed highly differential patterns of boredom development for both subjects. For mathematics, previous grade positively predicted later levels of boredom due to underchallenge, as well as its trajectory. This mirrors the role academic self-concept had to play in mathematics boredom due to underchallenge development, with the extension that initial performance incrementally exacerbated its increase over time. This is both good and bad, as reduced perceived challenge signals learning success. However, this should not come at the expense of increased boredom, due to its high aversiveness (Goetz et al., 2019). Separating authentic learning from its backlash regarding increased boredom, is, hence, an interventive challenge. Associated underlying mechanisms need further investigation. Surprisingly for the domain of German, previous grade enhanced boredom due to overchallenge, whereas it buffered them in the domain of mathematics. This could indicate a change in the architectural difficulty of German classes throughout early secondary school. While mathematics class is organized rather cumulatively (i.e., increased difficulty reverts to increased complexity), German class can be subject to more radical changes in tasks and assignments (i.e., increased difficulty reverts to novelty).

Alternative Explanations to Main Findings

Taken together, boys' tendency for higher boredom leaves the notion that gender can function as a carrier of specific vulnerability to experiences of boredom in the classroom. Alternate explanations to our findings, reflecting said vulnerability, are now briefly addressed. For instance, boredom itself has often been discussed as a personality trait. In contrast to academic conceptualizations, trait boredom, as captured in constructs like *boredom proneness* (Farmer & Sundberg, 1986) and *boredom susceptibility* (Zuckerman, 1979), has a more deeply rooted biological base and forms an overarching disposition to feeling bored, regardless of context or domain (for a timely and critical discussion of boredom proneness measurement, see Gana et al., 2019). Gender differences regarding trait boredom have been reported in the past and showed higher trait boredom in boys, as compared to girls (Butković & Bratko, 2003; Vodanovich & Kass, 1990). More recently, Czikmantori et al. (2021) showed trait task enjoyment to hold measurement invariance across genders and to negatively predict experiences of boredom. Contrary to trait boredom, trait task enjoyment could be a strong protective factor against boredom. However, gender differences herein need further investigation. As far as instruction is concerned, Parr et al. (2019) showed boys (as well as low-achieving students) to particularly respond to more interactive, dialogic forms of mathematics instruction and to experience less boredom as a consequence. A lack of dynamic teaching could hence explain increased male boredom due to underchallenge in mathematics. Another recent study by Hannover et al. (2022) showed teachers to act in less communal ways towards boys, who are more regularly ignored or lectured and less frequently addressed in understanding and interested ways in comparison to girls. These aspects of dyadic teacher-student-interactions could also have an important impact on students' emotional responses, potentially leading to increased boredom in boys, as we found it. In addition, the observed lack of enhanced boredom for girls in any form of boredom could be due to superior challenge adaptation and self-regulatory competencies. As this is a rather speculative note, future studies should examine gender differences in subject-specific self-regulation.

Implications

The present study highlights gender as a research variable and contextual predictor to academic emotion and its development, surpassing its usual role as a control variable. Consequently, existing work on gender differences in academia can help adapt academically rooted theories of motivation and emotion, such as CVT (Pekrun, 2006). That way, hypotheses regarding statistical effects of gender (i.e., gender differences) could be incorporated more explicitly into educational theorizing. Our latent growth curve framework furthermore stresses the role of developmental trajectories in psychological constructs, beyond more commonly adopted reciprocal effects models (e.g., Forsblom et al., 2021; Pekrun et al., 2017). While this approach partly neglects variable interplay, modeling growth in various constructs has clearer implications for their development over time. These long-term developmental aspects should also be integrated into educational theorizing, leading to specific hypotheses on the shape of construct trajectories over time. Aside from developmental perspectives, the present work shows the importance of differentiating between different forms of academic boredom (see also Acee et al., 2010; Daschmann et al., 2011) as they are differently associated with appraisals and performance. It furthermore highlights the intricacies of subject-specific assessment in educational research (cf. Goetz et al., 2006, 2007, 2010).

On a more practical note, exploring reasons for increasing boredom throughout secondary school could help prevent its detrimental consequences. We still lack knowledge about the reasons for boredom in class. Uncovering those could be achieved, for instance, by using qualitative experience sampling methodology (see Zirkel et al., 2015 for an overview of experience sampling methods), or by engaging in open conversation on academic emotions as part of a psychoeducational intervention during classes. Our results furthermore indicate that increased academic interest can reduce boredom intensity experiences. Academic interest, for instance, can be promoted by designing class content that caters to the interests of students. It is important for teachers and scholars to learn more about the age-specific interest patterns of students in and outside of school, for example by questionnaires concerning students' spare-time activities. Interest profiles could then be used for subject-specific task framing. A potential downside to promoting academic interest actively is given by its associations with increased growth rates in boredom due to underchallenge in our study, making adequate interest interventions especially difficult but also promising.

Conclusions

Results of this work underline the conceptual complexity of boredom as an achievement emotion. Measuring different forms of boredom (i.e., boredom intensity, boredom due to underchallenge, and boredom due to overchallenge) with subject-specific framing (i.e., either as mathematics or German academic emotion) resulted in differential gender differences. Boys generally seem to be at a greater risk of experiencing boredom at school than girls. This assumption should be put under scrutiny in future empirical studies. Given that this continues to be a persistent tendency, possible explanations to increased boredom should be explored. Furthermore, our developmental perspective showed that addressing long-term processes in psychological theorizing can substantially deepen one's understanding of an observed phenomenon. Longitudinal additions to theories are especially warranted when common developmental tendencies (e.g., long-term increases or decreases) for target constructs are known to exist. For all kinds of classroom-based instruction, boredom reduction needs to be a clear desiderate regardless of educational stage.

Chapter 5 – General Discussion

Unique to the present dissertation, extensive longitudinal results were obtained for two subject domains, namely mathematics and German, and three different forms of academic boredom, namely boredom intensity, boredom due to underchallenge, and boredom due to overchallenge. The developmental period of focus for all three studies spanned the first 3.5 years of German secondary education (i.e., grades 5-8).

After detailing individual studies in the previous chapters (see chapter 2-4), I will now summarize and interpret respective results in light of CVT as main theoretical framework (cf. Pekrun, 2019; Pekrun et al., 2023). I will furthermore discuss limitations of the present dissertation and address future directions for academic boredom research. Before ending this chapter by drawing final conclusions, I will specify implications for teaching practice.

Overview of Study Results

Results from Study 1

Study 1 examined different forms of subject-specific academic boredom and their relations to academic self-concepts longitudinally. Results included a construct validation for the 3F-model of boredom (cf. Feuchter & Preckel, 2022b), with boredom intensity, boredom due to underchallenge, and boredom due to overchallenge as same-order correlated factors. Besides, academic boredom and self-concept showed differential cross-sectional, and predictive relations. Academic self-concept was negatively correlated with concurrent boredom intensity and boredom due to overchallenge and positively correlated with concurrent boredom due to underchallenge. Furthermore, academic self-concept could occasionally predict different subject-specific forms of academic boredom in longitudinal analyses but not the other way round.

Results from Study 2

Study 2 addressed subject-specific academic boredom trajectories in early secondary school in regular vs. ability-grouped gifted classes. Results pointed towards increasing trajectories for all examined types of academic boredom from grades 5 through 8 (see also Meyer & Schlesier, 2021; Spaeth et al., 2015; Vierhaus et al., 2016). Ability-grouping showed limited effectiveness and efficacy at buffering these trajectories. Only the intensity of mathematics boredom showed reduced levels in gifted classes after 3.5 years of secondary school. Still, positive effects of ability grouping remain (cf. Steenbergen-Hu et al., 2016), even though its adequacy as a boredom intervention for gifted students needs to be questioned.

Results from Study 3

Study 3 investigated gender differences in subject-specific academic boredom development in early secondary school, while controlling for previous appraisals and academic achievement. Results indicated that few, one-sided gender differences herein exist (i.e., higher boredom in boys for both, mathematics and verbal domains, cf. Goetz & Frenzel, 2010; Zaccoletti et al., 2020). Including appraisal and achievement mediations had little impact on observed gender differences in academic boredom development but reflected scholar gender stereotypes in these constructs (cf. Vuletich et al., 2020). Appraisals and achievement differentially predicted subject-specific boredom trajectories. Academic interest had a more prominent role regarding the domain of German, while academic self-concept and previous grades were able to predict mathematics boredom development to a greater extent.

Implications for CVT

In the following, I will relate results from Studies 1-3 to different procedural steps of CVT’s theoretical framework (Pekrun, 2019; Pekrun et al., 2023, see also Figure 1.3 of the topical introduction). Findings in line with previous studies are expressed whereas new or unexpected findings are discussed, occasionally providing an outlook on potential future research.

Environment

The environment marks the starting point in CVT's process model (Pekrun, 2019). Regarding situational antecedents to academic boredom, I examined the influence of ability-grouped vs. regular classes as different achievement settings (cf. Study 2). Ability grouping, however, does not only constitute an educational setting but an educational intervention. In CVT's framework, educational interventions refer to regulatory processes associated with respective procedural steps. As an intervention, ability grouping can be described as a situation- and competence-oriented regulation technique. Results of Study 2 (Feuchter & Preckel, 2022b) showed that full-time ability grouping has little influence on academic boredom development in secondary school. This is aligned with previous results (Hornstra et al., 2017; Preckel et al., 2010). Merely mathematics boredom development (i.e., levels of boredom intensity and boredom due to overchallenge) showed reduced levels at the end of our investigation in special gifted classes. German boredom development was unimpacted by class type (see Table 5.1 for a summary).

Table 5.1

Selected Study Results Embedded into the CVT-Framework – Environment / Situation

	Mathematics						German					
	BO-I		BO-U		BO-O		BO-I		BO-U		BO-O	
Procedural Step	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory
Environment / Situation												
GC (vs. RC)	↓↓ (|β| = .16[a]; |β| = .21[b])				↓↓ (|β| = .18[a])							

Note. ↓↓ = moderate negative effect.

Predictive effect sizes (i.e., β) evaluated according to Keith (2019): β < .10 = small, β ≤ .25 = moderate, β > .25 = large.

Empty cells indicate that no significant results were obtained.

[a, b] Class type effects in original (a) and matched samples (b).

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. GC = special classes for the gifted. RC = regular classes.

Gender

In CVT, individual characteristics, like gender, form distal individual antecedents to emotional experience (Pekrun, 2019). As such, gender should primarily impact proximal appraisal antecedents rather than exert a direct effect on emotional experience itself. I tested these mediation assumptions, exploring gender differences in academic boredom development (cf. Study 3). Subject-specific findings were as follows (see Table 5.2 for a summary).

Mathematics. Boys showed higher levels in boredom due to underchallenge at the end of our investigation. This is well aligned with previous research (Daschmann et al., 2011; Goetz & Frenzel, 2010; Pekrun et al., 2017). However, gender differences in boredom due to underchallenge lingered beyond the inclusion of appraisals, showing only a partial mediation. Full mediation of gender differences in boredom due to underchallenge might have been achieved by including additional variables (e.g., gender-stereotypical self-perceptions, cf. Cvencek et al., 2011).

German. Boys showed higher levels in boredom intensity and boredom due to overchallenge at the end of our investigation. This tracks for general boredom in verbal domains (Raccanello et al., 2019; Zaccoletti et al., 2020). Challenge-related forms of academic boredom are understudied in verbal domains, making this a novel result. Respective gender differences remained when including appraisal antecedents, seconding partial mediations found in mathematics. Alternative explanations to found gender differences in boredom (besides stereotypes) include gender differences in dispositional constructs (Butković & Bratko, 2003) or in teacher-student-interactions (Hannover et al., 2022). Beyond boredom level differences, boys showed steeper increases in boredom due to overchallenge after the inclusion of appraisal mediations. This is a novel finding, showing that boys are at an increased risk of experiencing overchallenge in German class. To judge the robustness of this finding, future studies should address gender differences in the development of subject-specific challenge perceptions.

Table 5.2

Selected Study Results Embedded into the CVT-Framework – Gender

	Mathematics						German					
	BO-I		BO-U		BO-O		BO-I		BO-U		BO-O	
Procedural Step	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory
Gender[a]												
Boys (vs. Girls)			$\uparrow\uparrow$ ($\lvert\beta\rvert = .15$)				$\uparrow\uparrow$ ($\lvert\beta\rvert = .11$)				$\uparrow\uparrow$ ($\lvert\beta\rvert = .22$)	$\uparrow\uparrow$ ($\lvert\beta\rvert = .19$)

Note. $\uparrow\uparrow$ = moderate positive effect.

Predictive effect sizes (i.e., β) evaluated according to Keith (2019): $\beta < .10$ = small, $\beta \leq .25$ = moderate, $\beta > .25$ = large.

Empty cells indicate that no significant results were obtained.

[a] Gender effects after inclusion of appraisal and achievement mediations.

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge.

Appraisal

Situational control and value appraisals are proximal antecedents to emotional experience in CVT (Pekrun, 2019). Testing predictive assumptions longitudinally, I regressed different forms of academic boredom, and their development over time, on academic self-concept (as a marker of subjective control) and academic interest (as a marker of subjective value) (cf. Studies 1 and 3). Subject-specific findings were as follows (see Table 5.3 for a summary).

Mathematics. Academic self-concept negatively predicted boredom intensity and boredom due to overchallenge, and positively predicted boredom due to underchallenge (cf. Study 1). This is well aligned with previous results (Clem et al., 2021; Forsblom et al., 2021; Frenzel et al., 2007; Goetz & Frenzel, 2010; Peixoto et al., 2017). Path coefficients were not substantial across all waves of measurement, however. Similar to past studies (Clem et al., 2021; Forsblom et al., 2021), boredom did not predict academic self-concept, negating feedback loops offered by CVT. A novel finding, academic self-concept showed long-term effects on boredom development, predicting levels of all three forms of boredom at the end of our investigation (cf. Study 3). Unlike for adjacent waves of measurement, long-term predictions of boredom intensity levels by academic self-concept were positive. This shows that academic self-concept relates differently to boredom intensity, depending on the observed time frame. This might indicate a shift in students' academic boredom perceptions taking place during our investigation (see below). Future studies exploring the mathematics domain should compare short- and long-term effects of academic self-concept on boredom intensity systematically.

Academic interest negatively predicted levels of boredom intensity at the end of our investigation, also showing significant long-term effects (cf. Study 3). This result is well aligned with respective short-term predictions that were not investigated in this dissertation

(Goetz et al., 2020; Pekrun et al., 2010; Putwain et al., 2018). Academic interest also predicted the developmental trajectory of boredom due to underchallenge, yielding steeper increases in case of higher interest. On a speculative note, this could imply student disappointment when interest is high, potentially due to failed expectations in classroom instruction (e.g., outpacing curricula).

German. Mirroring results for the mathematics domain, academic self-concept negatively predicted boredom due to overchallenge, albeit not across every wave of measurement (cf. Study 1). Regarding potential feedback loops, no form of academic boredom could predict academic self-concept (cf. Clem et al., 2021). Also comparable to mathematics results, academic self-concept showed significant long-term predictions of challenge-related boredom levels (cf. Study 3). Predictions of boredom due to underchallenge were positive, predictions of boredom due to overchallenge were negative. Unlike for mathematics, boredom intensity levels in German were not significantly predicted by early academic self-concept. As associated research for the German domain is lacking, these are all novel findings. Future studies considering the verbal domain should further explore predictive properties of academic self-concept regarding boredom, especially considering challenge-related forms. Furthermore, cross-domain differences (e.g., mathematics vs. verbal) should be considered in future theorizing.

Academic interest showed long-term predictive effects regarding boredom intensity- as well as boredom due to underchallenge development (cf. Study 3). For both forms of academic boredom, final levels were predicted negatively while developmental slopes were predicted positively. This means that increased interest in German led to steeper increases in boredom intensity and boredom due to underchallenge. Found negative predictions are well aligned with previous works examining shorter time spans (e.g., Goetz et al., 2020). Associations with steeper boredom increases pose a novel finding that could be related to increased interest-based learning investment. While this investment protects against boredom intensity and boredom due to underchallenge, it may boost respective developmental increases, for instance, due to the instructional pace no longer matching the net learning rate. In the future, precise assumptions on developmental slopes for different forms of academic emotions should be incorporated into theory and investigated systematically.

Table 5.3

Selected Study Results Embedded into the CVT-Framework – Appraisal

	Mathematics						German					
	BO-I		BO-U		BO-O		BO-I		BO-U		BO-O	
Procedural Step	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory
Appraisal												
ASC (+)	↑↑ (|β| = .14)		↑↑ (|β| = .25)		↓↓ (|β| = .20)				↑↑↑ (|β| = .36)		↓↓ (|β| = .25)	
AI (+)	↓↓↓ (|β| = .43)			↑↑↑ (|β| = .32)			↓↓↓ (|β| = .34)	↑↑ (|β| = .23)	↓↓ (|β| = .19)	↑↑↑ (|β| = .32)		

Note. ↑↑ / ↓↓ = moderate positive / negative effect. ↑↑↑ / ↓↓↓ = large positive / negative effect.

Predictive effect sizes (i.e., β) evaluated according to Keith (2019): $\beta < .10$ = small, $\beta \leq .25$ = moderate, $\beta > .25$ = large.

Empty cells indicate that no significant results were obtained.

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept. AI = academic interest.

Emotion

Achievement emotions are at the center of CVT's process model and taxonomized in terms of valence, arousal, and object focus (Pekrun, 2019; Pekrun et al., 2023). Different subtypes (i.e., challenge-related forms) of achievement boredom are not specified in CVT but have been established in previous work (Acee et al., 2010; Daschmann et al., 2011; Westgate & Wilson, 2018). In Study 1, I therefore sought to validate a three-factor-model of academic boredom (3F-model), including boredom intensity (in reference to one-dimensional, generalized boredom concepts), boredom due to underchallenge, and boredom due to overchallenge. Subject-specific findings were as follows (see Table 5.4 for a summary).

Mathematics. The 3F-model showed superior factorial validity in comparison to a competing one factor model (1F-model). This findings underlines the existence of challenge-related forms of academic boredom (Acee et al., 2010; Daschmann et al., 2011) and opposes functional accounts viewing boredom as a unitary construct (Elpidorou, 2021). Providing additional convergent and discriminant validity evidence, I examined correlations of different forms of boredom with academic self-concept as a control appraisal antecedent (see also below). Academic self-concept was negatively correlated with both boredom intensity and boredom due to overchallenge, and positively correlated to boredom due to underchallenge. This is well aligned with related works (Goetz & Frenzel, 2010; Krannich et al., 2019; Preckel et al., 2010). Intercorrelations for different forms of boredom confirmed previous results, as well (Goetz & Frenzel, 2010; Krannich et al., 2019; Preckel et al., 2010), except for correlations between boredom due to under- and overchallenge, which only showed an expected negative correlation at the last wave of measurement. This could mark a process of emotional differentiation taking place over the course of our investigation period.

German. The 3F-model outperformed the 1F-model in terms of data approximation. This points to the existence of different forms of boredom for the verbal domain, too. As this dissertation is written, I know of no studies distinguishing different forms of boredom for verbal domains. To add to validity evidence for the 3F-model, different forms of boredom were correlated to academic self-concept and to each other. Similar to the mathematics domain, academic self-concept showed negative relations to boredom intensity and boredom due to overchallenge. This replicates previous findings regarding general boredom for verbal domains (Clem et al., 2021; Krannich et al., 2019). Correlations between academic self-concept and challenge-related forms of boredom constitute novel results. Unlike for mathematics, academic self-concept showed negative and positive correlations with boredom due to underchallenge. This subject-specific finding could indicate differential difficulty development for both subjects, with potentially lower initial difficulty in early secondary school German class. Providing intercorrelations for different forms of academic boredom in German is a novelty. Mirroring mathematics results, as well as prior domain-unspecific results (Acee et al., 2010), boredom intensity was positively correlated to both boredom due to under- and overchallenge. Like for mathematics, correlations for boredom due to under- and overchallenge shifted across time. This time, however, without changing in direction, yielding all-positive correlations. This could imply that aforementioned differentiation processes take place at different developmental stages for mathematics vs. German, with German potentially lagging behind.

Table 5.4

Selected Study Results Embedded into the CVT-Framework – Emotion

	Mathematics						German					
	BO-I		BO-U		BO-O		BO-I		BO-U		BO-O	
Procedural Step	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory
Emotion[a]												
BO-U (+)	↑↑↑ - 0 (ρ_{T1} = .80; ρ_{T4} = .09)						↑↑↑ - ↑↑↑ (ρ_{T1} = .83; ρ_{T4} = .45)					
BO-O (+)	↑↑↑ - ↑↑↑ (ρ_{T1} = .37; ρ_{T4} = .56)		↑ - ↓↓↓ (ρ_{T1} = .18; ρ_{T4} = -.35)				↑↑↑ - ↑↑↑ (ρ_{T1} = .44; ρ_{T4} = .46)		↑↑↑ - 0 (ρ_{T1} = .43; ρ_{T4} = .08)			

Note. ↑ = small positive effect. ↑↑↑ / ↓↓↓ = large positive / negative effect. 0 = nonsignificant effect.

Latent correlation effect sizes (i.e., ρ) evaluated according to Gignac & Szodorai (2016): $\rho \geq .15$ = small, $\rho \geq .25$ = moderate, $\rho \geq .35$ = large.

Empty cells indicate that no significant results were obtained.

[a] Inter-boredom-correlations, with effect sizes for beginning (i.e., T1) and end (i.e., T4) of the investigation.

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. GC = special classes for the gifted. RC = regular classes. T1, T4 = respective waves of measurement (T1: early 5th grade; T4: mid 8th grade).

Achievement

In CVT's framework, achievement appears at two different procedural steps – as part of the learning environment and as the final academic outcome (Pekrun, 2019). I used self-reported final scholar grades received in primary school as a predictor to boredom development in secondary school (cf. Study 3). Linking prior achievement feedback to future achievement emotions is representative of feedback loops implied by CVT. Subject-specific findings were as follows (see Table 5.5 for a summary).

Mathematics. Primary school final grades displayed long-term effects on boredom due to under- and overchallenge, both regarding their levels at the end of our investigation, and their developmental trajectories. Previous grade had a positive effect on boredom due to underchallenge, indicating that increased prior performance resulted in higher boredom levels and steeper increases. Conversely, previous grade negatively predicted levels and developmental trajectories of boredom due to overchallenge, indicating less boredom for higher prior performance. These findings, linking challenge-related forms of boredom to academic achievement are novel, yet plausible. Achievement-boredom-relations should be studied further, regarding different forms of academic boredom. Previous grade did not impact boredom intensity levels or development, which contrasts previous findings for the mathematics domain (Forsblom et al., 2021; Lichtenfeld et al., 2022; Pekrun et al., 2017).

German. Primary school final grades predicted boredom due to overchallenge levels and developmental slopes. Higher performance was associated with lower boredom due to overchallenge levels. Unlike for mathematics, higher performance also led to steeper increases in boredom due to overchallenge. Higher prior achievement hence exacerbated boredom due to overchallenge, which could be due to less predictability in the difficulty development for language classes, compared to more cumulative mathematics classes. Future studies should address subject-specific task difficulty development in search of a clearer understanding for achievement-boredom-relations, regarding different forms and subject domains.

Table 5.5

Selected Study Results Embedded into the CVT-Framework – Achievement

	Mathematics						German					
	BO-I		BO-U		BO-O		BO-I		BO-U		BO-O	
Procedural Step	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory	*M*(-Level)	Trajectory
Achievement												
Previous Grade (+)			↑↑ ($\|\beta\| = .16$)	↑↑ ($\|\beta\| = .16$)	↓↓ ($\|\beta\| = .24$)	↓↓ ($\|\beta\| = .21$)					↓↓ ($\|\beta\| = .10$)	↑↑ ($\|\beta\| = .14$)

Note. ↑↑ / ↓↓ = moderate positive / negative effect.

Predictive effect sizes (i.e., β) evaluated according to Keith (2019): $\beta < .10$ = small, $\beta \leq .25$ = moderate, $\beta > .25$ = large.

Empty cells indicate that no significant results were obtained.

BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. Previous grade coded inversely for higher values to imply higher achievement.

Future Directions

To limit repetition, this section does not contain limitations of individual studies (cf. discussion sections of chapters 2-4). Instead, I acknowledge conceptual and methodological alternatives to the way this dissertation was built. Consequently, the following paragraphs provide a scientific outlook and emphasise future research efforts building on the present dissertation.

Academic Boredom Measurement

AVG-project data used in this dissertation date back as far as 2005. Self-report measures of boredom have evolved (for an overview, see Vodanovich & Watt, 2016). Back when the AVG-project launched, the most prominent boredom measure was the *Achievement Emotions Questionnaire – Mathematics* (AEQ-M; Pekrun et al., 2005). I used two AEQ-M-items to measure boredom intensity in this dissertation. Still the most widely used academic boredom measure, the *Achievement Emotions Questionnaire* (AEQ; Pekrun et al., 2011) has since been refined and adapted for different age groups (e.g., for elementary school, Lichtenfeld et al., 2012; for preadolescent students, Peixoto et al., 2015). Using larger portions of the AEQ could have provided a more reliable and comprehensive general boredom factor, due to its multi-component approach, covering affective, cognitive, motivational, and physiological aspects (cf. Scherer, 2009). This was not intended, however, in the AVG-project's conception. Only the 2 AEQ-items used were available for anlayses.

Measures for different forms of academic boredom have also advanced since the early 2000s. The development of the *Academic Boredom Scale* (ABS; Acee et al., 2010) and the *Precursors to Boredom Scales* (PBS; Daschmann et al., 2011) brought about more differentiated boredom measures, acknowledging boredom due to under- and overchallenge as separate forms. For challenge-related forms of boredom, ABS- and PBS-measurement provide viable alternatives to self-developed items used in this dissertation. Especially the

PBS seems promising, as it combined generalized boredom (cf. boredom intensity), aspects of under- and overchallenge (cf. boredom due to under- and overchallenge), and characteristics of the learning environment (e.g., teacher dislike, monotony) into a comprehensive eight-factor-model. Furthermore, the PBS has been externally validated (Tze et al., 2014) and even translated into Chinese (Chen et al., 2021). However, PBS-measurement consists of 22 items, making it a rather long scale to be used for repeated assessments spanning several years.

Besides questionnaire measures, there have been increased calls for more objective boredom assessments (Goetz et al., 2019). Using physiological markers (e.g., heart rates, body temperature, electro-encephalogram) and observational methods (e.g., eye-tracking, or the *Facial Action Coding System*, Ekman et al., 2002) would increase measurement adequacy for physiological and expressive components. Nevertheless, cognitive, affective, and motivational components need measurement approaches reverting to introspection. Future works should provide multi-component measurements of boredom (as well as other achievement emotions) and explore the equivalence of self-report to objective data. More objective emotion measurement also implies an increased used of laboratory settings and experimental procedures in educational psychology. While this certainly strengthens interpretations and limits alternative explanations to obtained findings, randomization and direct manipulation are seldom possible in authentic field settings. Still, moving to the laboratory more often forms a desiderate to enhance the internal validity of educational research designs.

Widening the Scope

All three studies from this dissertation analysed the same project data – namely data from the AVG-project – in singular, empirical works. Therefore, data-related limitations apply to all three studies likewise (see also limitation sections of chapters 2-4). Beyond this, available project data contained a lot more information that could have been used in empirical examinations of academic boredom. For example, different types of achievement goals (i.e., approach and avoidance achievement motivation, competence motivation, see Elliot & Church, 1997), forming distal individual antecedents in CVT's framework, were also assessed in the AVG-project (for a meta-analysis, see Huang, 2011). In grades 6 and 8, classroom climate (Eder & Mayr, 2000) was included as an important environmental variable. Most crucially perhaps, relevant achievement outcomes were mostly neglected in this dissertation (for meta-analyses, see Camacho-Morles et al., 2021; Tze et al., 2016). Self-reported scholar grades from grade 5 on could have been incorporated into individual studies. Furthermore, this dissertation focused on early secondary school (i.e., T1-T4), while AVG-project data spanned a total of 15 years of data assessments (i.e., T1-T8), including two post-graduate waves of measurement (i.e., T7-8).[7] While academic boredom itself was only assessed in early secondary school, long-term effects of boredom predicting academic performance or choices in the distant future would have been an interesting avenue to explore empirically. Besides possible empirical expansions within AVG-project data, including integrative, nonempirical approaches, like systematic reviews or meta-analyses could have increased comprehensiveness of attained results. In a recent integrative work, for instance, Loderer et al. (2020) found similar functional mechanisms regarding achievement emotions for technology-based learning environments and classic face-to-face instruction – an interesting perspective, considering predominant field research in educational settings.

Many ideas listed here fall beyond the scope of the present dissertation as it stands. Nevertheless, I encourage other scholars to look into these matters in the future.

[7] An initial wave of measurement right after transitioning to 5^{th} grade (T0), with questionnaire assessments referring to previous experience in primary school was also part of the AVG-project (see also chapter 4). This initial wave, however, featured altered item wording for many repeated assessments.

Analytical Alternatives

This dissertation's process included fair amounts of methodological literature study. During this process, I came across various approaches to modelling longitudinal data and extract sound interpretations. I did explore many of these alternatives in search of the best method for my precise analytical purposes. Listed here are analytical directions, I did not follow in conduction of this dissertation.

First, I established the 3F-model of boredom by means of confirmatory factor analyses, comparing it to alternative one-factor-model in representation of unitary boredom conceptualizations (cf. Study 1). However, other alternative models, although lacking theoretical foundations, could also have been tested against (e.g., higher-order- or nested factor models, see Brunner et al., 2012). Furthermore, instrument construction protocols (Moosbrugger & Kelava, 2012) recommend exploratory factor analyses to precede confirmatory approaches.[8] However, boredom scales used in this dissertation were designed to be short-scales to begin with. Precisely following procedural steps of questionnaire construction would have required a larger initial item pool.

Second, I tested for measurement invariance across comparison groups using multi-group-comparison models (MGC; cf. Byrne, 2012; French & Finch, 2008; Kleinke et al., 2017) (cf. Studies 2 and 3). These same MGC models could also have been used to test differential boredom development for regular vs. special gifted classes (Study 2) or girls vs. boys (Study3). Reverting to MGC models for these analyses would have resulted in separate latent growth curves for comparison groups that could be compared directly. Using multiple-indicator-multiple-cause models (MIMIC; Muthén, 1989) instead – like I did – bears the

[8] I supervised two Bachelor's theses, testing the factorial structure of AVG-project boredom scales by exploratory factor analyses (Brucker, 2022; Jörgensen, 2022). Results of parallel analyses concerning both subject-domains indicated a three-factor-structure.

advantage of attaining standardized predictive effects sizes associated with respective group membership. This enhances cross-study comparability for two group-designs.

Third, methodological approaches to longitudinal data analysis are manifold (see Usami et al., 2019 for an overview). Especially with respect to reciprocal effects models, recent developments in the methods literature have emerged. Adding random intercepts to extract trait variance in studies of reciprocal causation is argued as the more viable alternative to classic REM-approaches (Hamaker et al., 2015; Lucas, 2022). Alternatively, trait-state-occasion models (TSO; Cole et al., 2005; Eid et al., 2017) are often used to examine intensive longitudinal data. While possible to run for any set of longitudinal data, such models usually require many measurement occasions spread over shorter time periods for optimal functioning. This is not the case for the AVG-project data.

Lastly, all studies conducted in this dissertation, featured nonequidistant waves of measurement. While this can be compensated for by choosing apt latent slope factor loadings in latent growth curve modelling approaches (LGCM; Meredith & Tisak, 1990) (cf. Studies 2 and 3), this limits generalizability for REMs (cf. Study 1). This issue can be addressed using continuous time modelling (Oud, 2007; Voelkle et al., 2012). Still, interpretations of results can be linked to different grade levels (instead of age, as would be the case if continuous time were established). This is an adequate solution for educational settings, as relevant developmental processes here often revert to different grade levels, rather than precise subject age.

Practical Implications

This dissertational project stresses the need for direct boredom intervention at school (Feuchter & Preckel, 2022b). That is why, in the following, I will propose avenues to boredom intervention, linked to the results of this dissertational project. For practical implications of individual studies, please refer to discussion sections of chapters 2-4.

Promoting awareness of boredom's high prevalence and detrimental consequences at school is the first step in boredom intervention (Macklem, 2015). To achieve this, educational research should be made available not only to other researchers but to studied populations (e.g., secondary school students), practitioners (e.g., teachers) and policy makers (e.g., ministries). To ensure proper reception, academic results need to be communicated adequately, breaking down findings in ways that relate to everyday scholar processes. After boredom is addressed as an educational problem, educating teachers and students about the intricacies of boredom and its development across secondary school is vital. Not just regarding boredom – but emotional development in all important achievement emotions should be taught at school. To minimize interference with existing scholar curricula, a subject-specific teaching approach could prove helpful. As achievement emotions are organized subject-specifically (Goetz et al., 2007), a limited amount of time should be allocated to teaching emotional development in every major scholar subject. That way, awareness and basic education regarding emotional processes at school could be achieved without the need for excessive reforms of existing teaching structures and routines (Gordon Biddle, 2021).

Returning to boredom intervention, control and value perceptions in class should be tackled specifically. Avoiding inadequate challenge, for instance, could reduce experiences of boredom in class drastically. This is especially difficult to achieve, as perceptions of challenge are highly subjective. Finding objective markers of task difficulty in class could provide a first step to solving this puzzle (e.g., by continuous student ratings of classroom-based challenge). Teachers should understand their own part in classroom-based difficulty regulation. Identifying singular aspects of tasks, rendering them more or less difficult, is essential to establish a comprehensible and transparent difficulty structure. Still, curricular constraints remain. Of course, not every student will ace a given subject just because they understand where its difficulty lies. Individual competence and learning strategies play an immense part in academic achievement, as well (Pekrun, 2017). That is why, increased intra-individual feedback should be combined with a relatable class difficulty architecture. For students to climb up the difficulty ranks progressively could prove highly motivating, regardless of relative performance by classroom standards (Hughes, 2014). Practitioners need to be aware, however, that challenge perceptions are subject to qualitative changes in early secondary school, making challenge-based interventions most practical as of grade 8 (cf. Study 1).

How can subject value be increased in class? A favorable way to achieve this may be to cater more towards student interests. Like challenge perceptions, student interests are like to differ substantially between individuals, and across time. Moreover, curricular responsibilities on the teachers' part limit room for changes in content. Before any interest intervention can be installed, however, teachers need to inquire about student interests in order to find approachable angles to adapting class content. Overlapping interests, representing large portions of the classroom, could be useful in reframing task content. By means of this rather superficial intervention, curricular directions can be followed while also making tasks more appealing to students. Besides interest interventions, real-life task value can be promoted further during class (Berweger et al., 2022; Gaspard et al., 2015). Communicating the importance of individual skills required in different subjects, beyond their necessity for graduation, is key. This could be achieved together with field experts promoting their line of work in selected lessons, while emphasizing required skills attained in different scholar subjects. Linking individual tasks to student interests while providing more life-like task framing could help to make classes more interesting.

Even if boredom intervention is executed well, boredom at school may still prevail. Developing adequate coping strategies is therefore of major importance to evade long-term increases in academic boredom, as well as its negative consequences in scholar and private life. Training students to see meaning in potentially boring tasks appears to be an adaptive way of reducing boredom at school (Nett et al., 2010). Such situational reappraisals can be made more accessible to students in combination with abovementioned challenge and interest interventions. Preferably, students are aware of the underlying value of tasks they perform at school (or also at home, for instance in preparation for the next lesson) at all times. Therefore, establishing a consensus regarding meaningfulness of scholar tasks forms the base for effective boredom relief. This may also include discarding tasks for which utility, interest, or meaningfulness cannot be readily argued.

Summary of Conclusions

To finalize the present dissertation, I will now highlight its most important takeaways.

First, academic boredom is complex. Findings from this dissertation empirically separated intensity aspects from challenge-related aspects of boredom experience. All three forms investigated here showed similarities (e.g., long-term increases) and differences (e.g., regarding their interrelations) in their development over time. The present dissertation is also a testament to domain-specific educational research. Emotional and motivational processes taking place in scholar environments are usually embedded in classroom contexts. Only when they are addressed subject-specifically, emotional processes can be understood in detailed fashion.

Second, boredom at school requires tailored intervention. Increasing developmental trajectories found in the present dissertation underline this. Full-time ability grouping – as an intervention concept addressed in this work – has been shown to be less effective than expected when it comes to reducing boredom. This also indicates a disconnect in intellectual giftedness and boredom experience in secondary education. Enhancing interestingness of scholar tasks while empowering students to learn in according to their individual capabilities could prove more effective than dividing the classroom.

Third, academic boredom is highly subjective and shows considerable individual differences. Regarding gender, for instance, boredom development differed beyond differences in proximal control and value appraisal, as well as previous achievement. Aside from such quantitative differences, an increased understanding of the qualitative differences in individual boredom experience could help fill existing gaps in the conceptual understanding of this immensely enigmatic emotion.

References

Acee, T. W., Kim, H., Kim, H. J., Kim, J.-I., Chu, H.-N. R., Kim, M., Cho, Y., & Wicker, F. W. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, *35*(1), 17–27. https://doi.org/10.1016/j.cedpsych.2009.08.002

Atkinson, J. W. (1964). *An introduction to motivation*. Van Nostrand.

Baber, N. (1994). International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH). *British Journal of Clinical Pharmacology*, *37*(5), 401–404. https://doi.org/10.1111/j.1365-2125.1994.tb05705.x

Baker, S. N. (2016). Uninspired or disengaged? A phenomenological investigation of gifted middle school students on probation [Ed.D., Liberty University]. In *ProQuest Dissertations and Theses* (1797944984). https://www.proquest.com/dissertations-theses/uninspired-disengaged-phenomenological/docview/1797944984/se-2?accountid=14652

Bar-On, R. (2007). The Impact of Emotional Intelligence on Giftedness. *Gifted Education International*, *23*(2), 122–137. https://doi.org/10.1177/026142940702300203

Becker, M., Neumann, M., Tetzner, J., Böse, S., Knoppick, H., Maaz, K., Baumert, J., & Lehmann, R. (2014). Is early ability grouping good for high-achieving students' psychosocial development? Effects of the transition into academically selective schools. *Journal of Educational Psychology*, *106*(2), 555–568. https://doi.org/10.1037/a0035425

Bench, S., & Lench, H. (2013). On the Function of Boredom. *Behavioral Sciences*, *3*(3), 459–472. https://doi.org/10.3390/bs3030459

Berweger, B., Born, S., & Dietrich, J. (2022). Expectancy-value appraisals and achievement emotions in an online learning environment: Within- and between-person relationships. *Learning and Instruction*, *77*, 101546. https://doi.org/10.1016/j.learninstruc.2021.101546

Bieg, S., Grassinger, R., & Dresel, M. (2019). Teacher humor: Longitudinal effects on students' emotions. *European Journal of Psychology of Education*, *34*(3), 517–534. https://doi.org/10.1007/s10212-018-0402-0

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The Role of Coding Time in Estimating and Interpreting Growth Curve Models. *Psychological Methods*, *9*(1), 30–52. https://doi.org/10.1037/1082-989X.9.1.30

Boehme, K. L., Goetz, T., & Preckel, F. (2017). Is it good to value math? Investigating mothers' impact on their children's test anxiety based on control-value theory. *Contemporary Educational Psychology*, *51*, 11–21. https://doi.org/10.1016/j.cedpsych.2017.05.002

Bollen, K. A. (2007). On the Origins of Latent Curve Models. In R. Cudeck & R. C. MacCallum (Eds.), *Factor Analysis at 100: Historical Developments and Future Directions* (pp. 79–98). Lawrence Erlbaum Associates Publishers.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press. https://www.guilford.com/books/Confirmatory -Factor-Analysis-for-Applied-Research/Timothy-Brown/9781462515363

Brucker, S. (2022). *Eine itemanalytische Untersuchung akademischer Langeweile im Mathematikunterricht* [Unpublished bachelor's thesis]. University of Trier.

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

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A Tutorial on Hierarchically Structured Constructs: Hierarchically Structured Constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Butković, A., & Bratko, D. (2003). Generation and Sex Differences in Sensation Seeking: Results of the Family Study. *Perceptual and Motor Skills*, *97*(3), 965–970. https://doi.org/10.2466/pms.2003.97.3.965

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming.* Routledge/Taylor & Francis Group. https://www.routledge.com/Structural-Equation-Modeling-with-Mplus-Basic - Concepts-Applications-and/Byrne/p/book/9781848728394

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity Achievement Emotions and Academic Performance: A Meta-analysis. *Educational Psychology Review*, *33*(3), 1051–1095. https://doi.org/10.1007/s10648-020-09585-3

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, J., Zhou, L., & Liu, L. (2021). Chinese version of the Precursors to Boredom Scales: Evaluation and psychometric properties. *Social Behavior and Personality: An International Journal*, *49*(9), 1–10. https://doi.org/doi:10.2224/sbp.10577

Clem, A.-L., Hirvonen, R., Aunola, K., & Kiuru, N. (2021). Reciprocal relations between adolescents' self-concepts of ability and achievement emotions in mathematics and literacy. *Contemporary Educational Psychology*, *65*, 101964. https://doi.org/10.1016/j.cedpsych.2021.101964

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local Regression Models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical Models in S*. Routledge.

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

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and Conceptual Problems With Longitudinal Trait-State Models: Introducing a Trait-State-Occasion Model. *Psychological Methods*, *10*(1), 3. https://doi.org/10.1037/1082-989X.10.1.3

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: The experience of play in work and leisure*. Jossey-Bass.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math-Gender Stereotypes in Elementary School Children: Gender Stereotypes. *Child Development, 82*(3), 766–779. https://doi.org/10.1111/j.1467-8624.2010.01529.x

Czikmantori, T., Hennecke, M., & Brandstätter, V. (2021). Task Enjoyment as an Individual Difference Construct. *Journal of Personality Assessment*, *103*(6), 818–832. https://doi.org/10.1080/00223891.2021.1882473

Daschmann, E. C., Goetz, T., & Stupnisky, R. H. (2011). Testing the predictors of boredom at school: Development and validation of the precursors to boredom scales. *British Journal of Educational Psychology*, *81*(3), 421–440. https://doi.org/10.1348/000709910X526038

Daschmann, E. C., Goetz, T., & Stupnisky, R. H. (2014). Exploring the antecedents of boredom: Do teachers know why students are bored? *Teaching and Teacher Education*, *39*, 22–30. https://doi.org/10.1016/j.tate.2013.11.009

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The Unengaged Mind: Defining Boredom in Terms of Attention. *Perspectives on Psychological Science*, *7*(5), 482–495. https://doi.org/10.1177/1745691612456044

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *2*(48), 90–101. https://doi.org/10.1037/0003-066X.48.2.90

Eder, F., & Mayr, J. (2000). *Linzer Fragebogen zum Schul- und Klassenklima für die 4. Bis 8. Klassenstufe* (1st ed.). Hogrefe.

Eid, M., Holtmann, J., Santangelo, P., & Ebner-Priemer, U. (2017). On the Definition of Latent-State-Trait Models With Autoregressive Effects: Insights From LST-R Theory. *European Journal of Psychological Assessment*, *33*(4), 285–295. https://doi.org/10.1027/1015-5759/a000435

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial Action Coding System*. Research Nexus.

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218–232. https://doi.org/10.1037/0022-3514.72.1.218

Elpidorou, A. (2014). The bright side of boredom. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01245

Elpidorou, A. (2017). Boredom in art. *Behavioral and Brain Sciences*, *40*, e359. Cambridge Core. https://doi.org/10.1017/S0140525X17001674

Elpidorou, A. (2018). The bored mind is a guiding mind: Toward a regulatory theory of boredom. *Phenomenology and the Cognitive Sciences*, *17*(3), 455–484. https://doi.org/10.1007/s11097-017-9515-1

Elpidorou, A. (2021). Is boredom one or many? A functional solution to the problem of heterogeneity. *Mind & Language*, *36*(3), 491–511. https://doi.org/10.1111/mila.12282

Ernst, E., & Pittler, M. H. (2006). Efficacy or effectiveness? *Journal of Internal Medicine*, *260*(5), 488–490. https://doi.org/10.1111/j.1365-2796.2006.01707.x

Fan, X., & Nowell, D. L. (2011). Using Propensity Score Matching in Educational Research. *Gifted Child Quarterly*, *55*(1), 74–79. https://doi.org/10.1177/0016986210390635

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond Effect on Academic Self-Concept: A Meta-Analysis. *Frontiers in Psychology*, *9*, 1569. https://doi.org/10.3389/fpsyg.2018.01569

Farmer, R., & Sundberg, N. D. (1986). Boredom Proneness–The Development and Correlates of a New Scale. *Journal of Personality Assessment*, *50*(1), 4–17. https://doi.org/10.1207/s15327752jpa5001_2

Feldhusen, J. F., & Kroll, M. D. (1991). Boredom or Challenge for the Academically Talented in School. *Gifted Education International*, *7*(2), 80–81. https://doi.org/10.1177/026142949100700207

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, *15*(5), 451–474. https://doi.org/10.1016/0091-7435(86)90024-1

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., Mościcki, E. K., Schinke, S., Valentine, J. C., & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*(3), 151–175. https://doi.org/10.1007/s11121-005-5553-y

Forsblom, L., Pekrun, R., Loderer, K., & Peixoto, F. (2021). Cognitive appraisals, achievement emotions, and students' math achievement: A longitudinal analysis. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000671

French, B. F., & Finch, W. H. (2008). Multigroup Confirmatory Factor Analysis: Locating the Invariant Referent Sets. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 96–113. https://doi.org/10.1080/10705510701758349

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*(4), 497–514. https://doi.org/10.1007/BF03173468

Frenzel, A. C., & Stephens, E. J. (2013). Emotions. In N. C. Hall & T. Goetz (Eds.), *Emotion, motivation, and self-regulation: A handbook for teachers* (First edition, pp. 1–56). Emerald.

Freund, V. A., Schulenberg, J. E., & Maslowsky, J. (2021). Boredom by Sensation-Seeking Interactions During Adolescence: Associations with Substance Use, Externalizing Behavior, and Internalizing Symptoms in a US National Sample. *Prevention Science*, *22*(5), 555–566. https://doi.org/10.1007/s11121-020-01198-0

Feuchter, M. D., & Preckel, F. (2020, December 11). *Can Ability Grouping Reduce Academic Boredom? – An Evaluation of Intervention Effectiveness and Efficacy* [Poster presentation]. 2. Interdisziplinäre Graduiertenkonferenz [Online], Trier.

Feuchter, M. D., & Preckel, F. (2021a, August 18-20). *Evaluation Ability Grouping's Potential to Reduce Gifted Students' Academic Boredom.* [Paper presentation]. JURE 2021 [Online], Gothenburg.

Feuchter, M. D., & Preckel, F. (2021b, September 14-16). *Akademische Langeweile und Hochbegabung – Evaluation von Förderklasseneffekten* [Presentation as part of a symposium]. PAEPSY 2021 [Online], Heidelberg.

Feuchter, M. D., & Preckel, F. (2021c, December 3). *Geschlechterunterschiede im schulischen Langeweileerleben aus Sicht der Kontroll-Wert-Theorie.* [Paper presentation]. 3. Interdisziplinäre Graduiertenkonferenz [Online], Trier.

Feuchter, M. D., & Preckel, F. (2022a, April 21). *Can Gender Differences in Academic Boredom Development Be Explained by Assumptions of Control-Value Theory?* [Paper presentation]. AERA Annual Meeting 2022, San Diego, CA.

Feuchter, M. D., & Preckel, F. (2022b). Reducing boredom in gifted education—Evaluating the effects of full-time ability grouping. *Journal of Educational Psychology*, *114*(6), 1477–1493. https://doi.org/10.1037/edu0000694

Feuchter, M. D., & Preckel, F. (2022c, September 12). *Gender Differences in Academic Boredom Trajectories in Secondary School* [Poster presentation]. 52. DGPs-Kongress, Hildesheim.

Feuchter, M. D., & Preckel, F. (2023, April 13-16). *The Dynamics of Different Forms of Academic Boredom in Secondary School* [Poster presentation]. AERA Annual Meeting 2023, Chicago, IL.

Gana, K., Broc, G., & Bailly, N. (2019). Does the Boredom Proneness Scale capture traitness of boredom? Results from a six-year longitudinal trait-state-occasion model. *Personality and Individual Differences*, *139*, 247–253. https://doi.org/10.1016/j.paid.2018.11.030

Gaspard, H., Dicke, A.-L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, *51*, 1226–1240. https://doi.org/10.1037/dev0000028

Geiser, C. (2011). *Datenanalyse mit Mplus*. Springer. https://doi.org/10 .1007/978-3-531-93192-0

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gjesme, T. (1977). General Satisfaction and Boredom at School as a Function of the Pupils' Personality Characteristics. *Scandinavian Journal of Educational Research*, *21*(1), 113–146. https://doi.org/10.1080/0031383770210106

Goetz, T. (2004). *Emotionales Erleben und selbstreguliertes Lernen bei Schülern im Fach Mathematik* [Students' emotions and self-regulated learning in mathematics], [Unpublished doctoral dissertation]. University of Munich. https://www.utzverlag.de/assets/pdf/40429les.pdf

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013a). Do Girls Really Experience More Anxiety in Mathematics? *Psychological Science*, *24*(10), 2079–2087. https://doi.org/10.1177/0956797613486989

Goetz, T., & Frenzel, A. C. (2006). Phänomenologie schulischer Langeweile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *38*(4), 149–153. https://doi.org/10.1026/0049-8637.38.4.149

Goetz, T., & Frenzel, A. C. (2010). Über- und Unterforderungslangeweile im Mathematikunterricht. *Empirische Pädagogik*, *24*(2), 113–134.

Goetz, T., Frenzel, A. C., Hall, N. C., Nett, U. E., Pekrun, R., & Lipnevich, A. A. (2014). Types of boredom: An experience sampling approach. *Motivation and Emotion*, *38*(3), 401–419. https://doi.org/10.1007/s11031-013-9385-y

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

Goetz, T., Frenzel, A. C., Lüdtke, O., & Hall, N. C. (2010). Between-Domain Relations of Academic Emotions: Does Having the Same Instructor Make a Difference? *The Journal of Experimental Education*, *79*(1), 84–101. https://doi.org/10.1080/00220970903292967

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

Goetz, T., Hall, N. C., & Krannich, M. (2019). Boredom. In K. A. Renninger & S. E. Hidi (Eds.), *The Cambridge Handbook of Motivation and Learning* (pp. 465–489). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781316823279.021

Goetz, T., Keller, M. M., Lüdtke, O., Nett, U. E., & Lipnevich, A. A. (2020). The dynamics of real-time classroom emotions: Appraisals mediate the relation between students' perceptions of teaching and their emotions. *Journal of Educational Psychology*, *112*(6), 1243–1260. https://doi.org/10.1037/edu0000415

Goetz, T., Lüdtke, O., Nett, U. E., Keller, M. M., & Lipnevich, A. A. (2013b). Characteristics of teaching and students' emotions in the classroom: Investigating differences across domains. *Contemporary Educational Psychology*, *38*(4), 383–394. https://doi.org/10.1016/j.cedpsych.2013.08.001

Goetz, T., & Nett, U. (2012). *Boredom in university students*. Codebook, University of Konstanz.

Goetz, T., Nett, U. E., Martiny, S. E., Hall, N. C., Pekrun, R., Dettmers, S., & Trautwein, U. (2012). Students' emotions during homework: Structures, self-concept antecedents, and achievement outcomes. *Learning and Individual Differences*, *22*(2), 225–234. https://doi.org/10.1016/j.lindif.2011.04.006

Goetz, T., Pekrun, R., Hall, N., & Haag, L. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction. *British Journal of Educational Psychology*, *76*(2), 289–308. https://doi.org/10.1348/000709905X42860

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

Gordon Biddle, K. A. (2021). Social and Emotional Learning. In A. Garcia-Nevarez & K. A. Gordon Biddle (Eds.), *Developmentally Appropriate Curriculum and Instruction*. Routledge. https://doi.org/10.4324/9780429353147

Haag, L., & Goetz, T. (2012). Mathe ist schwierig und Deutsch aktuell: Vergleichende Studie zur Charakterisierung von Schulfächern aus Schülersicht. *Psychologie in Erziehung Und Unterricht*, *59*(1), 32–46. https://doi.org/10.2378/peu2012.art03

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. https://doi.org/10.1037/a0038889

Hannover, B., & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction, 14*(1), 51–67. https://doi.org/10.1016/j.learninstruc.2003.10.002

Hannover, B., Koeppen, K., & Kreutzmann, M. (2022). Die Qualität des Lehrkraftverhaltens in Lehrkraft-Kind-Dyaden: Unterschiede zwischen Kindern in Abhängigkeit ihrer Noten, ihres Geschlechts, ihrer sprachlichen Herkunft und eines sonderpädagogischen Förderbedarfs. *Zeitschrift für Pädagogische Psychologie*, *36*(1–2), 65–84. https://doi.org/10.1024/1010-0652/a000327

Harasymchuk, C., & Fehr, B. (2012). Development of a prototype-based measure of relational boredom. *Personal Relationships, 19*(1), 162–181. https://doi.org/10.1111/j.1475-6811.2011.01346.x

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. Bis 12. Klassen* [Cognitive abilities test of students from 4th to 12th grade]. Hogrefe.

Herrmann, J., Schmidt, I., Kessels, U., & Preckel, F. (2016). Big fish in big ponds: Contrast and assimilation effects on math and verbal self-concepts of students in within-school gifted tracks. *British Journal of Educational Psychology*, *86*(2), 222–240. https://doi.org/10.1111/bjep.12100

Hill, A. B., & Perkins, R. E. (1985). Towards a model of boredom. *British Journal of Psychology*, *76*(2), 235–240. https://doi.org/10.1111/j.2044-8295.1985.tb01947.x

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). *The MatchIt package* (3.0.2) [Computer software]. http://gking.harvard.edu/matchit

Hornstra, L., van der Veen, I., & Peetsma, T. (2017). Effects of full-time and part-time high-ability programs on developments in students' achievement emotions. *High Ability Studies*, *28*(2), 199–224. https://doi.org/10.1080/13598139.2017.1332575

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, C. (2011). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review*, *23*(3), 359–388. https://doi.org/10.1007/s10648-011-9155-x

Hughes, G. (2014). *Ipsative Assessment: Motivation through Marking Progress*. Springer.

Hunter, J. A., & Eastwood, J. D. (2021). Understanding the relation between boredom and academic performance in postsecondary students. *Journal of Educational Psychology*, *113*(3), 499–515. https://doi.org/10.1037/edu0000479

IBM Corporation. (2022). *IBM SPSS Statistics* (Version 29). IBM Corporation.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review*, *99*(3), 561–565. https://doi.org/10.1037/0033-295X.99.3.561

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*(2), 239–251. https://doi.org/10.1093/biomet/57.2.239

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. https://doi.org/10.1007/BF02291366

Jörgensen, I. (2022). *Akademische Langeweile im Fach Deutsch—Eine item- & faktoranalytische Untersuchung* [Unpublished bachelor's thesis]. University of Trier.

Keith, T. Z. (2006). *Multiple Regression and Beyond*. Pearson Education. https://www.pearson.com/uk/educators/higher-education-educators/ program/Keith-Multiple-Regression-and-Beyond/PGM489595.html

Keith, T. Z. (2019). *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling* (3rd ed.). Routledge/Taylor & Francis Group. https://www.routledge.com/Multiple-Regression-and-Beyond-An-Introduction-to-Multiple-Regression-and/Keith/p/book/9781138061446

Kellam, S. G., & Langevin, D. J. (2003). A framework for understanding “evidence” in prevention research and programs. *Prevention Science*, *4*(3), 137–153. https://doi.org/10.1023/A:1024693321963

Keller, L., Preckel, F., Eccles, J. S., & Brunner, M. (2021). Top-performing math students in 82 countries: An integrative data analysis of gender differences in achievement, achievement profiles, and achievement motivation. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000685

Kleinke, K., Schlüter, E., & Christ, O. (2017). *Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung* (2nd ed.). Walter de Gruyter GmbH. https://doi.org/10.1515/9783486989458

Košir, K., Horvat, M., Aram, U., & Jurinec, N. (2016). Is being gifted always an advantage? Peer relations and self-concept of gifted students. *High Ability Studies*, *27*(2), 129–148. https://doi.org/10.1080/13598139.2015.1108186

Krannich, M., Goetz, T., Lipnevich, A. A., Bieg, M., Roos, A.-L., Becker, E. S., & Morger, V. (2019). Being over- or underchallenged in class: Effects on students’ career aspirations via academic self-concept and boredom. *Learning and Individual Differences*, *69*, 206–218. https://doi.org/10.1016/j.lindif.2018.10.004

Larson, R. W., & Richards, M. H. (1991). Boredom in the Middle School Years: Blaming Schools versus Blaming Students. *American Journal of Education*, *99*(4), 418–443. https://doi.org/10.1086/443992

LePera, N. (2011). Relationships between boredom proneness, mindfulness, anxiety, depression, and substance use. *The New School Psychology Bulletin*, *8*(2), 15–25.

Lichtenfeld, S., Pekrun, R., Marsh, H. W., Nett, U. E., & Reiss, K. (2022). Achievement emotions and elementary school children's academic performance: Longitudinal models of developmental ordering. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000748

Lichtenfeld, S., Pekrun, R., Stupnisky, R. H., Reiss, K., & Murayama, K. (2012). Measuring students' emotions in the early years: The Achievement Emotions Questionnaire-Elementary School (AEQ-ES). *Learning and Individual Differences*, *22*(2), 190–201. https://doi.org/10.1016/j.lindif.2011.04.009

Litster, K., & Roberts, J. (2011). The self-concepts and perceived competencies of gifted and non-gifted students: A meta-analysis. *Journal of Research in Special Educational Needs*, *11*(2), 130–140. https://doi.org/10.1111/j.1471-3802.2010.01166.x

Little, C. A. (2012). Curriculum as motivation for gifted students. *Psychology in the Schools*, *49*(7), 695–705. https://doi.org/10.1002/pits.21621

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading*, *32*(2), 199–214. https://doi.org/10.1111/j.1467-9817.2008.01389.x

Lucas, R. E. (2022). *It's Time To Abandon The Cross-Lagged Panel Model* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/pkec7

Macklem, G. L. (2015). *Boredom in the Classroom: Addressing Student Motivation, Self-Regulation, and Engagement in Learning*. Springer.

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology*, *47*, 166–173. https://doi.org/10.1037/h0062855

Marsh, H. W. (1990). *Self Description Questionnaire II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: A test manual and research monograph.* Psychological Corporation.

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

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods*, *9*(4), 403–425. https://doi.org/10.1037/1082-989X.9.4.403

McDonald, R. P. (1999). *Test theory: A unified treatment.* Lawrence Erlbaum Associates Publishers.

Melton, A. M. A., & Schulenberg, S. E. (2009). A Confirmatory Factor Analysis of the Boredom Proneness Scale. *The Journal of Psychology*, *143*(5), 493–508. https://doi.org/10.3200/JRL.143.5.493-508

Mercer-Lynn, K. B., Bar, R. J., & Eastwood, J. D. (2014). Causes of boredom: The person, the situation, or both? *Personality and Individual Differences*, *56*, 122–126. https://doi.org/10.1016/j.paid.2013.08.034

Mercer-Lynn, K. B., Flora, D. B., Fahlman, S. A., & Eastwood, J. D. (2013). The Measurement of Boredom: Differences Between Existing Self-Report Scales. *Assessment*, *20*(5), 585–596. https://doi.org/10.1177/1073191111408229

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122. https://doi.org/10.1007/BF02294746

Meyer, S., & Schlesier, J. (2021). The development of students' achievement emotions after transition to secondary school: A multilevel growth curve modelling approach. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-021-00533-5

Mikulas, W. L., & Vodanovich, S. J. (1993). The Essence of Boredom. *The Psychological Record*, *43*(1). https://www.proquest.com/docview/1301204061/citation/9702D668DE9F45B6PQ/1

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

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2nd ed.). Springer.

Moosbrugger, H., & Schermelleh-Engel, K. (2012). Exploratorische (EFA) und Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 325–343). Springer. https://doi.org/10.1007/978-3-642-20072-4_13

Mostafa, T., & Schwabe, M. (2019). *Programme for International Student Assessment (PISA)—Results from PISA 2018—Country Note Germany*. OECD. https://www.oecd.org/pisa/publications/PISA2018_CN_DEU.pdf

Mulaik, S. A. (2009). Factor Scores and Factor Indeterminacy. In *Foundations of Factor Analysis* (2nd ed.). Chapman and Hall/CRC.

Müller, P., & Schäfer, S. (2017). Latent Mean (Comparison). In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (1st ed., pp. 1–7). Wiley. https://doi.org/10.1002/9781118901731.iecrm0132

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585. https://doi.org/10.1007/BF02296397

Muthén, L. K., & Muthén, B. O. (1998-2019). *Mplus* (8.4) [Computer software]. Author.

National Accociation for Gifted Children. (n.d.). *Definitions of giftedness*. Retrieved July 14, 2021, from https://www.nagc.org/resources-publications/resources/what-giftedness

Nett, U. E., Goetz, T., & Daniels, L. M. (2010). What to do when feeling bored? *Learning and Individual Differences*, *20*(6), 626–638. https://doi.org/10.1016/j.lindif.2010.09.004

Nett, U. E., Goetz, T., & Hall, N. C. (2011). Coping with boredom in school: An experience sampling perspective. *Contemporary Educational Psychology*, *36*(1), 49–59. https://doi.org/10.1016/j.cedpsych.2010.10.003

Obergriesser, S., & Stoeger, H. (2015). The role of emotions, motivation, and learning behavior in underachievement and results of an intervention. *High Ability Studies*, *26*(1), 167–190. https://doi.org/10.1080/13598139.2015.1043003

Obergriesser, S., & Stoeger, H. (2016). The influence of emotions and learning preferences on learning strategy use before transition into high-achiever track secondary school. *High Ability Studies*, *27*(1), 5–38. https://doi.org/10.1080/13598139.2015.1100980

OECD. (2012). *Closing the Gender Gap: Act Now*. OECD Publishing. https://doi.org/10.1787/9789264179370-en

Oud, J. H. L. (2007). Continuous Time Modeling of Reciprocal Relationships in the Cross–Lagged Panel Design. In S. M. Boker & M. J. Wenger (Eds.), *Data Ana.ytic Techniques for Dynamical Systems* (pp. 87–129). Erlbaum Associates Hillsdale, NJ.

Parr, A., Amemiya, J., & Wang, M.-T. (2019). Student learning emotions in middle school mathematics classrooms: Investigating associations with dialogic instructional practices. *Educational Psychology*, *39*(5), 636–658. https://doi.org/10.1080/01443410.2018.1560395

Peixoto, F., Mata, L., Monteiro, V., Sanches, C., & Pekrun, R. (2015). The Achievement Emotions Questionnaire: Validation for Pre-Adolescent Students. *European Journal of Developmental Psychology*, *12*(4), 472–481. https://doi.org/10.1080/17405629.2015.1040757

Peixoto, F., Sanches, C., Mata, L., & Monteiro, V. (2017). “How do you feel about math?”: Relationships between competence and value appraisals, achievement emotions and academic achievement. *European Journal of Psychology of Education*, *32*(3), 385–405. https://doi.org/10.1007/s10212-016-0299-4

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2017). Emotion and Achievement During Adolescence. *Child Development Perspectives*, *11*(3), 215–221. https://doi.org/10.1111/cdep.12237

Pekrun, R. (2019). Achievement Emotions—A control-value theory perspective. In R. Patulny, A. Bellocchi, R. E. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in Late Modernity* (pp. 142–157). Routledge.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549. https://doi.org/10.1037/a0019243

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., & Perry, R. P. (2005). *Achievement Emotions Questionnaire (AEQ)—User's Manual*. Department of Psychology, University of Munich.

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, *106*(3), 696–710. https://doi.org/10.1037/a0036006

Pekrun, R., Jullien, S., Zirngibl, A., Blum, W., Götz, T., vom Hofe, R., Jordan, A., Kleine, M., & Wartha, S. (2003). *PALMA – Projekt zur Analyse der Leistungsentwicklung in Mathematik: Skalenhandbuch Erhebungswelle II* [Unpublished project codebook]. Department of Psychology, University of Munich.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., Marsh, H. W., Elliot, A. J., Stockinger, K., Perry, R. P., Vogl, E., Goetz, T., van Tilburg, W. A. P., Lüdtke, O., & Vispoel, W. P. (2023). A three-dimensional taxonomy of achievement emotions. *Journal of Personality and Social Psychology*, *124*(1), 145–178. https://doi.org/10.1037/pspp0000448

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG priority programme* (pp. 17–37). Waxmann.

Pennington, C. R., Kaye, L. K., Qureshi, A. W., & Heim, D. (2021). Do gender differences in academic attainment correspond with scholastic attitudes? An exploratory study in a UK secondary school. *Journal of Applied Social Psychology*, *51*(1), 3–16. https://doi.org/10.1111/jasp.12711

Petersen, J. (2018). Gender Difference in Verbal Performance: A Meta-analysis of United States State Performance Assessments. *Educational Psychology Review*, *30*(4), 1269–1281. https://doi.org/10.1007/s10648-018-9450-x

Plucker, J. A., Robinson, N. M., Greenspon, T. S., Feldhusen, J. F., McCoach, D. B., & Subotnik, R. F. (2004). It’s not how the pond makes you feel, but rather how high you can jump. *American Psychologist*, *59*, 268–269. https://doi.org/10.1037/0003-066X.59.4.268

Preckel, F., & Brüll, M. (2010). The benefit of being a big fish in a big pond: Contrast and assimilation effects on academic self-concept. *Learning and Individual Differences*, *20*(5), 522–531. https://doi.org/10.1016/j.lindif.2009.12.007

Preckel, F., Götz, T., & Frenzel, A. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology*, *80*(3), 451–472. https://doi.org/10.1348/000709909X480716

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., Scherrer, V., & Schneider, W. (2019). High-Ability Grouping: Benefits for Gifted Students’

Achievement Development Without Costs in Academic Self-Concept. *Child Development*, *90*(4), 1185–1201. https://doi.org/10.1111/cdev.12996

Putwain, D. W., Pekrun, R., Nicholson, L. J., Symes, W., Becker, S., & Marsh, H. W. (2018). Control-Value Appraisals, Enjoyment, and Boredom in Mathematics: A Longitudinal Latent Interaction Analysis. *American Educational Research Journal*, *55*(6), 1339–1368. https://doi.org/10.3102/0002831218786689

Putwain, D. W., Wood, P., & Pekrun, R. (2022). Achievement emotions and academic achievement: Reciprocal relations and the moderating influence of academic buoyancy. *Journal of Educational Psychology*, *114*(1), 108–126. https://doi.org/10.1037/edu0000637

R Core Team. (2021). *R: A language and environment for statistical computing* (4.1.0) [Computer software]. R Foundation for Statistical Computing.

Raccanello, D., Brondino, M., Moè, A., Stupnisky, R., & Lichtenfeld, S. (2019). Enjoyment, Boredom, Anxiety in Elementary Schools in Two Domains: Relations With Achievement. *The Journal of Experimental Education*, *87*(3), 449–469. https://doi.org/10.1080/00220973.2018.1448747

Raffaelli, Q., Mills, C., & Christoff, K. (2018). The knowns and unknowns of boredom: A review of the literature. *Experimental Brain Research*, *236*(9), 2451–2462. https://doi.org/10.1007/s00221-017-4922-7

Ragheb, M. G., & Merydith, S. P. (2001). Development and validation of a multidimensional scale measuring free time boredom. *Leisure Studies*, *20*(1), 41–59. https://doi.org/10.1080/02614360122569

Reijseger, G., Schaufeli, W. B., Peeters, M. C. W., Taris, T. W., Beek, I. van, & Ouweneel, E. (2013). Watching the paint dry at work: Psychometric examination of the Dutch

Boredom Scale. *Anxiety, Stress, & Coping*, *26*(5), 508–525. https://doi.org/10.1080/10615806.2012.720676

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, *74*(4), 445–458. https://doi.org/10.1037/amp0000356

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. (2014). *Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package*. RAND Corporation.

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. (2015). *The twang package* (1.6) [Computer software]. https://CRAN.R-project.org/package=twang

Robinson, W. P. (1975). Boredom At School. *British Journal of Educational Psychology*, *45*(2), 141–152. https://doi.org/10.1111/j.2044-8279.1975.tb03239.x

Rogers, K. B. (2007). Lessons Learned About Educating the Gifted and Talented: A Synthesis of the Research on Educational Practice. *Gifted Child Quarterly*, *51*(4), 382–396. https://doi.org/10.1177/0016986207306324

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 15. http://dx.doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, *79*(387), 9. https://doi.org/10.2307/2288398

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38. https://doi.org/10.2307/2683903

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Ruthig, J. C., Perry, R. P., Hladkyj, S., Hall, N. C., Pekrun, R., & Chipperfield, J. G. (2008). Perceived control and emotions: Interactive effects on performance in achievement settings. *Social Psychology of Education*, *11*(2), 161–180. https://doi.org/10.1007/s11218-007-9040-0

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

Scherer, K. R. (2009). Emotions are emergent processes: They require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3459–3474. https://doi.org/10.1098/rstb.2009.0141

Scherrer, V., & Preckel, F. (2019). Development of Motivational Variables and Self-Esteem During the School Career: A Meta-Analysis of Longitudinal Studies. *Review of Educational Research*, *89*(2), 211–258. https://doi.org/10.3102/0034654318819127

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Annals of Internal Medicine*, *152*(11), 726–732. https://doi.org/10.7326/0003-4819-152-11-201006010-00232

Schwartze, M. M., Frenzel, A. C., Goetz, T., Pekrun, R., Reck, C., Marx, A. K. G., & Fiedler, D. (2021). Boredom Makes Me Sick: Adolescents' Boredom Trajectories and Their Health-Related Quality of Life. *International Journal of Environmental Research and Public Health*, *18*(12), 6308. https://doi.org/10.3390/ijerph18126308

Sedlmeier, P., & Renkewitz, F. (2013). *Forschungsmethoden und Statistik* (2nd ed.). Pearson.

Sommers, J., & Vodanovich, S. J. (2000). Boredom proneness: Its relationship to psychological- and physical-health symptoms. *Journal of Clinical Psychology*, *56*, 149–155. https://doi.org/10.1002/(SICI)1097-4679(200001)56:1<149::AID-JCLP14>3.0.CO;2-Y

Spaeth, M., Weichold, K., & Silbereisen, R. K. (2015). The development of leisure boredom in early adolescence: Predictors and longitudinal associations with delinquency and depression. *Developmental Psychology, 51*(10), 1380–1394. https://doi.org/10.1037/a0039480

Stambaugh, T. W. (2017). Gifted Students and Mental Health: The Role of Boredom, Belonging, Friendship, Service Delivery, and Academic Challenge [Ph.D., Trevecca Nazarene University]. In *ProQuest Dissertations and Theses* (2122116662). https://www.proquest.com/dissertations-theses/gifted-students-mental-health-role-boredom/docview/2122116662/se-2?accountid=14652

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses. *Review of Educational Research*, *86*(4), 849–899. https://doi.org/10.3102/0034654316675417

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking Giftedness and Gifted Education: A Proposed Direction Forward Based on Psychological Science.

Psychological Science in the Public Interest, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056

Sundberg, N., & Bisno, H. (1983). Boredom at life transitions—Adolescence and old age. *Meeting of the Western Psychological Association, San Francisco, CA*.

Sutter-Brandenberger, C. C., Hagenauer, G., & Hascher, T. (2018). Students' self-determined motivation and negative emotions in mathematics in lower secondary education—Investigating reciprocal relations. *Contemporary Educational Psychology*, *55*, 166–175. https://doi.org/10.1016/j.cedpsych.2018.10.002

Tam, K. Y. Y., Poon, C. Y. S., Hui, V. K. Y., Wong, C. Y. F., Kwong, V. W. Y., Yuen, G. W. C., & Chan, C. S. (2020). Boredom begets boredom: An experience sampling study on the impact of teacher boredom on student boredom and motivation. *British Journal of Educational Psychology*, *90*(S1), 124–137. https://doi.org/10.1111/bjep.12309

Terman, L. M. (1922). A new approach to the study of genius. *Psychological Review*, *29*(4), 310. https://doi.org/10.1037/h0071072

Thompson, M. S., & Green, S. B. (2013). Evaluating Between-Group Differences in Latent Variable Means. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 163–218). Information Age Publishing. https://www.infoagepub.com/products/Structural-Equation-Modeling-2nd-Edition

Titz, W. (2001). *Emotionen von Studierenden in Lernsituationen: Explorative Analysen und Entwicklung von Selbstberichtskalen – Anhang* [Unpublished doctoral dissertation]. https://epub.uni-regensburg.de/9862/

Todman, M. (2013). The dimensions of state boredom: Frequency, duration, unpleasantness, consequences and causal attributions. *Educational Research International*, *1*(1), 32–40.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison Wesley.

Tze, V. M. C., Daniels, L. M., & Klassen, R. M. (2014). Examining the factor structure and validity of the English Precursors to Boredom Scales. *Learning and Individual Differences*, *32*, 254–260. https://doi.org/10.1016/j.lindif.2014.03.018

Tze, V. M. C., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the Relationship Between Boredom and Academic Outcomes: A Meta-Analysis. *Educational Psychology Review*, *28*(1), 119–144. https://doi.org/10.1007/s10648-015-9301-y

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

van Hooft, E. A. J., & van Hooff, M. L. M. (2018). The state of boredom: Frustrating or depressing? *Motivation and Emotion*, *42*(6), 931–946. https://doi.org/10.1007/s11031-018-9710-6

van Tilburg, W. A. P., & Igou, E. R. (2017). Boredom begs to differ: Differentiation from other negative emotions. *Emotion*, *17*(2), 309–322. https://doi.org/10.1037/emo0000233

Vierhaus, M., Lohaus, A., & Wild, E. (2016). The development of achievement emotions and coping/emotion regulation from primary to secondary school. *Learning and Instruction*, *42*, 12–21. https://doi.org/10.1016/j.learninstruc.2015.11.002

Vodanovich, S. J., & Kass, S. (1990). A Factor Analytic Study of the Boredom Proneness Scale. *Journal of Personality Assessment*, *55*(1), 115–123. https://doi.org/10.1207/s15327752jpa5501&2_11

Vodanovich, S. J., & Watt, J. D. (2016). Self-Report Measures of Boredom: An Updated Review of the Literature. *The Journal of Psychology*, *150*(2), 196–228. https://doi.org/10.1080/00223980.2015.1074531

Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176–192. https://doi.org/10.1037/a0027543

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Vuletich, H. A., Kurtz-Costes, B., Cooley, E., & Payne, B. K. (2020). Math and language gender stereotypes: Age and gender differences in implicit biases and explicit beliefs. *PLOS ONE*, *15*(9), e0238230. https://doi.org/10.1371/journal.pone.0238230

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. John Wiley & Sons.

Watt, J. D., & Vodanovich, S. J. (1999). Boredom proneness and psychosocial development. *The Journal of Psychology*, *133*(3), 303–314. https://doi.org/10.1080/00223989909599743

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573.

Westgate, E. C. (2020). Why Boredom Is Interesting. *Current Directions in Psychological Science*, *29*(1), 33–40. https://doi.org/10.1177/0963721419884309

Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement. *Psychological Review*, *125*(5), 689. https://doi.org/10.1037/rev0000097

Weybright, E. H., Caldwell, L. L., Ram, N., Smith, E. A., & Wegner, L. (2015). Boredom Prone or Nothing to Do? Distinguishing Between State and Trait Leisure Boredom and Its Association with Substance Use in South African Adolescents. *Leisure Sciences*, *37*(4), 311–331. https://doi.org/10.1080/01490400.2015.1014530

Weybright, E. H., Schulenberg, J., & Caldwell, L. L. (2020). More Bored Today Than Yesterday? National Trends in Adolescent Boredom From 2008 to 2017. *Journal of Adolescent Health*, *66*(3), 360–365. https://doi.org/10.1016/j.jadohealth.2019.09.021

Wickham, H. (2016). *ggplot 2: Elegant Graphics for Data Analysis* (3.3.5) [Computer software]. Springer. https://ggplot2.tidyverse.org

Wigfield, A., Ecces, J., & Möller, J. (2020). How dimensional comparisons help to understand linkages between expectancies, values, performance and choice. *Educational Psychology Review*, *32*(6), 657–680. https://doi.org/10.1007/s10648-020-09524-2

Wigfield, A., & Wagner, A. L. (2005). Competence, Motivation, and Identity Development during Adolescence. In *Handbook of competence and motivation* (pp. 222–239). Guilford Publications.

Williams, M. E. (2015). A phenomenological exploration of the perceptions of gifted adolescents coping with stress [Ed.D., Lamar University - Beaumont]. In *ProQuest Dissertations and Theses* (1728125175). https://www.proquest.com/dissertations-theses/phenomenological-exploration-perceptions-gifted/docview/1728125175/se-2?accountid=14652

Willis, J. (2014). Neuroscience Reveals That Boredom Hurts. *Phi Delta Kappan*, *95*(8), 28–32. https://doi.org/10.1177/003172171409500807

Winner, E. (1998). Uncommon talents: Gifted children, prodigies, and savants. *Scientific American Presents: Exploring Intelligence*, *9*(4), 32–37.

Wirthwein, L., Bergold, S., Preckel, F., & Steinmayr, R. (2019). Personality and school functioning of intellectually gifted and nongifted adolescents: Self-perceptions and parents' assessments. *Learning and Individual Differences*, *73*, 16–29. https://doi.org/10.1016/j.lindif.2019.04.003

Zaccoletti, S., Altoè, G., & Mason, L. (2020). Enjoyment, anxiety and boredom, and their control-value antecedents as predictors of reading comprehension. *Learning and Individual Differences*, *79*, 101869. https://doi.org/10.1016/j.lindif.2020.101869

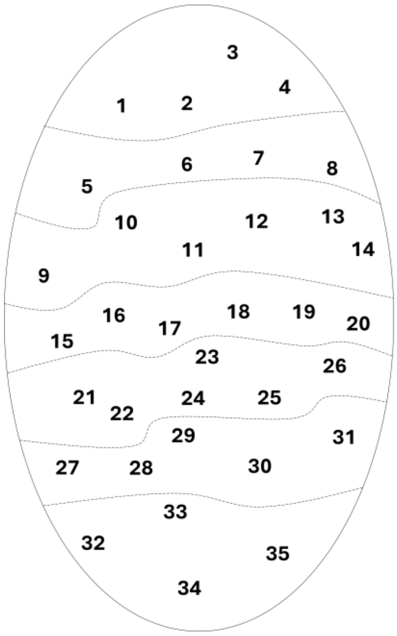
Zirkel, S., Garcia, J. A., & Murphy, M. C. (2015). Experience-Sampling Research Methods and Their Potential for Education Research. *Educational Researcher*, *44*(1), 7–16. https://doi.org/10.3102/0013189X14566879

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal. 1979*. Erlbaum Associates Hillsdale, NJ.

Appendix 1: Picture Sources for Many Faces of Boredom-Collage

Figure A1.1

Legend to Many Faces of Boredom-Collage



Note. See list of picture sources below. Pictures were used noncommercially and for design purposes only. Date of retrieval for all pictures was March 6th, 2023.

List of picture sources:

1. https://www.vecteezy.com/png/14604084-girl-bored-face-cartoon-cute
2. https://openclipart.org/detail/297180/sleeping-student
3. https://globalsymbols.com/symbolsets/arasaac/symbols/23511?locale=en
4. https://mystickermania.com/sticker-packs/tom-and-jerry/tom-and-jerry-bored-tom
5. https://www.iconfinder.com/icons/6815019/bored_emoticon_girl_sticker_thinking_tired_vee_icon
6. https://www.flaticon.com/free-icon/boring_5965301
7. https://freesvg.org/vector-image-of-bored-green-alligator
8. https://boredapetronclub.com/
9. https://lyfas.com/testdo-you-get-bored-often-check-your-boredom-proneness-with-validated-online-mental-health-self-screening-and-assessment-test-with-cardiometabolicautonomicstess-risk//
10. https://www.litmos.com/de-DE/resources/downloads/no-boring-learning-infographic
11. https://en.wikipedia.org/wiki/Boredom#/media/File:La_Touche_Lennui_1893.jpg
12. https://peakeducationalresources.blogspot.com/2012/08/a-different-way-of-looking-at-boredom.html
13. https://pngtree.com/so/bored

14. https://pngtree.com/free-png-vectors/bored
15. https://www.boringstartupstuff.com/
16. https://www.iconfinder.com/icons/391399/bleh_bored_boring_emoticon_icon
17. https://www.clipartmax.com/max/m2i8H7H7G6m2A0H7/
18. https://www.vexels.com/png-svg/preview/143934/bored-face-emoji
19. https://youtooz.com/products/bored-squidward
20. https://www.flaticon.com/free-icon/boring_3121559
21. https://www.thecollienois.com/how-to-tell-if-your-dog-is-bored/
22. https://news.reflexmath.com/lp/worksheetsb.php
23. https://www.iconfinder.com/icons/3589449/behavior_bore_boring_dull_lazy_x_icon
24. https://creazilla.com/nodes/76774-boredom-clipart
25. https://www.flaticon.com/free-icon/bored_3220820
26. https://www.psychologytoday.com/us/blog/how-do-life/201408/bored
27. https://creazilla.com/nodes/21093-bored-bee-clipart
28. https://www.deviantart.com/paulthedualartist/art/Super-Mario-Toad-bored-2D-922871866
29. https://www.flaticon.com/free-icon/boring_155261
30. https://commons.wikimedia.org/wiki/File:Bored_Cartoon_Man_Using_A_Computer.svg
31. https://boredhoard.com/
32. https://kuku-keke.com/en/archives/466031
33. https://www.acap.edu.au/newsletters/five-signs-its-time-for-a-career-change/
34. https://www.pinterest.de/pin/771804454881111972/
35. https://freesvg.org/bored-owl-head-vector-image

Appendix 2: Supplemental Material to Study 1

Table A2.1

Manifest (Upper Array) and Latent Correlations (Lower Array) of Study Variables

	BO-I$_{T1}$	BO-U$_{T1}$	BO-O$_{T1}$	ASC$_{T1}$	BO-I$_{T2}$	BO-U$_{T2}$	BO-O$_{T2}$	ASC$_{T2}$	BO-I$_{T3}$	BO-U$_{T3}$	BO-O$_{T3}$	ASC$_{T3}$	BO-I$_{T4}$	BO-U$_{T4}$	BO-O$_{T4}$	ASC$_{T4}$
								Mathematics								
BO-I$_{T1}$		**.547**	**.264**	**-.266**	**.463**	**.281**	**.166**	**-.221**	**.308**	**.171**	**.171**	**-.170**	**.185**	.008	**.088**	**-.121**
BO-U$_{T1}$	**.796**		**.131**	**.070**	**.315**	**.446**	**.096**	.049	**.222**	**.341**	**.082**	.014	**.083**	**.141**	-.018	.047
BO-O$_{T1}$	**.374**	**.177**		**-.348**	**.197**	.031	**.502**	**-.307**	**.106**	-.025	**.335**	**-.197**	**.116**	-.030	**.167**	**-.124**
ASC$_{T1}$	**-.346**	**.095**	**-.453**		**-.242**	**.113**	**-.307**	**.688**	**-.129**	**.232**	**-.268**	**.504**	**-.145**	**.260**	**-.266**	**.410**
BO-I$_{T2}$	**.611**	**.492**	**.279**	**-.292**		**.515**	**.333**	**-.286**	**.342**	**.160**	**.177**	**-.161**	**.258**	.043	**.140**	**-.121**
BO-U$_{T2}$	**.484**	**.632**	.067	**.125**	**.733**		**.147**	**.088**	**.238**	**.398**	.042	.067	**.163**	**.269**	-.012	**.076**
BO-O$_{T2}$	**.242**	**.136**	**.721**	**-.387**	**.453**	**.188**		**-.435**	**.186**	-.019	**.440**	**-.308**	**.153**	**-.107**	**.324**	**-.218**
ASC$_{T2}$	**-.273**	.046	**-.381**	**.758**	**-.358**	**.110**	**-.543**		**-.157**	**.261**	**-.321**	**.621**	**-.174**	**.278**	**-.324**	**.479**
BO-I$_{T3}$	**.276**	**.231**	**.162**	**-.146**	**.443**	**.348**	**.258**	**-.194**		**.344**	**.353**	**-.325**	**.323**	.045	**.109**	**-.150**
BO-U$_{T3}$	**.164**	**.352**	-.040	**.279**	**.248**	**.573**	-.006	**.321**	**.454**		-.032	**.264**	**.096**	**.452**	**-.151**	**.258**
BO-O$_{T3}$	**.160**	.060	**.418**	**-.295**	**.287**	.075	**.579**	**-.400**	**.453**	-.055		**-.491**	**.166**	**-.161**	**.352**	**-.292**
ASC$_{T3}$	**-.177**	.035	**-.254**	**.514**	**-.224**	.078	**-.364**	**.683**	**-.381**	**.342**	**-.579**		**-.268**	**.326**	**-.418**	**.618**
BO-I$_{T4}$	**.133**	**.090**	**.088**	**-.138**	**.189**	**.135**	**.140**	**-.191**	**.378**	**.146**	**.220**	**-.321**		**.071**	**.462**	**-.458**
BO-U$_{T4}$	.037	**.190**	**-.093**	**.267**	.055	**.316**	**-.104**	**.322**	.077	**.606**	**-.203**	**.406**	.093		**-.275**	**.498**
BO-O$_{T4}$	**.072**	-.026	**.198**	**-.255**	**.101**	-.052	**.274**	**-.344**	**.112**	**-.200**	**.451**	**-.487**	**.560**	**-.354**		**-.608**
ASC$_{T4}$	**-.090**	.062	**-.149**	**.354**	**-.100**	**.113**	**-.210**	**.467**	**-.152**	**.322**	**-.330**	**.671**	**-.535**	**.614**	**-.696**	
								German								
BO-I$_{T1}$		**.591**	**.331**	**-.339**	**.444**	**.344**	**.174**	**-.230**	**.270**	**.203**	**.196**	**-.171**	**.172**	.015	**.107**	**-.102**
BO-U$_{T1}$	**.827**		**.324**	**-.084**	**.390**	**.471**	**.181**	**-.115**	**.265**	**.329**	**.171**	**-.087**	**.131**	**.147**	**.123**	-.054

BO-O$_{T1}$	**.438**	**.431**		**-.304**	**.258**	**.197**	**.442**	**-.249**	**.209**	**.109**	**.381**	**-.225**	**.146**	.005	**.262**	**-.208**
ASC$_{T1}$	**-.409**	**-.110**	**-.375**		**-.206**	-.055	**-.238**	**.585**	**-.142**	.064	**-.226**	**.455**	**-.170**	**.116**	**-.232**	**.378**
BO-I$_{T2}$	**.530**	**.528**	**.342**	**-.223**		**.585**	**.379**	**-.311**	**.362**	**.231**	**.215**	**-.213**	**.178**	**.100**	**.133**	**-.126**
BO-U$_{T2}$	**.488**	**.661**	**.277**	-.032	**.796**		**.298**	-.031	**.267**	**.347**	**.134**	-.040	**.126**	**.155**	**.078**	-.030
BO-O$_{T2}$	**.248**	**.253**	**.594**	**-.290**	**.498**	**.399**		**-.347**	**.228**	**.112**	**.427**	**-.270**	**.082**	-.007	**.277**	**-.198**
ASC$_{T2}$	**-.259**	**-.126**	**-.297**	**.662**	**-.357**	-.037	**-.411**		**-.205**	.052	**-.294**	**.574**	**-.122**	**.138**	**-.183**	**.438**
BO-I$_{T3}$	**.254**	**.270**	**.207**	**-.151**	**.445**	**.392**	**.309**	**-.228**		**.511**	**.290**	**-.325**	**.317**	**.161**	**.145**	**-.148**
BO-U$_{T3}$	**.211**	**.329**	**.114**	.044	**.325**	**.505**	**.167**	.079	**.669**		**.110**	**.104**	**.159**	**.287**	.008	**.112**
BO-O$_{T3}$	**.156**	**.151**	**.358**	**-.243**	**.282**	**.213**	**.592**	**-.349**	**.369**	**.154**		**-.400**	**.104**	-.068	**.354**	**-.259**
ASC$_{T3}$	**-.182**	**-.110**	**-.227**	**.441**	**-.245**	-.064	**-.324**	**.662**	**-.376**	**.124**	**-.485**		-,157	,121	-,250	,502
BO-I$_{T4}$	**.101**	**.100**	**.079**	**-.086**	**.170**	**.136**	**.115**	**-.131**	**.377**	**.218**	**.133**	**-.212**		**.339**	**.366**	**-.418**
BO-U$_{T4}$	**.050**	**.094**	-.005	**.068**	**.083**	**.155**	-.012	**.104**	**.214**	**.348**	-.065	**.152**	**.449**		.061	**.098**
BO-O$_{T4}$	**.080**	.067	**.184**	**-.144**	**.145**	.090	**.303**	**-.210**	**.195**	.034	**.511**	**-.298**	**.464**	.075		**-.435**
ASC$_{T4}$	**-.094**	-.047	**-.134**	**.259**	**-.127**	-.013	**-.194**	**.388**	**-.180**	**.125**	**-.301**	**.584**	**-.477**	**.121**	**-.519**	

Note. Significant correlations ($p < .05$) in boldface. Concurrent correlations framed. Latent correlations taken from Mplus's tech 4-outputs for 4-Variable REMs. Underlying sample sizes for manifest vs. latent correlation may differ due to differential treatment of missing data (pairwise deletion for manifest correlations; FIML for latent correlations). BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

Table A2.2

Standardized Latent (Residual) Correlations and Path Coefficients from REMs

Coefficients	Mathematics 3-Variable-REM	Mathematics 4-Variable-REM	German 3-Variable-REM	German 4-Variable-REM
		Correlations[a]		
T1				
r (BO-I, BO-U)	**.811 (.033)*****	**.796 (.038)*****	**.810 (.040)*****	**.827 (.041)*****
r (BO-I, BO-O)	**.372 (.051)*****	**.374 (.055)*****	**.424 (.065)*****	**.438 (.065)*****
r (BO-U, BO-O)	**.196 (.067)****	**.177 (.070)***	**.435 (.045)*****	**.431 (.045)*****
r (BO-I, ASC)		**-.346 (.047)*****		**-.409 (.045)*****
r (BO-U, ASC)		**.095 (.036)****		**-.110 (.042)****
r (BO-O, ASC)		**-.453 (.043)*****		**-.375 (.046)*****
T2				
r (BO-I, BO-U)	**.735 (.060)*****	**.731 (.055)*****	**.754 (.076)*****	**.756 (.082)*****
r (BO-I, BO-O)	**.499 (.068)*****	**.478 (.069)*****	**.435 (.076)*****	**.445 (.073)*****
r (BO-U, BO-O)	**.268 (.078)****	**.272 (.091)****	**.384 (.078)*****	**.393 (.075)*****
r (BO-I, ASC)		**-.249 (.057)*****		**-.313 (.056)*****
r (BO-U, ASC)		.059 (.064)$^{n.s.}$		.058 (.071)$^{n.s.}$
r (BO-O, ASC)		**-.500 (.065)*****		**-.311 (.045)*****
T3				
r (BO-I, BO-U)	**.514 (.060)*****	**.482 (.060)*****	**.664 (.057)*****	**.666 (.061)*****
r (BO-I, BO-O)	**.394 (.048)*****	**.410 (.046)*****	**.260 (.042)*****	**.259 (.047)*****
r (BO-U, BO-O)	.002 (.067)$^{n.s.}$	-.007 (.072)$^{n.s.}$	.100 (.069)$^{n.s.}$	.108 (.073)$^{n.s.}$
r (BO-I, ASC)		**-.393 (.050)*****		**-.322 (.057)*****
r (BO-U, ASC)		**.223 (.066)****		**.153 (.055)****
r (BO-O, ASC)		**-.527 (.044)*****		**-.376 (.059)*****
T4				
r (BO-I, BO-U)	**.176 (.082)***	.148 (.080)$^{n.s.}$	**.446 (.089)*****	**.451 (.087)*****
r (BO-I, BO-O)	**.592 (.045)*****	**.578 (.046)*****	**.486 (.045)*****	**.492 (.047)*****
r (BO-U, BO-O)	**-.221 (.091)***	**-.223 (.088)***	**.174 (.074)***	**.159 (.076)***
r (BO-I, ASC)		**-.537 (.049)*****		**-.494 (.046)*****
r (BO-U, ASC)		**.542 (.077)*****		**.015 (.075)**$^{n.s.}$
r (BO-O, ASC)		**-.608 (.043)*****		**-.483 (.053)*****
		Path coefficients		
T1 → T2				
β (BO-I → BO-I)	**.625 (.128)*****	.398 (.210)$^{n.s.}$	**.317 (.106)****	.200 (.156)$^{n.s.}$
β (BO-I → BO-U)	.013 (.164)$^{n.s.}$	.053 (.254)$^{n.s.}$	-.177 (.132)$^{n.s.}$	-.218 (.209)$^{n.s.}$
β (BO-I → BO-O)	-.157 (.143)$^{n.s.}$	-.322 (.241)$^{n.s.}$	-.031 (.089)$^{n.s.}$	-.150 (.127)$^{n.s.}$
β (BO-I → ASC)		.065 (.184)$^{n.s.}$		.247 (.166)$^{n.s.}$
β (BO-U → BO-I)	-.015 (.113)$^{n.s.}$	.186 (.190)$^{n.s.}$	.205 (.106)$^{n.s.}$	**.315 (.151)***
β (BO-U → BO-U)	**.628 (.147)*****	**.586 (.228)***	**.805 (.134)*****	**.839 (.201)*****
β (BO-U → BO-O)	.148 (.139)$^{n.s.}$	.289 (.232)$^{n.s.}$	.032 (.094)$^{n.s.}$	.122 (.124)$^{n.s.}$
β (BO-U → ASC)		-.070 (.162)$^{n.s.}$		-.237 (.156)$^{n.s.}$
β (BO-O → BO-I)	.036 (.052)$^{n.s.}$	.024 (.053)$^{n.s.}$	**.115 (.045)***	.092 (.057)$^{n.s.}$
β (BO-O → BO-U)	-.049 (.063)$^{n.s.}$	-.021 (.062)$^{n.s.}$	-.005 (.050)$^{n.s.}$	-.001 (.058)$^{n.s.}$
β (BO-O → BO-O)	**.747 (.048)*****	**.695 (.058)*****	**.603 (.059)*****	**.559 (.069)*****

β (BO-O → ASC)		-.046 (.043)$^{n.s.}$		-.031 (.065)$^{n.s.}$
β (ASC → BO-I)		-.161 (.103)$^{n.s.}$		-.072 (.080)$^{n.s.}$
β (ASC → BO-U)		.078 (.112)$^{n.s.}$		-.029 (.083)$^{n.s.}$
β (ASC → BO-O)		-.211 (.116)$^{n.s.}$		**-.129 (.063)***
β (ASC → ASC)		**.766 (.089)*****		**.725 (.079)*****
T2 → T3				
β (BO-I → BO-I)	**.353 (.126)****	.301 (.188)$^{n.s.}$	**.291 (.113)***	.227 (.171)$^{n.s.}$
β (BO-I → BO-U)	-.311 (.167)$^{n.s.}$	-.231 (.215)$^{n.s.}$	-.226 (.140)$^{n.s.}$	-.180 (.237)$^{n.s.}$
β (BO-I → BO-O)	.063 (.120)$^{n.s.}$	.065 (.164)$^{n.s.}$	-.026 (.140)$^{n.s.}$	-.106 (.206)$^{n.s.}$
β (BO-I → ASC)		.067 (.120)$^{n.s.}$		.101 (.134)$^{n.s.}$
β (BO-U → BO-I)	.094 (.104)$^{n.s.}$	.124 (.155)$^{n.s.}$	.111 (.106)$^{n.s.}$	.174 (.167)$^{n.s.}$
β (BO-U → BO-U)	**.798 (.147)*****	**.706 (.195)*****	**.672 (.123)*****	**.642 (.218)****
β (BO-U → BO-O)	-.054 (.107)$^{n.s.}$	-.058 (.150)$^{n.s.}$	-.006 (.132)$^{n.s.}$	.071 (.198)$^{n.s.}$
β (BO-U → ASC)		-.050 (.098)$^{n.s.}$		-.095 (.125)$^{n.s.}$
β (BO-O → BO-I)	.080 (.061)$^{n.s.}$	.062 (.067)$^{n.s.}$	**.136 (.037)*****	.083 (.048)$^{n.s.}$
β (BO-O → BO-U)	-.035 (.069)$^{n.s.}$	.075 (.068)$^{n.s.}$	.030 (.062)$^{n.s.}$	.019 (.066)$^{n.s.}$
β (BO-O → BO-O)	**.577 (.062)*****	**.509 (.072)*****	**.614 (.065)*****	**.552 (.074)*****
β (BO-O → ASC)		.004 (.063)$^{n.s.}$		-.062 (.049)$^{n.s.}$
β (ASC → BO-I)		-.066 (.093)$^{n.s.}$		-.106 (.081)$^{n.s.}$
β (ASC → BO-U)		**.201 (.100)***		.046 (.096)$^{n.s.}$
β (ASC → BO-O)		-.094 (.103)$^{n.s.}$		-.158 (.082)$^{n.s.}$
β (ASC → ASC)		**.715 (.059)*****		**.669 (.074)*****
T3 → T4				
β (BO-I → BO-I)	**.357 (.090)*****	.200 (.118)$^{n.s.}$	**.414 (.075)*****	**.361 (.102)*****
β (BO-I → BO-U)	**-.206 (.072)****	-.155 (.085)$^{n.s.}$	.017 (.095)$^{n.s.}$	.103 (.119)$^{n.s.}$
β (BO-I → BO-O)	-.041 (.077)$^{n.s.}$	**-.204 (.087)***	.051 (.091)$^{n.s.}$	.049 (.106)$^{n.s.}$
β (BO-I → ASC)		.074 (.085)$^{n.s.}$		.008 (.094)$^{n.s.}$
β (BO-U → BO-I)	.015 (.079)$^{n.s.}$	.167 (.116)$^{n.s.}$	-.044 (.075)$^{n.s.}$	-.004 (.106)$^{n.s.}$
β (BO-U → BO-U)	**.690 (.079)*****	**.637 (.101)*****	**.365 (.096)*****	**.280 (.124)***
β (BO-U → BO-O)	-.125 (.067)$^{n.s.}$	.047 (.080)$^{n.s.}$	-.067 (.087)$^{n.s.}$	-.070 (.113)$^{n.s.}$
β (BO-U → ASC)		.047 (.078)$^{n.s.}$		.057 (.099)$^{n.s.}$
β (BO-O → BO-I)	.052 (.062)$^{n.s.}$	-.055 (.054)$^{n.s.}$	-.010 (.066)$^{n.s.}$	-.047 (.072)$^{n.s.}$
β (BO-O → BO-U)	-.089 (.052)$^{n.s.}$	-.036 (.065)$^{n.s.}$	-.124 (.064)$^{n.s.}$	-.093 (.074)$^{n.s.}$
β (BO-O → BO-O)	**.483 (.061)*****	**.315 (.057)*****	**.514 (.060)*****	**.487 (.072)*****
β (BO-O → ASC)		.052 (.072)$^{n.s.}$		-.041 (.063)$^{n.s.}$
β (ASC → BO-I)		**-.334 (.089)*****		-.098 (.071)$^{n.s.}$
β (ASC → BO-U)		.108 (.096)$^{n.s.}$		.111 (.077)$^{n.s.}$
β (ASC → BO-O)		**-.398 (.077)*****		-.034 (.083)$^{n.s.}$
β (ASC → ASC)		**.713 (.079)*****		**.560 (.074)*****

Note. Model estimated standard errors in brackets. T1 – T4 = waves of measurement. BO-I = boredom intensity. BO-U = boredom due to underchallenge. BO-O = boredom due to overchallenge. ASC = academic self-concept.

$^{n.s.}$ $p \geq .05$. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant coefficients at p <.05 in boldface.

[a] Correlations reported here represent latent correlations for exogenous variables or latent residual correlations for endogenous variables, respectively.

Appendix 3: Supplemental Material to Study 2

In the following, we document the propensity score matching (PSM) procedure applied in the present study (see also Tables A3.1 and A3.2). Additionally, we present result tables for auxiliary analyses (see Tables A3.3 – A3.7).

What is PSM?

Propensity score matching methods originated in research on treatment effects (e.g., Rosenbaum & Rubin, 1983, 1984, 1985) and have become increasingly prevalent over the past years in the social sciences. The main intention behind any PSM-study is a retrospective transformation of field data into quasi-experimental data through balancing. This can be helpful when experiments with randomized groups cannot be conducted, e.g., within an educational system. Due to various ethical reasons, study participants cannot be randomly assigned to attend a special gifted class, for example. Propensity scores (PS) as “conditional probability that expresses how likely a participant is to be assigned or to select the treatment condition given certain observed baseline characteristics” (Thoemmes & Kim, 2011, p. 92) help form the basis for matching participants who are similar in these underlying characteristics but ended up in different groups due to sampling issues. Given a large enough original sample and a sufficient number of matching variables, comparable control groups can be readily attained through PSM. In our case, PSM helped us find students who would be apt to attend a gifted class but did not. Performing analyses on the PSM-subsample of the original field sample, hence strengthens conclusions on the actual (i.e., isolated) treatment effect, in our case attending a special class for gifted students. Therefore, it is imperative to include all available information as matching variables in the PS estimation to rule out possible biases in treatment effects and eliminate alternative explanations to findings from PSM-research.

How was PSM applied?

As stated in the main manuscript, we conducted PSM over two separate steps using two different statistical packages for R Statistics, namely the “twang” (Ridgeway et al., 2015, for details on package functions, see 2014) and “MatchIt” (Ho et al., 2011a, for details on package functions, see 2011b) packages. This two-step approach was necessary because of the differential capabilities of the two packages. “MatchIt”, for instance, only functions correctly, if applied on a full set of data without any missing values on any variable. “Twang”, on the other hand, can handle missing data reliably, treating missing values as separate sources of information for propensity score (PS) estimation. However, “twang” does not feature 1:1 nearest neighbor-matching, which was our preferred matching scheme. Hence, we combined both software packages to achieve the best possible matching result given the incomplete nature of our longitudinal assessment data.

In their review of PSM-studies, Thoemmes and Kim (2011) formulate 14 criteria for information that should be presented when conducting PSM-based research. Since covering these details would overload the main manuscript, we decided to sum up relevant information on the present study in this supplemental material (see Table A3.2). However, not all of the 14 points were applicable to our study. For instance, covering all variables collected throughout the entire [project name] is outside the scope of this study. Therefore, we only included information on variables that were included in the PSM process. Furthermore, details on sample stratification or weighting were not applicable, since we only estimated PS for matching purposes. Please refer to the main manuscript’s method section for a brief summary, and to Tables A3.1 and A3.2 for details on matching variable assessment, as well as a list of relevant details to the matching process.

Table A3.1

Information from Scale Data

Variable	Number of items	Measure (Source)	n_{T1}	M_{T1}	SD_{T1}	Reliability_{T1}[a]	Relative influence on PS estimation (%)
Cognitive Ability (IQ)[b]	140	Cognitive Ability Test for Grades 4-12, short version (KFT 4-12+R; Heller & Perleth, 2000)	919	110.62	12.42	.931	31.25
Social self-concept of acceptance	3	Items originated in the "Development of Adolescents"-project (Fend & Prester, 1986)	1,261	1.61	.80	.770	.00
Social self-concept of assertiveness	3	Items originated in the "Development of Adolescents"-project (Fend & Prester, 1986)	1,271	2.30	.99	.711	1.44
Self-esteem	4	Rosenberg Self-Esteem Scale, revised German version (von Collani & Herzberg, 2003)	1,249	4.10	.67	.749	.68
$ASC_{general}$	3	Self-Descriptions Questionnaire, short version (SDQ II-S; Marsh, 1990b, 2007)	1,276	4.07	.68	.791	.86
ASC_{math}	5	Self-Descriptions Questionnaire, short version (SDQ II-S; Marsh, 1990b, 2007)	1,216	4.12	.79	.865	.77
ASC_{German}	5	Self-Descriptions Questionnaire, short version (SDQ II-S; Marsh, 1990b, 2007)	1,246	3.96	.79	.858	.51
AI_{math}	6	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007)	1,248	3.52	.93	.864	.54
AI_{German}	6	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007)	1,248	3.44	.91	.851	1.03
AP-AM_{math}	4	Achievement Goals Questionnaire, German adaptation (Elliot & Church, 1997; Pekrun et al., 2007)	1,248	2.98	1.05	.792	.01
AP-AM_{German}	4	Achievement Goals Questionnaire, German adaptation (Elliot & Church, 1997; Pekrun et al., 2007)	1,251	3.00	1.05	.819	.60
AV-AM_{math}	4	Achievement Goals Questionnaire, German adaptation (Elliot & Church, 1997; Pekrun et al., 2007)	1,249	3.46	1.09	.838	1.10
AV-AM_{German}	4	Achievement Goals Questionnaire, German adaptation (Elliot & Church, 1997; Pekrun et al., 2007)	1,257	3.46	1.13	.874	.20
CM_{math}	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007)	1,279	3.87	1.01	.725	.71
CM_{German}	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007)	1,272	3.74	1.06	.771	2.08
BO-I_{mat}	2	Achievement Emotions Questionnaire – Mathematics (AEQ-M; Pekrun et al., 2005; see also Pekrun et al., 2003)	1,270	1.73	.92	.708	.00
BO-I_{German}	2	Achievement Emotions Questionnaire – Mathematics (AEQ-M; Pekrun et al., 2005; see also Pekrun et al., 2003)	1,274	1.81	.96	.789	.20
BO-U_{math}	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007, see also 2003)	1,265	2.02	1.00	.705	.55
BO-U_{German}	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007, see also 2003)	1,270	2.00	1.01	.760	.39

BO-O$_{math}$	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007, see also 2003)	1,274	1.62	.86	.761	.00
BO-O$_{German}$	2	Items originated in the Project for the Analysis of Learning and Achievement in Mathematics (PALMA; Pekrun et al., 2007, see also 2003)	1,275	1.56	.84	.797	.00

Note. ASC = Academic self-concept. AI = Academic interest. AP-AM = Approach achievement motivation. AV-AM = Avoidance achievement motivation. CM = Competence motivation. BO-I = Intensity of boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge.

[a] Cronbach's α was used as primary reliability coefficient. For two-item scales, Spearman-Brown coefficients (ρ) were calculated instead (Eisinga et al., 2013).

[b] Sample size and Cronbach's α reported here are based on the 90-minute short version. Additionally, 21 students completed the full version of the KFT 4-12+R (α = .86). Cronbach's α of this subsample, however, is not entirely trustworthy because it was based on a reduced item pool (i = 116 instead of 195), with several item variances yielding a value of zero.

For all measures besides IQ, participants responded on 5-point Likert-type rating scales, ranging from 1 (strongly disagree) to 5 (strongly agree). By replacing the word "mathematics" with "German," German-specific item framing was provided.

Table A3.2

PSM Details as Recommended by Thoemmes & Kim (2011)

Thoemmes & Kim (2011) criteria		Relevant info for present study
1	List of variables collected	n/a (outside of the scope of this study)
2	List of variables that were used to estimate PS	See Table 1 (for scale reliabilities, see Table S1)
3	Method to determine set of covariates used for estimation	Parsimonious model (i.e. all featured variables were included in PS estimation)
4	Inclusion of polynomial or interaction terms	Up to three-way interaction depth
5	Estimation method for PS	Boosted regression
6	Conditioning strategy	Matching
7	Region of common support	Unmatched: PS_{min} = .012, PS_{max} = .988, range of PS = .976; matched: PS_{min} = .025, PS_{max} = .929, range of PS = .904
8	Details on matching scheme	
8.1	Type of matching algorithm	Nearest neighbor, with exact matching on school[a]
8.2	Number of units that were matched with each other	1:1
8.3	Matching with or without replacement	Without replacement
8.4	Caliper width	Caliper = 0.9
9	Details of stratification	n/a (stratification not included in conditioning strategy)
10	Details on weighting	n/a (weighting not included in conditioning strategy)
11	Sample size before and after matching	Unmatched: n_{RC} = 1,471, n_{GC} = 390; matched: $n_{RC,\ matched} = n_{GC,\ matched}$ = 273
12	Standardized differences before and after matching	See Table 1
13	Point estimate of treatment effect and associated *SE*	See Table 4
14	Inclusion of covariates in outcome model	none

Note: PS = propensity score. n/a = not applicable. Caliper width in *SD*-units of estimated PS.

RC = regular classes. GC = special classes for the gifted.

PS were estimated in “twang”, using the “ES.mean”-stopping rule for the average treatment effect on the treated (Ridgeway et al., 2014).

In order to control for context effects, only students from the same school were matched via exact matching (Ho et al., 2011b).

PSM was conducted based on T1 data (for details on waves of measurement, see method section of main manuscript)

Table A3.3

Fit Indexes of Confirmatory Factor Analyses of First-Order Factor Models of Boredom

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	ω_{BO-I}	ω_{BO-U}	ω_{BO-O}
Full sample, RC and GC									
Math									
T1 (*n* = 1,294)	16.691*	6	1.5618	.991	.037	.020	.711	.706	.763
T2 (*n* = 1,307)	57.233**	6	1.4393	.964	.081	.033	.748	.705	.761
T3 (*n* = 1,282)	85.741**	6	1.3372	.956	.102	.051	.811	.743	.789
T4 (*n* = 1,264)	195.861**	6	.8674	.888	.158	.066	.758	.763	.835
German									
T1 (*n* = 1,290)	14.518*	6	2.1408	.993	.033	.015	.796	.760	.800
T2 (*n* = 1,302)	2.684	6	.8473	1.000	.000	.007	.786	.717	.827
T3 (*n* = 1,277)	5.833	6	.4421	1.000	.000	.014	.822	.569	.785
T4 (*n* = 1,253)	32.608**	6	1.2121	.982	.059	.026	.821	.768	.752
Full sample, RC only									
Math									
T1 (*n* = 1,025)	11.365	6	1.6020	.994	.030	.016	.703	.703	.770
T2 (*n* = 1,033)	50.493**	6	1.4001	.961	.085	.034	.740	.695	.756
T3 (*n* = 1,001)	60.855**	6	1.3245	.961	.096	.050	.807	.727	.791
T4 (*n* = 987)	137.003**	6	.8552	.890	.149	.059	.770	.749	.825
German									
T1 (*n* =1,021)	11.791	6	2.0368	.993	.031	.014	.779	.744	.787
T2 (*n* = 1,029)	2.582	6	2.1508	1.000	.000	.007	.767	.723	.840
T3 (*n* = 996)	3.677	6	.7203	1.000	.000	.013	.818	.715	.780
T4 (*n* = 977)	22.522**	6	1.2909	.986	.053	.025	.819	.755	.513
Full sample, GC only									
Math									
T1 (*n* = 269)	20.621**	6	1.0883	.950	.095	.043	.751	.720	.747
T2 (*n* = 274)	10.424	6	1.1545	.987	.052	.033	.791	.744	.805
T3 (*n* = 281)	44.669**	6	.8505	.922	.151	.062	.825	.794	.859
T4 (*n* = 277)	51.244**	6	1.0130	.918	.165	.082	.681	.802	.871
German									
T1 (*n* = 269)	7.713	6	2.1362	.995	.033	.031	.850	.817	.857
T2 (*n* = 273)	9.145	6	1.0091	.991	.044	.017	.857	.695	.768
T3 (*n* = 281)	4.764	6	.5744	1.000	.000	.018	.837	.651	.802
T4 (*n* = 276)	16.050*	6	.8684	.980	.078	.033	.830	.806	.827
Matched sample, RC and GC									
Math									
T1 (*n* = 370)	14.261*	6	.9233	.979	.061	.029	.713	.751	.740
T2 (*n* = 374)	19.067**	6	1.2480	.963	.076	.039	.752	.704	.788
T3 (*n* = 375)	68.685**	6	.8851	.887	.167	.066	.774	.801	.768
T4 (*n* = 369)	59.243**	6	.9091	.927	.155	.066	.749	.792	.879
German									
T1 (*n* = 370)	21.148**	6	.7147	.972	.083	.030	.861	.837	.870
T2 (*n* = 373)	9.935	6	1.1349	.991	.042	.020	.846	.693	.820
T3 (*n* = 374)	6.005	6	1.3472	1.000	.001	.020	.844	.693	.758
T4 (*n* = 363)	45.326**	6	.6765	.938	.134	.045	.811	.759	.773

Note. BO-I = Intensity of boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge.

RC = regular classes. GC = special classes for the gifted.

* $p < .05$. ** $p < .01$.

McDonald's ω was calculated as $\frac{(\sum_{i=1}^{p} \lambda_{ij})^2}{(\sum_{i=1}^{p} \lambda_{ij})^2 + \sum_{i=1}^{p} e_i}$, where λ_{ij} is the standardized factor loading of item i on factor j, and e_{ij} is the standardized item residual for item i regarding factor j (Brunner et al., 2012; see also McDonald, 1999).

Table A3.4

Investigation of Measurement Invariance of the First-Order Factor Model of Boredom Over Time

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Full sample, RC and GC												
Math (*n* = 1,802)												
configural	439.943**	150	1.2163	.966	.033	.042						
metric	465.133**	159	1.2299	.964	.033	.044	configural vs. metric	25.378**	9	-.002	.000	+.002
scalar	486.939**	168	1.2344	.963	.032	.044	metric vs. scalar	22.080**	9	-.001	-.001	.000
German (*n* = 1,797)												
configural	213.968**	150	1.3001	.992	.015	.024						
metric	222.818**	159	1.3212	.992	.015	.024	configural vs. metric	9.688	9	.000	.000	.000
scalar	239.580**	168	1.3205	.991	.015	.025	metric vs. scalar	16.801	9	-.001	.000	+.001
Full sample, RC only												
Math (*n* = 1421)												
configural	357.342**	150	1.2051	.969	.031	.041						
metric	377.325**	159	1.2167	.967	.031	.043	configural vs. metric	20.183*	9	-.002	.000	+.002
scalar	406.665**	168	1.2203	.964	.032	.043	metric vs. scalar	28.945**	9	-.003	+.001	.000
German (*n* = 1417)												
configural	191.134*	150	1.2719	.993	.014	.025						
metric	198.838*	159	1.2912	.994	.013	.026	configural vs. metric	8.455	9	+.001	-.001	+.001
scalar	214.656**	168	1.2929	.993	.014	.026	metric vs. scalar	15.714	9	-.001	+.001	.000
Full sample, GC only												
Math (*n* = 381)												
configural	287.850**	150	1.0015	.941	.049	.064						
metric	291.594**	159	1.0299	.944	.047	.064	configural vs. metric	8.003	9	+.003	-.002	.000
scalar	299.477**	168	1.0462	.944	.045	.065	metric vs. scalar	9.744	9	.000	-.002	+.001
German (*n* = 380)												
configural	213.239**	150	1.0768	.971	.033	.043						
metric	222.345**	159	1.1246	.971	.032	.044	configural vs. metric	10.635	9	.000	-.001	+.001
scalar	227.892**	168	1.1297	.973	.031	.044	metric vs. scalar	6.067	9	+.002	-.001	.000
Matched sample												
Math (*n* = 528)												

configural	270.860**	150	1.0597	.957	.039	.055						
metric	270.646**	159	1.0844	.960	.036	.056	configural vs. metric	4.317	9	+.003	-.003	+.001
scalar	279.348**	168	1.0822	.960	.035	.056	metric vs. scalar	8.456	9	.000	-.001	.000
German (*n* = 525)												
configural	248.204**	150	1.0536	.968	.035	.039						
metric	249.296**	159	1.0956	.970	.033	.038	configural vs. metric	6.472	9	+.002	-.002	-.001
scalar	262.678**	168	1.0943	.969	.033	.039	metric vs. scalar	13.366	9	-.001	.000	+.001

Note. RC = regular classes. GC = special classes for the gifted. * $p < .05$. ** $p < .01$.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

Table A3.5

Investigation of Measurement Invariance of the First-Order Factor Model of Boredom Across Class Types in the Full Sample

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
T1												
Math (n_{RC} = 1,025; n_{GC} = 269)												
configural	27.789^{**}	12	1.4628	.987	.045	.024						
metric	31.797^{**}	15	1.4060	.986	.042	.027	configural vs. metric	3.442	3	-.001	-.003	+.003
scalar	37.852^{**}	18	1.3356	.983	.041	.030	metric vs. scalar	5.946	3	-.003	-.001	+.003
German (n_{RC} = 1,021; n_{GC} = 269)												
configural	21.338^{*}	12	1.8976	.993	.035	.019						
metric	28.478^{*}	15	1.7940	.990	.037	.026	configural vs. metric	7.682	3	-.003	+.002	+.005
scalar	31.880^{*}	18	1.6572	.989	.035	.025	metric vs. scalar	1.790	3	-.001	-.002	-.001
T2												
Math (n_{RC} = 1,033; n_{GC} = 274)												
configural	68.219^{**}	12	1.2127	.957	.085	.034						
metric	67.414^{**}	15	1.2447	.960	.073	.034	configural vs. metric	0.860	3	+.003	-.012	.000
scalar	76.395^{**}	18	1.1962	.955	.070	.035	metric vs. scalar	7.836^{*}	3	-.005	-.003	+.001
German (n_{RC} = 1,029; n_{GC} = 273)												
configural	9.740	12	1.5176	1.000	.000	.010						
metric	12.668	15	1.5057	1.000	.000	.016	configural vs. metric	2.944	3	.000	.000	+.006
scalar	17.586	18	1.4128	1.000	.000	.015	metric vs. scalar	6.086	3	.000	.000	-.001
T3												
Math[a] (n_{RC} = 1,001; n_{GC} = 281)												
configural	116.116^{**}	13	1.0309	.935	.111	.052						
metric	108.859^{**}	16	1.1689	.942	.095	.055	configural vs. metric	4.268	3	+.007	-.016	+.003
scalar	114.638^{**}	19	1.1431	.940	.089	.056	metric vs. scalar	3.777	3	-.002	-.006	+.001
German (n_{RC} = 996; n_{GC} = 281)												
configural	9.555	12	1.4059	1.000	.000	.014						
metric	14.134	15	1.4366	1.000	.000	.021	configural vs. metric	4.407	3	.000	.000	+.007
scalar	17.704	18	1.3670	1.000	.000	.020	metric vs. scalar	3.824	3	.000	.000	-.001
T4												
Math (n_{RC} = 987; n_{GC} = 277)												
configural	194.673^{**}	12	0.8685	.889	.155	.065						

metric	146.872**	15	1.1742	.920	.118	.066	configural vs. metric	0.703	3	+.031	+.003	+.001
scalar	168.098**	18	1.1417	.909	.115	.068	metric vs. scalar	19.874**	3	-.011	-.003	+.002
partially scalar[b]	150.431**	17	1.1771	.919	.111	.066	metric vs. partially scalar[b]	3.850	2	-.001	-.007	.000
German (n_{RC} = 977; n_{GC} = 276)												
configural	32.805**	12	1.3111	.985	.053	.027						
metric	34.423**	15	1.3366	.986	.045	.028	configural vs. metric	2.085	3	+.001	-.008	+.001
scalar	37.413**	18	1.2842	.986	.041	.028	metric vs. scalar	1.991	3	.000	-.004	.000

Note. RC = regular classes. GC = special classes for the gifted. * $p < .05$. ** $p < .01$.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

[a] Residual variance for item bom_o2_3 was set to 0 within special classes for the gifted after showing a nonsignificant negative value in the original configural model (Chen et al., 2001).

[b] Intercept for item bom_u2_4 was allowed to vary across class types upon revision of model modification indexes.

Table A3.6

Investigation of Measurement Invariance of the First-Order Factor Model of Boredom Across Class Types in the Matched Sample

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
T1												
Math (n_{RC} = 182; n_{GC} = 188)												
configural	15.848	12	1.1031	.990	.042	.036						
metric	15.639	15	1.1558	.998	.015	.037	configural vs. metric	0.434	3	+.008	-.027	+.001
scalar	20.328	18	1.1360	.994	.026	.041	metric vs. scalar	4.838	3	-.004	+.011	+.004
German[a] (n_{RC} = 182; n_{GC} = 188)												
configural	20.597	14	1.1788	.989	.050	.033						
metric	38.838**	17	1.3527	.962	.083	.082	configural vs. metric	13.056**	3	-.027	+.033	+.049
partially metric[b]	21.085	16	1.2758	.991	.041	.037	configural vs. partially metric[b]	1.341	2	+.002	-.009	+.004
partially scalar[b]	24.872	19	1.2205	.990	.041	.038	partially metric[b] vs. partially scalar[b]	3.734	3	-.001	.000	+.001
T2												
Math (n_{RC} = 196; n_{GC} = 178)												
configural	28.068**	12	0.9804	.958	.085	.040						
metric	23.242	15	1.2015	.978	.054	.040	configural vs. metric	0.195	3	+.020	-.031	.000
scalar	24.588	18	1.1675	.983	.044	.041	metric vs. scalar	0.783	3	+.005	-.010	+.001
German (n_{RC} = 196; n_{GC} = 177)												
configural	16.502	12	1.1551	.990	.045	.023						
metric	16.543	15	1.2184	.997	.023	.026	configural vs. metric	0.744	3	+.007	-.022	+.003
scalar	20.365	18	1.1771	.995	.027	.027	metric vs. scalar	3.931	3	-.002	+.004	+.001
T3												
Math[c] (n_{RC} = 189; n_{GC} = 186)												
configural	61.933**	13	1.0283	.900	.142	.067						
metric	58.792**	16	1.1412	.913	.119	.071	configural vs. metric	2.090	3	+.013	-.023	+.004
scalar	64.365**	19	1.1116	.908	.113	.074	metric vs. scalar	4.671	3	-.005	-.006	+.003
German (n_{RC} = 188; n_{GC} = 186)												
configural	16.476	12	1.2865	.990	.045	.026						
metric	22.410	15	1.2721	.983	.051	.041	configural vs. metric	6.020	3	-.007	+.006	+.015
scalar	24.133	18	1.2185	.986	.043	.041	metric vs. scalar	0.945	3	+.003	-.008	.000
T4												
Math[d] (n_{RC} = 190; n_{GC} = 179)												

configural	57.385**	13	1.0057	.923	.136	.072						
metric	50.232**	16	1.1871	.941	.108	.071	configural vs. metric	.972	3	+.018	-.028	-.001
scalar	59.934**	19	1.1678	.929	.108	.072	metric vs. scalar	9.729*	3	-.012	.000	+.001
partially scalar[e]	51.994**	18	1.1695	.941	.101	.070	metric vs. partially scalar[e]	1.144	2	.000	-.007	-.001
German[f] (n_{RC} = 185; n_{GC} = 178)												
configural	25.418*	14	1.1850	.974	.067	.049						
metric	35.609**	17	1.1924	.957	.078	.065	configural vs. metric	10.058*	3	-.017	+.011	-.016
partially metric[g]	31.408*	16	1.1835	.965	.073	.059	configural vs. partially metric[g]	6.011*	2	-.009	+.006	-.010
partially scalar[g]	34.251*	19	1.1361	.965	.067	.063	partially metric[g] vs. partially scalar[g]	1.971	3	.000	-.006	+.004

Note. RC = regular classes. GC = special classes for the gifted. * $p < .05$. ** $p < .01$.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

[a] Residual variance for item bog_o2_1 was set to 0 within both class types after showing nonsignificant negative values in the original configural model (Chen et al., 2001).

[b] Factor loading for item bog_o2_1 was allowed to vary across class types upon revision of model modification indexes.

[c] Residual variance for item bom_o2_3 was set to 0 within special classes for the gifted after showing a nonsignificant negative value in the original configural model (Chen et al., 2001).

[d] Residual variance for item bom_u1_4 was set to 0 within special classes for the gifted after showing a nonsignificant negative value in the original configural model (Chen et al., 2001).

[e] Intercept for item bom_u2_4 was allowed to vary across class types.

[f] Residual variances for items bog2_4 and bog_u2_4 were set to 0 within regular classes after showing nonsignificant negative values in the original configural model (Chen et al., 2001).

[g] Factor loading for item bog2_4 was allowed to vary across class types.

Table A3.7

Mean Factor Scores and Standard Deviations for Latent Variables (Full Sample and by Class Type Subsamples)

	Math											
	Full sample, RC and GC (*n* = 1,802)				Full sample, RC only (*n* = 1,421)				Full sample, GC only (*n* = 381)			
Boredom	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
BO-I	1.76 (.62)	1.81 (.65)	2.08 (.82)	2.48 (.74)	1.78 (.62)	1.85 (.67)	2.09 (.81)	2.56 (.74)	1.69 (.62)	1.65 (.63)	2.02 (.88)	2.19 (.69)
BO-U	2.05 (.67)	2.06 (.66)	2.12 (.70)	2.25 (.71)	2.04 (.65)	2.10 (.65)	2.09 (.67)	2.24 (.69)	2.06 (.75)	1.93 (.71)	2.23 (.82)	2.29 (.81)
BO-O	1.64 (.60)	1.75 (.64)	1.94 (.71)	2.27 (.87)	1.67 (.61)	1.80 (.65)	1.98 (.72)	2.36 (.86)	1.53 (.54)	1.58 (.61)	1.76 (.70)	1.98 (.88)
	German											
	Full sample, RC an GC (*n* = 1,797)				Full sample, RC only (*n* = 1,471)				Full sample, GC only (*n* = 380)			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
BO-I	1.82 (.69)	1.96 (.73)	2.24 (.83)	2.45 (.81)	1.82 (.66)	1.98 (.72)	2.26 (.82)	2.47 (.81)	1.84 (.79)	1.89 (.78)	2.20 (.87)	2.38 (.85)
BO-U	2.01 (.70)	2.05 (.66)	2.25 (.66)	2.38 (.66)	2.00 (.68)	2.07 (.67)	2.26 (.66)	2.39 (.64)	2.05 (.82)	1.93 (.60)	2.25 (.69)	2.34 (.73)
BO-O	1.58 (.60)	1.70 (.69)	1.75 (.63)	1.76 (.59)	1.62 (.61)	1.74 (.72)	1.79 (.64)	1.79 (.58)	1.41 (.57)	1.54 (.57)	1.62 (.59)	1.66 (.63)

Note. RC = regular classes. GC = special classes for the gifted. *BO-I = Intensity of Boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge.*

Factor scores had the same metric as item indicators due to effects coding (Little et al., 2006), spanning a possible range of 1 – 5.

Appendix 4: Supplemental Material to Study 3

Table A4.1

Item Wording for Subject-Specific Boredom and Appraisal Measurement

Item	Text
	BO-I
1	"I find [mathematics / German] to be boring"
2	"I find it hard to stay awake during [mathematics / German] class out of sheer boredom"
	BO-U
1	"When I'm bored in [mathematics / German] class, this is because the subject matter is so easy"
2	"When I'm bored in [mathematics / German] class, this is because the teacher goes on about trivial points"
	BO-O
1	"When I'm bored in [mathematics / German] class, this is because I cannot follow the teacher"
2	"When I'm bored in [mathematics / German] class, this is because the [mathematics / German] subject matter is too difficult for me"
	AI
1	"I'm interested in [mathematics / German]"
2	"[Mathematics / German] class is fun to me"
3	"Engaging in [mathematics / German] is among my favorite activities"
4	"Often times after class, I'm already curious about the next [mathematics / German]-lesson"
	ASC
1	"I get good grades in [mathematics / German]"
2	"[Mathematics / German] is one of my best subjects"
3	"I've always been good at [mathematics / German]"
4	"I learn fast in [mathematics / German]"

Note. Item wording translated from original German wording. Item responses were given on 5-point rating scales, ranging from 1 – *strongly disagree* to 5 – *strongly agree*. BO-I = Intensity of boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge. AI = academic interest. ASC = academic self-concept.

Table A4.2

Fit Indexes for Confirmatory Factor Analyses (CFA) of 3F-model of Boredom

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	ω_{BO-I}	ω_{BO-U}	ω_{BO-O}
					Entire sample				
Math									
T1 (*n* = 1,025)	11.290	6	1.6127	.994	.029 [.000-.055]	.016	.703	.703	.770
T2 (*n* = 1,032)	49.932***	6	1.4093	.960	.084 [.064-.107]	.034	.739	.695	.755
T3 (*n* = 999)	60.944***	6	1.3267	.961	.096 [.075-.118]	.050	.808	.726	.791
T4 (*n* = 987)	136.164***	6	.8605	.892	.148 [.127-.170]	.059	.770	.749	.825
German									
T1 (*n* = 1,021)	11.713	6	2.0503	.994	.031 [.000-.056]	.014	.779	.744	.787
T2 (*n* = 1,028)	2.571	6	2.1593	1.000	.000 [.000-.022]	.007	.767	.724	.840
T3 (*n* = 994)	3.597	6	1.8273	1.000	.000 [.000-.030]	.013	.818	.714	.779
T4 (*n* = 977)	22.384**	6	1.2988	.986	.053 [.031-.077]	.025	.819	.755	.734
					Males only				
Math									
T1 (*n* = 526)	7.839	6	1.5817	.996	.024 [.000-.065]	.020	.672	.687	.768
T2 (*n* = 524)	31.751***	6	1.3202	.953	.091 [.061-.123]	.032	.760	.695	.717
T3 (*n* = 510)	23.976***	6	1.3826	.971	.077 [.046-.110]	.045	.797	.688	.754
T4 (*n* = 485)	56.415***	6	.9638	.896	.132 [.102-.164]	.063	.738	.726	.766
German									
T1 (*n* = 523)	6.394	6	2.2261	.999	.011 [.000-.059]	.016	.776	.743	.788
T2 (*n* = 521)	2.233	6	2.3020	1.000	.000 [.000-.025]	.013	.726	.713	.854
T3 (*n* = 505)	4.587	6	1.7004	1.000	.000 [.000-.050]	.014	.798	.700	.720
T4 (*n* = 476)	15.716*	6	1.1317	.979	.058 [.024 -.094]	.030	.818	.728	.704
					Females only				
Math									
T1 (*n* = 499)	5.858	6	1.4911	1.000	.000 [.000-.057]	.016	.744	.722	.771
T2 (*n* = 508)	24.192***	6	1.0564	.970	.077 [.047-.111]	.034	.735	.677	.792
T3 (*n* = 489)	38.224***	6	1.2452	.958	.105 [.075-.138]	.051	.826	.778	.839
T4 (*n* = 502)	85.021***	6	.6625	.904	.162 [.132-.193]	.051	.810	.780	.877
German									
T1 (*n* = 498)	7.730	6	1.3298	.997	.024 [.000-.066]	.013	.774	.744	.807
T2 (*n* = 507)	7.093	6	1.4538	.998	.019 [.000-.063]	.013	.819	.729	.823
T3 (*n* = 489)	17.462**	6	.8744	.985	.063 [.030-.098]	.026	.839	.726	.823
T4 (*n* = 501)[a]	16.904*	7	1.3800	.987	.053 [.021-.086]	.026	.824	.796	.817

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

SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019).

90%-CI for RMSEA are provided in brackets.

BO-I = Intensity of boredom. BO-U = Boredom due to underchallenge. BO-O = Boredom due to overchallenge.

McDonald's ω was calculated as $\frac{(\sum_{i=1}^{p}\lambda_{ij})^2}{(\sum_{i=1}^{p}\lambda_{ij})^2+\sum_{i=1}^{p}e_i}$, where λ_{ij} is the standardized factor loading of item i on factor j, and e_{ij} is the standardized item residual for item i regarding factor j (Brunner et al., 2012; see also McDonald, 1999)

[a] Residual variance for item "When I'm bored in German class, this is because I cannot follow the teacher" was set to 0 after showing a nonsignificant negative value of -.071 in the original model (Chen et al., 2001).

Table A4.3

Fit Indexes of CFA for Academic Self-concept and Academic Interest at T1

Model	*n*	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Factor determinacy[a]	ω
				Entire sample					
Mathematics									
ASC	1,025	2.759	2	1.4512	.999	.019 [.000-.068]	.006	.939	.877
AI	1,025	17.441^{***}	2	1.7313	.984	.087 [.052-.126]	.017	.937	.872
German									
ASC	1,021	1.985	2	2.5274	1.000	.000 [.000-.062]	.006	.944	.888
AI	1,023	9.671^{**}	2	1.8346	.993	.061 [.027-.102]	.016	.930	.855
				Males only					
Mathematics									
ASC	526	2.738	2	1.7174	.999	.026 [.000-.095]	.010	.935	.869
AI	526	6.672^{*}	2	2.5218	.990	.067 [.015-.126]	.019	.930	.857
German									
ASC	523	1.395	2	2.5910	1.000	.000 [.000-.078]	.008	.937	.874
AI	524	10.738^{**}	2	1.6963	.980	.091 [.043-.148]	.024	.927	.845
				Females only					
Mathematics									
ASC	499	2.843	2	0.9083	.999	.029 [.000-.098]	.007	.937	.873
AI	499	11.989^{**}	2	1.1055	.983	.100 [.051-.158]	.015	.942	.882
German									
ASC	498	0.665	2	2.0816	1.000	.000 [.000-.064]	.004	.951	.900
AI	499	1.519	2	1.4295	1.000	.000 [.000-.082]	.007	.933	.863

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

SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019).

90%-CI for RMSEA are provided in brackets.

ASC = academic self-concept. AI = academic interest.

McDonald's ω was calculated as $\frac{(\sum_{i=1}^{p} \lambda_{ij})^2}{(\sum_{i=1}^{p} \lambda_{ij})^2 + \sum_{i=1}^{p} e_i}$, where λ_{ij} is the standardized factor loading of item i on factor j, and e_{ij} is the standardized item residual for item i regarding factor j (Brunner et al., 2012; see also McDonald, 1999).

[a] Factor determinacies based on respective complete-data patterns.

Table A4.4

Investigation of Measurement Invariance of the 3F-model of Boredom Across Genders

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
T1												
Mathematics ($n_{♂}$ = 526, $n_{♀}$ = 499)												
configural	14.708	12	1.4369	.997	.021 [.000-.052]	.018						
metric	21.179	15	1.4686	.993	.028 [.000-.054]	.027	configural vs. metric	6.249	3	-.004	+.007	+.009
scalar	31.352*	18	1.3919	.985	.038 [.013-.060]	.026	metric vs. scalar	12.431**	3	-.008	+.010	-.001
German ($n_{♂}$ = 523, $n_{♀}$ = 498)												
configural	15.185	12	1.6144	.997	.023 [.000-.053]	.015						
metric	23.838	15	1.5779	.991	.034 [.000-.059]	.026	configural vs. metric	9.148*	3	-.006	+.011	+.011
scalar	25.844	18	1.4790	.992	.029 [.000-.053]	.027	metric vs. scalar	0.619	3	+.001	-.005	+.001
T2												
Mathematics ($n_{♂}$ = 524, $n_{♀}$ = 508)												
configural	56.017***	12	1.2045	.957	.084 [.063-.107]	.033						
metric	61.316***	15	1.2968	.955	.077 [.058-.098]	.040	configural vs. metric	7.228	3	-.002	-.007	+.007
scalar	94.940***	18	1.2253	.925	.091 [.073-.109]	.044	metric vs. scalar	42.424***	3	-.300	+.014	+.004
partially scalar[a]	66.825***	17	1.2581	.951	.075 [.057-.095]	.040	metric vs. partially scalar	4.709	2	-.004	-.002	.000
German ($n_{♂}$ = 521, $n_{♀}$ = 507)												
configural	8.938	12	1.7287	1.000	.000 [.000-.034]	.013						
metric	18.558	15	1.7223	.997	.021 [.000-.049]	.031	configural vs. metric	9.732*	3	-.003	+.021	+.018
scalar	20.831	18	1.6031	.997	.017 [.000-.045]	.030	metric vs. scalar	1.422	3	.000	-.004	-.001
T3												
Mathematics ($n_{♂}$ = 510, $n_{♀}$ = 489)												
configural	74.423***	12	1.0850	.951	.102 [.081-.125]	.048						
metric	78.173***	15	1.1697	.950	.092 [.072-.112]	.053	configural vs. metric	7.087	3	-.001	-.010	+.005
scalar	118.194***	18	1.1172	.921	.106 [.088-.124]	.059	metric vs. scalar	47.511***	3	-.029	+.014	+.006
partially scalar[b]	82.951***	16	1.1511	.947	.092 [.073-.111]	.053	metric vs. partially scalar	4.639*	2	-.003	.000	.000
German ($n_{♂}$ = 505, $n_{♀}$ = 489)												
configural	17.177	12	1.3430	.995	.029 [.000-.058]	.021						
metric	17.906	15	1.3516	.997	.020 [.000-.049]	.021	configural vs. metric	0.818	3	+.002	-.009	.000
scalar	20.065	18	1.3037	.998	.015 [.000-.044]	.023	metric vs. scalar	1.839	3	+.001	-.005	+.002
T4												

Mathematics ($n_♂$ = 485, $n_♀$ = 502)													
configural	132.505***	12	0.8354	.916	.143 [.121-.165]	.058							
metric	118.293***	15	1.0224	.928	.118 [.099-.138]	.059	configural vs. metric	5.789	3	+.012	-.025	+.001
scalar	127.893***	18	1.0424	.923	.111 [.094-.130]	.062	metric vs. scalar	10.831*	3	-.005	-.007	+.003
German ($n_♂$ = 476, $n_♀$ = 501)													
configural	33.539***	12	1.2188	.981	.061 [.037-.085]	.029							
metric	40.424***	15	1.2547	.977	.059 [.037-.081]	.034	configural vs. metric	7.039	3	-.004	-.002	+.005
scalar	48.290***	18	1.2093	.973	.059 [.039-.079]	.033	metric vs. scalar	7.815	3	-.004	.000	-.001

Note. ♂ = male. ♀ = female. * $p < .05$. ** $p < .01$. *** $p < .001$.

SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019).

90%-CI for RMSEA are provided in brackets.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

[a] Intercept for item "I find it hard to stay awake during mathematics class out of sheer boredom" was allowed to vary across sexes upon revision of model modification indexes.

[b] Intercepts for items "When I'm bored in mathematics class, this is because the teacher goes on about trivial points" and "I find it hard to stay awake during mathematics class out of sheer boredom" were allowed to vary across sexes upon revision of model modification indexes.

Table A4.5

Investigation of Measurement Invariance of Appraisals Across Genders at T1

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Factor determinacy[a]		Model comparison					
							♂	♀	Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Mathematics														
ASC ($n_♂$ = 526, $n_♀$ = 499)														
configural	5.081	4	1.4336	.999	.023 [.000-.074]	.009	.935	.937						
metric	8.287	7	1.3685	.999	.019 [.000-.060]	.033	.935	.936	configural vs. metric	3.165	3	.000	-.004	+.024
scalar	20.900*	10	1.2944	.989	.046 [.017-.074]	.055	.935	.936	metric vs. scalar	14.010**	3	-.010	+.027	+.022
partially scalar[b]	14.242	9	1.3186	.995	.034 [.000-.065]	.045	.935	.937	metric vs. partially scalar	6.503*	2	-.004	+.015	+.012
AI ($n_♂$ = 526, $n_♀$ = 499)														
configural	20.761***	4	1.4488	.987	.090 [.054-.131]	.017	.930	.942						
metric	26.704***	7	1.2483	.984	.074 [.046-.105]	.031	.930	.942	configural vs. metric	3.319	3	-.003	-.016	+.014
scalar	42.320***	10	1.1912	.974	.079 [.056-.105]	.041	.930	.941	metric vs. scalar	16.141**	3	-.010	-.005	+.010
partially scalar[c]	35.688***	9	1.2008	.979	.076 [.051-.103]	.033	.930	.942	metric vs. partially scalar	9.202*	2	-.005	+.002	+.002
German														
ASC ($n_♂$ = 523, $n_♀$ = 498)														
configural	2.343	4	2.1338	1.000	.000 [.000-.052]	.007	.937	.951						
metric	4.283	7	1.7188	1.000	.000 [.000-.039]	.023	.937	.951	configural vs. metric	2.027	3	.000	.000	+.016
scalar	6.256	10	1.5051	1.000	.000 [.000-.032]	.023	.937	.951	metric vs. scalar	2.041	3	.000	.000	.000
AI ($n_♂$ = 524, $n_♀$ = 499)														
configural	15.277**	4	1.3345	.989	.074 [.037-.115]	.018	.927	.933						
metric	17.367*	7	1.2322	.990	.054 [.022-.086]	.021	.926	.933	configural vs. metric	0.924	3	+.001	-.020	+.003
scalar	23.952**	10	1.1673	.986	.052 [.025-.079]	.024	.926	.933	metric vs. scalar	6.457	3	-.004	-.002	+.003

Note. ♂ = male. ♀ = female. * $p < .05$. ** $p < .01$. *** $p < .001$.

SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019).

ASC = Academic Self-concept.

90%-CI for RMSEA are provided in brackets.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

[a] Factor determinacies based on respective complete-data patterns.

[b] Intercept for item “I get good grades in mathematics” was allowed to vary across genders.

[c] Intercept for item “Mathematics class is fun to me” was allowed to vary across genders.

Table A4.6

Investigation of Measurement Invariance of the 3F-model of Boredom Over Time

Model	χ^2	*df*	SCF	CFI	RMSEA	SRMR	Model comparison					
							Compare	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Math (n = 1418)												
configural	355.097***	150	1.2122	.969	.031 [.027-.035]	.041						
metric	374.785***	159	1.2242	.967	.031 [.027-.035]	.043	configural vs. metric	19.915*	9	-.002	.000	-.002
scalar	404.000***	168	1.2278	.964	.031 [.028-.035]	.043	metric vs. scalar	28.821***	9	-.003	.000	.000
German (n = 1414)												
configural	189.804*	150	1.2794	.994	.014 [.006-.019]	.025						
metric	197.280*	159	1.2988	.994	.013 [.005-.019]	.025	configural vs. metric	8.256	9	.000	-.001	.000
scalar	213.114*	168	1.3006	.993	.014 [.007-.019]	.026	metric vs. scalar	15.723	9	-.001	+.001	+.001

Note. * $p < .05$. *** $p < .001$.

SCF = Scaling correction factor for MLR estimator in Mplus (Muthén & Muthén, 1998-2019).

90%-CI for RMSEA are provided in brackets.

$\Delta\chi^2$-values were calculated using Satorra & Bentlers' (2010) correction formula for MLR.

Appendix 5: American Psychological Association Licence Terms and Conditions

AMERICAN PSYCHOLOGICAL ASSOCIATION LICENSE
TERMS AND CONDITIONS

Aug 17, 2023

This Agreement between Mr. Markus Feuchter ("You") and American Psychological Association ("American Psychological Association") consists of your license details and the terms and conditions provided by American Psychological Association and Copyright Clearance Center.

License Number	5611260329828
License date	Aug 17, 2023
Licensed Content Publisher	American Psychological Association
Licensed Content Publication	Journal of Educational Psychology
Licensed Content Title	Reducing boredom in gifted education—Evaluating the effects of full-time ability grouping.
Licensed copyright line	Copyright © 2022, American Psychological Association
Licensed Content Author	Feuchter, Markus D.; Preckel, Franzis
Licensed Content Date	Aug 1, 2022
Licensed Content Volume	114
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	Academic institution
Format	Print, Electronic

Portion	page
Number of pages requested	17
Page range of portion	1477-1493
Rights for	Main product and any product related to main product
Duration of use	life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
Title	Reducing boredom in gifted education—Evaluating the effects of full-time ability grouping
Institution name	University of Trier
Expected presentation date	Sep 2023
Portions	1477-1493
Requestor Location	Mr. Markus Feuchter Liebfrauenstraße 5 Trier, 54290 Germany Attn: Mr. Markus Feuchter

Total 0.00 EUR

Terms and Conditions

RightsLink Plus Terms and Conditions
American Psychological Association

The publisher for this copyrighted material is the American Psychological Association ("APA"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by the Copyright Clearance Center, Inc. ("CCC") at the time that you opened your CCC account.

APA hereby grants you a non-exclusive license to use this material, subject to the conditions below.

1. Limited License.
   Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed with 12 months from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to a period of 12 months.

2. Credit Lines and Translation Disclaimer.
   The reproduced material must include one of the following credit lines:

   For English language content: Copyright © [year] by [Copyright Owner]. Reproduced [or Adapted] with permission. [Full bibliographic citation.]

   For translated language content: This material originally appeared in English as [full bibliographic citation]. Copyright © [year] by [Copyright Owner]. Translated and reproduced [or adapted] with permission. [Copyright Owner] is not responsible for the quality or accuracy of this translation. This translation cannot be reproduced or distributed further without prior written permission.

3. Author Permission.
   You must obtain the author's (or, in case of multiple authorship, one author's) permission. A good faith effort must be made to locate the author and obtain the author's permission. Author permission is not required for the reuse of small excepts or 1 to 3 figures or 1 to 3 tables; or for course pack or electronic reserve use.

4. Geographic Rights.
   Licenses may be exercised anywhere in the world.

5. Book Edition Limit.
   If the APA content will be reproduced in a book, then permission is limited to the life of the current edition.

6. Online Use Limitations.
   APA full-text articles or book chapters may only be posted on a secure and restricted web site. The credit line specified in Section 2 shall appear on the first screen on

which the APA content appears and shall be amended to include the following additional wording: No further reproduction or distribution is permitted without written permission from the American Psychological Association.

7. Reservation of Rights.
APA reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent on Payment.
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by APA or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and APA reserves the right to take any and all action to protect its copyright in the materials. "$0.00" fee means that APA waived the fee.

9. License Revocation.
APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to CCC for denied permissions.

10. Warranties.
APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original to APA and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

11. Indemnity.
You hereby indemnify and agree to hold harmless APA and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. No Transfer of License.
This license is personal to you, but may be assigned or transferred by you to your employer or a business associate if you give prompt written notice of the assignment or transfer to the APA. No such assignment or transfer shall relieve you of the obligation to pay the designated license fee on a timely basis (although payment by the identified assignee can fulfill your obligation).

13. No Amendment except in Writing.

This license may not be amended except in a writing signed by both parties (or, in the case of APA, by CCC on APA's behalf).

14. Objection to Contrary Terms.
APA hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and APA (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

15. Jurisdiction.
This license transaction shall be governed by and construed in accordance with the laws of the District of Columbia. You hereby agree to submit to the jurisdiction of the federal courts located in the District of Columbia for purposes of resolving any disputes that may arise in connection with this licensing transaction.

Other Conditions:

v1.2

Questions? customercare@copyright.com.

References (Appendices 2-4)

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A Tutorial on Hierarchically Structured Constructs: Hierarchically Structured Constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Chen, F., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper Solutions in Structural Equation Models: Causes, Consequences, and Strategies. *Sociological Methods and Research*, *29*(4), 468–508. https://doi.org/10.1177/0049124101029004003

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, *72*(1), 218–232. https://doi.org/10.1037/0022-3514.72.1.218

Fend, H., & Prester, H. (1986). *Dokumentation der Skalen des Projekts "Entwicklung im Jugendalter"* [Scale documentation of the project "development of adolescents"], [Unpublished project codebook]. Sozialwissenschaftliche Fakultät der Universität Konstanz.

Heller, K., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. Bis 12. Klassen* [Cognitive abilities test of students from 4th to 12th grade]. Hogrefe.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011a). *The MatchIt package* (3.0.2) [Computer software]. http://gking.harvard.edu/matchit

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011b). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software, Articles*, *42*(8), 1–28. https://doi.org/10.18637/jss.v042.i08

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Marsh, H. W. (1990b). *Self Description Questionnaire II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept: A test manual and research monograph.* Psychological Corporation.

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

McDonald, R. P. (1999). *Test theory: A unified treatment.* Lawrence Erlbaum Associates Publishers.

Muthén, L. K., & Muthén, B. O. (1998-2019). *Mplus* (8.4) [Computer software]. Author.

Pekrun, R., Goetz, T., & Perry, R. P. (2005). *Achievement Emotions Questionnaire (AEQ)—User's Manual*. Department of Psychology, University of Munich.

Pekrun, R., Jullien, S., Zirngibl, A., Blum, W., Götz, T., vom Hofe, R., Jordan, A., Kleine, M., & Wartha, S. (2003). *PALMA – Projekt zur Analyse der Leistungsentwicklung in Mathematik: Skalenhandbuch Erhebungswelle II* [Unpublished project codebook]. Department of Psychology, University of Munich.

Pekrun, R., vom Hofe, R., Blum, W., Frenzel, A., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools: The final report on the DFG priority programme* (pp. 17–37). Waxmann.

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. (2014). *Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package*. RAND Corporation.

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. (2015). *The twang package* (1.6) [Computer software]. https://CRAN.R-project.org/package=twang

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 15. http://dx.doi.org/10.1093/biomet/70.1.41

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, *79*(387), 9. https://doi.org/10.2307/2288398

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38. https://doi.org/10.2307/2683903

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

Thoemmes, F. J., & Kim, E. S. (2011). A Systematic Review of Propensity Score Methods in the Social Sciences. *Multivariate Behavioral Research*, *46*(1), 90–118. https://doi.org/10.1080/00273171.2011.540475

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *24*(1), 3–7.

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere die Richtigkeit der vorangegangenen Erklärung und bin mir der strafrechtlichen Folgen einer Falschaussage bewusst.

Trier, 28.08.2023	Markus Feuchter
Ort, Datum	Unterschrift