Interactions of Gender with Predictors of Academic Achievement

Cathy Hauspie[1], Stijn Schelfhout[1, 2, 3], Nicolas Dirix[1], Lot Fonteyne[4], Arnaud Szmalec[1, 5, 6], and Wouter Duyck[1, 3, 7]

[1] Ghent University, Faculty of Psychology and Educational Sciences, Department of Experimental Psychology, Henri Dunantlaan 2, 9000 Ghent, Belgium.

[2] Ghent University, Faculty of Psychology and Educational Sciences, Department of Work, Organization and Society, Henri Dunantlaan 2, 9000 Ghent, Belgium.

[3] Ghent University, Faculty of Psychology and Educational Sciences, Interdepartmental Research Group Vocational and Personnel Psychology, Henri Dunantlaan 2, 9000 Ghent, Belgium.

[4] Ghent University, Department of Educational Policy, Student Counseling Office, Sint-Pietersnieuwstraat 33, 9000 Ghent, Belgium.

[5] Psychological Sciences Research Institute, Université catholique de Louvain, Place du Cardinal Mercier 10, 1348 Louvain-la-Neuve, Belgium.

[6] Institute of Neuroscience, Université catholique de Louvain, Avenue E. Mounier 53 & 73, 1200 Brussels, Belgium.

[7] NVAO, The Accreditation Organisation of the Netherlands and Flanders, Parkstraat 83, 2514 JG Den Haag, The Netherlands

Accepted for publication in *Contemporary Educational Pyschology* on April 4th, 2023

Abstract

Predictive models of academic achievement are used in various (often high stakes) applications, including selection and study orientation procedures for higher education. Considering the far-reaching consequences of their outcomes, these models should show as little bias for irrelevant factors as possible. While numerous studies have researched the impact of gender on the isolated individual predictors of academic achievement, no studies yet have explored how gender affects program-specific prediction models of academic achievement. As such, the present study examined whether prediction models exhibit gender differences in the accuracy of their predictions, and how such differences relate to the gender balance within a study program. Besides that, we developed gender-specific prediction models of academic achievement in order to examine how these models differ in terms of which predictors are included, and whether they make more accurate predictions. Data was examined from a large sample of first year students across 16 programs in an open access higher education system (*N* = 5,016). Results revealed interactions between gender and several predictors of academic achievement. While the models exhibited little difference in the accuracy of their predictions for male and female students, analyses showed that using gender-specific models substantially improved our predictions. We also found that male and female models of academic achievement differ greatly in terms of the predictors included in their composition, irrespective of the gender balance in a study program.

Keywords: gender differences; academic achievement prediction; higher education

The prediction of academic achievement has been a field of interest within educational research for over a century (Petrides et al., 2005), and even the first intelligence tests were developed specifically for educational purposes (Binet, 1903). Ever since, a great deal of research has focused on developing and improving prediction models of academic achievement by identifying and testing a variety of predictors (Schneider & Preckel, 2017). Such models serve different practical applications, depending on the educational context in which they are used. In an open access system, there are no limitations based on previous achievement (e.g., no entry exams) (OECD, 2020). In such a system, modelling future academic achievement is crucial for the development of study orientation tools. These tools inform and support students in their study choice process by helping them decide which study programs fit their interests, and whether they are attainable (XXXX et al., 2017). In contrast, a closed access system has more strict admission procedures, such as high-stakes testing, which is common in Anglo-Saxon education (e.g., the SAT: Scholastic Aptitude test) (OECD, 2020). In such a context, predictive models of academic achievement are used for selection procedures in higher education, where counseling offices make binding decisions on who can or cannot enter their university (programs) (Kuncel et al., 2001).

Various cognitive, affective, and demographic factors are typically included in predictive models of academic achievement (e.g., cognitive ability, academic self-efficacy, test anxiety, gender, and many more) (XXXX et al., 2022). The variable of interest in this study is gender. Importantly, sex and gender are not interchangeable constructs. The former is biologically determined by one's physical attributes (e.g., genitalia, hormones, …) and is generally referred to as a binary construct (i.e., male or female). The latter is socially determined, and concerns behaviors, activities and societal roles typically associated with males or females. Gender is considered to be a spectrum, but most studies only consider the most common gender identities, namely male and female (Bass et al., 2018; Diamond, 2020; Torgrimson & Minson, 2005). As the gender construct provides a more nuanced understanding of an individual's identity and of the societal and cultural influences that affect one's experiences, whereas sex is a more limited construct in that regard (American

Psychological Association, 2022), we have chosen to only consider gender in this study. Research has reported gender differences in some of the predictors used in predictive models of academic achievement. For instance, studies have shown that female students score higher on test anxiety measures than their male counterparts (Cassady & Johnson, 2002; Núñez-Peña et al., 2016). Over time, the literature has also reported some specific interaction effects of gender and a given predictor on academic achievement that appear in such models (Mellon et al., 1980; Steinmayr & Spinath, 2008). For example, research has shown that intrinsic motivation is a stronger predictor of academic achievement for male students (Cortright et al., 2013; Freudenthaler et al., 2008). Unfortunately, the majority of research has studied such effects in isolation, only looking at one factor and its interaction with gender, not controlling for the myriad of other supposedly important predictors of academic achievement. No studies yet have researched whether gender also affects the predictive value of such factors in the context of comprehensive prediction models of academic achievement. Examining the interactions between gender and predictors of academic achievement in this framework is valuable, as prediction models allow us to account for the complex interrelationships among a system of variables (Ruban & McCoach, 2005). Thus, based on the literature regarding the interaction effects between gender and separate predictor values, the first research goal of the present study is to further evaluate the effect of gender in the context of prediction models of academic achievement.

As mentioned earlier, predictive models of academic achievement are used in several high-stakes applications, such as study orientation and selection procedures. If gender actually affects the composition of these prediction models or the accuracy of its predictions, the current models fail to take this into account. As a result, students' projected academic achievement could be over- or underestimated, which could lead to students being oriented towards or selected for the wrong majors. This could hold disadvantageous repercussions for their career in postsecondary education. An example of such a gender bias was reported in the study by Tulbure and Gavrilla (2019), who reported a high frequency of overachievement in women, and a high frequency of underachievement in men, based on what could be predicted by intelligence tests. Such effects could be due to real gender differences in study motivation (Freudenthaler et al., 2008), but also to a previously undetected interactions effects of gender with predictors in models of academic achievement. Our second research goal is thus to explore whether predictive models of academic achievement exhibit gender differences in the accuracy of their predictions.

Besides gender bias in predictions of academic achievement, there have also been reports of gender bias in study interest orientation tools. Indeed, studies have shown that there are not only substantial gender differences in predictor values, but also in the study choice process (XXXX et al., 2021; Stoet & Geary, 2020). For instance, XXXX and colleagues (2021) reported that females (vs. males) have more explicit knowledge about their profile of vocational interests and they also use this knowledge more explicitly. Furthermore, research has shown that study choices for the field of science, technology, engineering, and mathematics (STEM) are less determined by STEM preparation (e.g., knowledge of mathematics) for females, as they evaluate their cognitive capabilities much more modestly than men (Buchmann et al., 2008; Nix et al., 2015; XXXX et al., 2021). In sum, gender affects to what extent various cognitive and affective factors play a role in the study choice process. Subsequently, the third research goal of the present study is to explore whether gender also affects which factors are included in predictive models of academic achievement.

If it appears that gender does indeed affect the composition of the factors included in such prediction models, it is crucial to assess whether gender-specific prediction models of academic achievement would allow for more accurate predictions. Therefore, our fourth research goal is to compare the predictions of gender-specific models and general prediction models. Finally, our fifth and final research goal is to examine how the results of our second, third and fourth research goal relate to the gender imbalance in a study program.

Effects of Gender on Predictors of Academic Performance

Research has identified a main effect of gender on a variety of predictors. The first research goal of the present study is to add to the existing scientific evidence by evaluating the effect of gender as a predictor and its subsequent interactions with other predictors of academic achievement. These predictors can be subdivided into three main categories: cognitive variables, affective variables, and demographic data (Richardson et al., 2012). First, cognitive measures include skills that refer to our ability to comprehend, process and work out complex ideas. These skills are used in writing, reading and numeracy, among others, and are typically compiled under the term 'cognitive ability' (Pierre et al., 2014; Rohde & Thompson, 2007). Second, affective measures refer to socio-emotional skills that are developed and socially determined over the course of our life. Variables such as personality and motivation are examples of the factors included in this category (XXXX et al., 2017; Pierre et al., 2014; Schneider & Preckel, 2017). Finally, demographic data includes information on student's background characteristics that may also influence academic achievement (e.g., gender, migration status, language, …) (Richardson et al., 2012).

Cognitive Predictors

Research shows that the correlation between cognitive ability (often operationalized as intelligence) and academic success ranges from r = .30 to r = .70, making cognitive ability one of the strongest predictors of academic achievement (Kriegbaum et al., 2018; Petrides et al., 2005; Rohde & Thompson, 2007; Roth et al., 2015). While most studies agree that there are no main effects of gender on general cognitive ability (Naderi et al., 2010), research on the interaction effect between gender and general cognitive ability on academic achievement is rather mixed. While Ruffing and colleagues (2015) report that cognitive ability is a stronger predictor of academic performance for men, other studies contradict the existence of interactions between gender and general or domain-specific intelligence in the prediction of academic achievement (Freudenthaler et al., 2008; Spinath et al., 2014). Spinath and colleagues (2014) for example, reported that females' better school performance cannot or can only partly be explained by their better verbal abilities. Similarly, male students' advantage in numerical abilities cannot or only partly account for their better scores on mathematics tests. Research has reported that such gender-specific differences in particular aspects of intelligence, such as verbal or numerical abilities, can be attributed to a combination of both biological as well as sociocultural variables (Nisbett et al., 2012).

Measures of prior study performance, which correlate greatly with cognitive ability, are also included in this category (Roth et al., 2015). This measure is often quantified as (high school) GPA (Grade Point Average) and is one of the most used predictors of academic achievement (XXXX et al., 2017; Hodara & Lewis, 2017; Poole et al., 2012; Richardson et al., 2012; XXXX et al., 2019).

Affective Predictors

The interaction of gender with affective predictors of academic achievement has been thoroughly researched, but the results are often inconclusive. In what follows, we provide examples on the most preeminent affective predictors of academic achievement, which all demonstrate incremental validity over and beyond measures of cognitive ability (XXXX et al., 2017; Schneider & Preckel, 2017).

Personality is a prominent affective predictor of academic achievement (XXXX et al., 2017; Petrides et al., 2005). More specifically, research has shown that the trait conscientiousness is an important predictor of academic achievement, as higher scores on this personality trait are associated with higher grades (Carvalho, 2016; Kling et al., 2013; Nguyen et al., 2005; Poropat, 2009; Trapmann et al., 2008). The personality literature describes conscientiousness as a construct related to facets such as orderliness and dependability (Duckworth et al., 2007). Conscientiousness encompasses various personality traits, such as self-control (the regulation of one's own conflicts between short- and long-term goals) and grit (the perseverance to work towards long-term goals and maintaining effort and interest, regardless of contingent obstacles), which are often included in prediction models as unique predictors of academic achievement (Duckworth et al., 2007, 2019). Studies report that the main effect of gender on conscientiousness is small to non-existent (de Fruyt et al., 2008; Steinmayr & Spinath, 2008), but research on the interaction effects between gender and personality traits is rather contradictory. On the one hand, de Fruyt and Mervielde (1996) reported that conscientiousness is a stronger predictor of degree attainment for males. On the other hand, other studies found no interaction effects between gender and the predictive value of personality traits on academic achievement (Mellon et al., 1980; Steinmayr & Spinath, 2008).

Metacognition can be described as the knowledge of one's own motivation and ability to use self-regulatory techniques during studying. This construct is positively associated with academic performance (Kitsantas et al., 2008; Richardson et al., 2012). For instance, research shows that flawed metacognition is associated with a superficial approach to studying (r = .42). Such an approach to learning then correlates with poorer academic achievement (r = -.33) (Spada & Moneta, 2014). The literature on the existence of gender differences in the use of metacognitive strategies is inconclusive (Callan et al., 2016). To our knowledge, research on gender differences in the predictive value of metacognition does not exist as of today.

Studies show that increasing levels of academic self-efficacy (i.e., the confidence in one's ability to attain the desired academic goals) correlates positively with academic performance and persistence (Komarraju & Nadler, 2013; Pirmohamed et al., 2017; Robbins et al., 2004; Turner et al., 2009). The meta-analysis by Huang (2013) showed that female students tend to score higher on language arts self-efficacy, while males display higher mathematics, computer and social sciences self-efficacy. When looking at academic self-efficacy in a broader sense, Vantieghem and colleagues (2014) report that females tend to score higher, while Huang (2013) reports that the gender differences vary with age. To our knowledge, literature on how these gender differences are reflected in the predictive value of self-efficacy does not exist yet.

Test anxiety is a construct negatively correlated with academic success (Cassady & Johnson, 2002; Credé & Kuncel, 2008; Núñez-Peña et al., 2016). One experiences test anxiety when feeling anxious in settings related to learning and evaluation (Credé & Kuncel, 2008). Several studies report that females score higher on test anxiety measures (Cassady & Johnson, 2002; Núñez-Peña et al., 2016). However, research shows that there are no gender differences in the predictive value of test anxiety on academic achievement (Torrecilla-Sánchez et al., 2019).

According to the Self Determination Theory (SDT), different types of motivation are supposed to drive students' behavior. A student can display autonomous (driven by internal factors) and controlled (driven by external factors) motivation at the same time (Deci & Ryan, 2008; Kusurkar et al., 2013; Ratelle et al., 2007; Vansteenkiste et al., 2009). A multitude of studies report that motivation predicts school achievement, with positive effects specifically for autonomous motivation (Freudenthaler et al., 2008; Kriegbaum et al., 2018; Soenens & Vansteenkiste, 2005). While some studies show that intrinsic motivation is a stronger predictor of academic performance for men (Cortright et al., 2013; Freudenthaler et al., 2008), Steinmayr and Spinath (2008) did not find such interaction effects. The present study addresses this conflicting evidence.

Finally, vocational interests are a proven unique predictor of academic achievement (XXXX et al., 2017; XXXX et al., 2019), although the explained variance remains limited. One of the most prominent models to depict vocational interests in higher education, is the theory of vocational personalities and environments proposed by Holland in 1997 (Nauta, 2010).The essence of the theory is a hexagonal model, consisting of six interest dimensions: realistic, investigative, artistic, social, enterprising, and conventional (abbreviated as RIASEC) (de Fruyt & Mervielde, 1996; Päßler & Hell, 2012). The model allows to measure the congruence between a person's profile and that of his environment, also known as the person-environment fit (PE fit) (Astin & Holland, 1961; Holland, 1997; Nauta, 2010; XXXX et al., 2019). Research has shown that the better the PE fit, the higher the grades and persistence in school (Nauta, 2010; Nye et al., 2012; Päßler & Hell, 2012; Rounds & Su, 2014). Gender differences in vocational interest profiles have been thoroughly researched. In a meta-analysis by Su and colleagues (2009), an effect size of d = .93 was reported for gender differences on the things-people interest dimension. Men score higher on the things-oriented and realistic dimensions, while women tend to score higher on the person-oriented and social dimensions (Nauta, 2010; Rounds & Su, 2014; Su et al., 2009). As vocational interests are heavily influenced by socialization processes and environmental influences, such as parental expectations, these gender differences are to be expected (Ion et al., 2019). Besides that, research by XXXX and colleagues (2021) showed that women in STEM majors tend to show a better interest fit with their study programs in comparison to their male counterparts. In sum, gender not only influences vocational interest profiles, but also how well a person fits his or her environment. Unfortunately, no studies yet have investigated whether gender also affects the predictive value of vocational interests.

Clearly, the importance of affective predictors has been well established. Nevertheless, the existence of interactions with gender is unclear in several cases, such as for motivation, vocational interests, and metacognition. Furthermore, the interactions were often only explored when the predictors were studied in isolation, meaning that only one factor and its interaction with gender were associated with academic achievement, not controlling for several other potentially relevant predictors of academic achievement. For example, this was the case in the studies on test anxiety (Torrecilla-Sánchez et al., 2019) and motivation (Kusurkar et al., 2013).

Demographic Predictors

Finally, the last category of predictors of academic achievement concerns demographic variables, such as socioeconomic status (SES), gender or age (Richardson et al., 2012). Because of the continuously growing amount of students in higher education, student populations have become increasingly diverse (OECD, 2020). Consequently, research that explores how such demographic factors affect the prediction of academic achievement is indispensable. In an example of Sirin (2005), SES shows a medium to strong association with academic achievement, although other research also suggests that it is important to separate SES from other variables, such as mother’s cognitive abilities, when making such claims (Marks & O’Connell, 2021a, 2021b). A study by Connolly (2006) has shown that there is no evidence in favor of gender differences in the predictive value of demographic predictors, such as social class or ethnic groups, on educational attainment.

The Present Study

The literature on academic achievement is in need of further research into the interactions between gender and cognitive, affective, and demographic predictors, as empirical data on this topic are either lacking or inconclusive. Furthermore, most studies only examined the effect of gender when predictors were studied in isolation. As such, our first research goal is to evaluate interactions between gender and predictors of academic achievement in program-specific prediction models. If gender interacts with how strongly a factor should contribute to the prediction of academic achievement, the prediction models that are currently used, which are typically generalized across gender, could result in imprecise predictions. For instance, if intrinsic motivation is a stronger predictor for men than for women (Cortright et al., 2013; Freudenthaler et al., 2008), but is attributed the same weight for both genders in the prediction model, men's academic achievement would be underestimated, thus resulting in less accurate predictions. In line with this, we present the following hypothesis:

H_1: *Cognitive, affective, and demographic variables interact with gender in the prediction of academic achievement.*

Besides that, it is imperative to explore how gender affects predictions of academic achievement. Research has already shown that gender differences exist in the prediction of academic achievement (e.g., women achieve higher college grades than predictions based on an intelligence test would suggest (Tulbure & Gavrilla, 2019). As such, our second research goal is to assess whether a gender-based accuracy difference in the predictions of our models exists, and thus if the general models perform as well for females as they do for males:

H_2: *The program-specific prediction models do not demonstrate the same accuracy in their predictions for both genders.*

Such accuracy differences could be the result of the finding that different factors are predictive of academic achievement for men and women within the same study program. Such differences are currently undetected, thus possibly resulting in accuracy differences between genders in the prediction of academic achievement. With our third research goal, we explore whether different factors are predictive of academic achievement for men and women:

H_3: *The composition of the program-specific prediction models for males is different than that for females.*

If a gender difference exists in the composition of the predictive models of academic achievement, exploring whether gender-specific models make more accurate predictions is desirable. For our fourth research goal, we aim to examine whether gender-specific models make the most accurate predictions:

H_4: *The gender- and program-specific prediction models make more accurate predictions and explain more variance than the program-specific prediction models across genders.*

However, it is crucial to recognize that these prediction models are determined by the students in the used samples. Literature has shown that various study fields are associated with highly gendered narratives (thus impacting the gender balance in these programs), and how vocational interests might be the critical factor behind this (Su et al., 2009). Indeed, vocational interests are heavily influenced by gendered societal roles and expectations (Ion et al., 2019), while also being an important predictor of study choice (XXXXX et al., 2021). If certain samples are mainly male- or female-dominated, chances are that such a gender imbalance is at least partly responsible for any observed effects in the previous hypotheses. As such, we aim to control for the effect of the gender imbalance of each study program with our fifth and final research goal:

H_5: *The gender imbalance of a study program correlates with the results of H2, H3 and H4.*

Method and Materials

The XXXX[1] Project in an Open Access Study Environment

We performed secondary analyses on data gathered within the longitudinal, university-wide XXXX project, developed at XXXX, a renowned and well-established Western-European university (ARWU top 100 of the Shanghai ranking of worldwide universities). The data were obtained within an open access study environment: all students with a degree from secondary school can start any program in higher education, except medicine, dentistry, and performance arts (OECD, 2020). The

[1] Potential references to the contributing authors or the host institution have been masked by 'XXXX'.

XXXX project was developed to remedy low success rates in higher education: over the time period of 2016-2018, only 36% of all first-year students at XXXX were able to succeed in all courses, which is necessary to stay on track for timely degree attainment. The problem of low success rates is not specific to XXXX, as 30% of the students in OECD countries leave higher education without obtaining a degree (OECD, 2017). The goal of this project is thus to improve these success rates, by equipping (aspiring) university students with fitting program-specific advice. To develop such advice, an algorithm has been established based on an extensive historical dataset (with more than 70,000 entries as of 2021) containing former students' test results on an internet-based self-assessment tool and exam scores. The tests included in the self-assessment tool measure a multitude of predictors of academic achievement, as previously established in the introduction. As such, to provide program-specific advice to new students, their test results are fed into this algorithm. If the comparison to previous scores indicates that the student has a very low likelihood of succeeding, he or she is advised to improve their basic abilities or to reorient towards a more attainable or suitable study program. High response rates are acquired, as participation in the XXXX project is strongly encouraged among students at the start of their first academic year. After the students finish their first year, their exam results are linked to their original data, which ensures a continued improvement of the prediction models (XXXX et al., 2017; XXXX et al., 2022).

Data

The present dataset includes test results of the XXXX project and the consecutive exam results of first year university students, collected within the XXXX project over the time period 2016-2018. As only study programs with $n > 120$ students were included[2], this results in data from N =

[2] First, we took a sample size of 50 as a starting point. This baseline allowed us to take a conservative approach and filter out any study programs with very little amounts of students. Second, we wanted to include 10 participants per predictor that would be included in the model, based on the literature by Peduzzi and colleagues (1995, 1996). To get an estimate of how many predictors are generally included in a prediction model of academic achievement, when the predictor pool is this large, we based ourselves on XXXX (2019). In his study, a similar number of predictors functioned as input into an AIC algorithm, to develop separate prediction models for 21 study programs. From these final 21 prediction models, a mean of 5 predictors per model was recovered, with a standard deviation of 2. As such, as we include 10 participants per predictor, we

5,016 students across 16 university programs (an overview is presented in Table 1).

Ethics Statement

The Ethical Commission of XXXX has granted approval to “XXXX" of which the present study is an integral part, with reference XXXX. This study was carried out in accordance with the recommendations of the Ethical Commission of XXXX. All subjects gave their online informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethical Commission of XXXX.

Measures

An overview of the set of variables used to construct the predictive models is given in Table 2. To decide which variables should be included in our analyses, we based ourselves on the reasoning by Shmueli (2010). She proposed three main criteria for choosing which variables to include in predictive models: the quality of the association between the predictors and the response, the data quality, and the availability of the predictors at the time of prediction. Based on our extensive literature review, as presented in the introduction, we selected the variables that showed the best predictive association with academic achievement and are thus expected to provide incremental predictive value. Second, we checked these predictors’ availability in our dataset and we found that all the variables we selected through our literature review were present in this dataset. Finally, we know that the data on these measures is of good quality, as these measures have been successfully used in study orientation devices from 2015 on (XXXX et al., 2017; XXXX et al., 2022). As such, we decided to include them in the present study. Given the extensive nature of the item pool, we refer to Appendix A for a full description of these predictors and their reliability. Given the extensive nature of the item pool, we refer to Appendix A for a full description of these predictors and their reliability.

---

added another 50 participants to the minimum sample size, based on the mean number of predictors in a model we calculated. We wanted to be extra conservative, so we also added another 20 participants to the minimum sample size, based on the standard deviation of 2 predictors per model. In sum, the following calculation was made: 50 (minimum sample size) + 5 (mean number of predictors) x 10 + 2 (standard deviation of number of predictors) x 10 = 120.

Procedures and Analyses

First, predictive models for GPA were built. In this study, GPA is the weighted average of students' local grades, with the ECTS (European Credit Transfer and Accumulation System) credits being used as weights. Generally, in order to estimate how much population variance in academic achievement a combination of predictors explains, researchers opt for linear or logistic regression models. These regression models allow researchers to reveal the total explained variance of their model, along with the inputs of particular predictors incremental to other known predictors (XXXX et al., 2022). This explanative approach has been used by most studies presented earlier. Yet, even when a model displays a large amount of explained population variance, a good prediction of individual student results is not necessarily guaranteed (Shmueli, 2010). As the present study aims to further improve individual student prediction of academic achievement, we modelled academic achievement using the Akaike's Information Criterion (AIC) stepwise selection procedure. This procedure ensures that the best possible prediction model is chosen, by minimizing the model's prediction errors and consequently, the information loss (Burnham & Anderson, 2002). However, research by Hurvich and Tsai (1989) has shown that small sample sizes elicit an AIC model characterized by a potentially high degree of negative bias. Because the sample sizes of our individual study programs are rather limited in some cases, it was thus imperative to prevent overfitting (Shmueli, 2010). As such, we opted to use a more stringent version of the AIC (AICc) where a correction based on the sample size of each study program is included for every predictor that joins the model (Cavanaugh, 1997). By comparing all predictor combinations imaginable to one another, the grouping with the smallest probability of information loss is chosen as the final model. Afterwards, the model's prediction is compared to the actual performance of each student separately. Finally, the selected model is the one that delivers the smallest prediction error across all students. While these models were primarily developed for the prediction of individual students' performance, they can also be used to explain the population variance of study results in their programs (XXXX et al., 2021).

Typically, models are run across programs, in lieu of building models specific to study programs. However, several studies found that adapting a program-specific method was beneficial compared to a more general approach, as this method can significantly improve the individual student prediction of academic achievement (XXXX et al., 2017; XXXX et al., 2021). Program-specific predictive models were thus constructed across genders (the general models), as well as for male and female students separately (the gender-specific models). The prediction models across programs are also reported, which are based on the data from each program, except medicine and rehabilitation science and physiotherapy, as these programs do not include data on comprehensive reading. Besides that, the predictor list of the models across programs does not include data on chemistry and physics, as most programs are not tested on these predictors (for more information, we refer to Appendix A).

For our first research goal, we counted in which percentage of the program-specific models across genders each predictor category was present. Afterwards, we checked how many of these programs also demonstrated interactions between the present predictor categories and gender. For our second research goal, we compared the prediction errors of the general model for male data with the prediction errors for female data, using independent samples t-tests. For our third research goal, we checked to which degree the male and female program-specific models corresponded. Furthermore, we compared the prediction errors of the male-specific model on male data, with the prediction errors of the female-specific model on female data, using independent samples t-tests. We also assessed differences in the explained sample variance for each model comparison. For our fourth research goal, we explored whether gender-specific prediction models make better predictions than the models across gender. As such, we compared the prediction errors on male data of the general and the gender-specific models, using paired samples t-tests. Besides that, we compared the explained sample variances of the models. The same was done for the female data. Finally, we performed an exploratory analysis where we compared the explained variance of the general model (first on the data across genders and second on the gender-specific data) with that of the gender-specific models (on the gender-specific data). Finally, for our fifth research goal, we determined the correlation between the outcomes of the second, third and fourth research goal, and the gender proportion of each program.

Considering the multiple hypotheses tested in this study, we need to remain wary of increased type I error rates. However, every analysis that required t-tests was performed for each of the 16 study programs separately. For this, the entire dataset was split up into 16 different subsamples, with each subsample consisting of the students of a particular study program. Subsequently, the data was split up into male and female data, resulting in 32 subsamples. The t-tests that thus accompanied each hypothesis were mainly based on different subsamples of the dataset. A maximum of 3 t-tests per subsample was performed in total. As our samples were quite large, and we included a correction for small sample sizes through the AICc procedure, we did not perform any additional corrections. Furthermore, we always report the effect sizes[3], which are independent from significance levels and allow us to assess the magnitude of each effect.

Results

The correlation matrix of the predictors and the (program-specific) prediction models are reported in Appendix B and C, respectively. For H_1, we first checked to what extent each predictor category was present in the program-specific general models. Afterwards, we explored how many of the models that included a certain predictor category, also displayed significant interactions between that category and gender. The analyses were only performed on the program-specific models, thus excluding the model across programs. The results are reported in Table 3. We found that cognitive and demographic predictors were present in all 16 prediction models, and affective predictors were found in 14 models. Cognitive predictors displayed the most interaction effects with gender (in 7 models), followed by affective predictors (in 4 models), while interaction effects between gender and demographic predictors were only present in 1 model. These results are in

[3] A Cohen's *d* was used to report these effect sizes, with 0.01 = very small effect, 0.20 = small effect, 0.50 = medium effect, 0.80 = large effect, 1.20 = very large effect and 2.00 = huge effect (Sawilowsky, 2009).

support of H_1. For a more detailed explanation on which predictors displayed interactions with gender, we refer to the discussion. Furthermore, we refer to Appendix C for a detailed overview of the configuration of each program-specific prediction model.

The results of the analyses to test H_2 are documented in Table 4. In order to compare the performance of the general model on male data with its performance on female data, we used independent samples t-tests to analyze differences in the elicited mean prediction errors. Significant differences in these prediction errors were found in programs 1, 9 and 13. The effect sizes for these significant programs ranged from small (d = 0.28) to medium (d = 0.57). For programs 1 and 9, the prediction errors were significantly higher in male data. The opposite was true for program 13, which displayed significantly higher prediction errors in female data. In line with H_5, we then determined the correlation between the effect sizes (excluding that of the model across programs) with the gender proportion of each program (which is documented in Table 1). A correlation of $r(14) = .49$, $p = .054$ was revealed across 16 programs. In other words, the more gender-skewed program populations are, the less accurate the predictions of the general model will be.

Afterwards, we developed gender- and program-specific prediction models of academic achievement. To test H_3, we checked the correspondence between the male and female predictive models of each study program. The correspondence measure was calculated by dividing the total amount of shared predictors between the male and female model, by the total amount of predictors present across both models. The results are reported in Table 6. The mean correspondence (without the models across programs) is M = 0.22 with a standard deviation of SD = 0.12. To examine H_5, we then correlated the correspondence measures with the gender proportion of each study program, excluding the model across programs. A fairly strong correlation of $r(14) = -.52$, $p = .039$ was observed, meaning that a stronger gender imbalance in the population of a study program leads to more diverse gender-specific prediction models.

We then performed an exploratory analysis to compare the performance of the gender-specific models on the gender-specific data, in a similar way as with H_2. We used independent samples t-tests to analyze differences in the elicited mean prediction errors. We also included differences in the explained variance. The results are reported in Table 7. Significant differences in the prediction errors were found in programs 3, 7, 9, 15, 16 and across programs. The effect sizes ranged from very small (d = 0.09) to medium (d = 0.51) and correlated strongly with the absolute value of the R^2 differences: $r(14) = .75$, $p < .001$ (calculated without the model across programs). For programs 3, 7, 16 and across programs, the prediction errors of the male specific model were significantly higher. The opposite was true for programs 9 and 15, which displayed significantly higher prediction errors for the female specific model. We determined the correlation between the effect sizes (excluding that of the model across programs) with the gender proportion of each program (see Table 1). A non-significant correlation of $r(14) = .49$, $p = .054$ was revealed across 16 programs. In other words, the more one gender dominates the student body of a program, the larger the difference between the size of the prediction errors of the gender-specific models.

For H_4, analyses were performed to check which models produce the smallest prediction errors and explain the most variance: the general or the gender-specific models. First, we compared the performance when both models use gender-specific data. As such, we used paired samples t-tests, of which the results are reported in Table 8. When we compared the performance on male data, analyses showed that the prediction errors of the male model were smaller than those of the general model each time. Significant values were found in programs 1, 4, 6, 7, 8, 13, 16 and across programs, with effect sizes ranging from small (d = 0.13) to large (d = 0.75). We also compared the explained variances, which were always higher in the male model, with the difference ranging from 1% to 27%. The prediction errors on female data were compared using the same techniques. The results showed that the prediction errors of the female model were generally smaller than those of the general model, except in programs 2, 3 and 16. Significant values were found in programs 6, 13, 15, 16 and across programs, with effect sizes varying from small (d = 0.16) to somewhat large (d = 0.67). The differences in R^2 ranged from 0% to 22%, and the R^2 was higher in the female model in 75% of the programs. In line with H_5, an analysis was then performed to determine the correlation between the effect sizes reported in Table 8 and the gender proportion of each program (see Table 1). We found a non-significant correlation of $r(30) = .32$, $p = .072$. Similar results were found for the correlation between absolute R^2 differences and gender proportion: $r(30) = .36$, $p = .04$. In sum, a stronger gender imbalance in the population of a study program leads to an increase in the difference in prediction errors between the general and the gender-specific model, and a bigger difference in explained variance.

For the second part of the analyses for H_4, we compared the explained variance of the general model on the data across genders with the explained variance of the gender-specific models on the gender-specific data. The results are reported in Table 9. The R^2 difference between the male and general models ranged from 1% to 32%, with the male model demonstrating a higher explained variance in 75% of the programs. The R^2 difference between the female and general models ranged from 1% to 22%, with the female model explaining more variance in 37.5% of the programs.

Discussion

Various studies have established the existence of gender differences in the relative importance of several predictors of academic achievement (Mellon et al., 1980; Steinmayr & Spinath, 2008). However, most research has studied the effect of gender on these predictors in isolation, rather than in the context of prediction models of academic achievement. The present study aimed to address this gap in the literature, as these models are used in several practical applications of considerable influence (e.g., study orientation or selection procedures for higher education). We developed program-specific AIC models of academic achievement (first-year GPA) of a large sample of students, using cognitive, affective, and demographic predictors. First, we explored to what extent each predictor category interacts with gender in the program-specific prediction models. Second, we investigated whether the general predictive models of academic achievement make equally accurate predictions for males as they do for females. Third, we studied whether different factors were predictive of academic achievement for male and female students, per program individually. Fourth, we determined which model made the most accurate predictions: the gender-specific one, or the model across genders. Finally, we related the findings of the previous research goals to the gender imbalance of the various study programs.

The current study corroborated previous findings by demonstrating the importance of the three predictor categories in almost all general prediction models. For the gender-specific prediction models, we found that both the cognitive as well as the affective predictors were again present in almost all male and female models. However, we found that the demographic predictors were present to a lesser extent in female prediction models. Indeed, demographic predictors were only included in 68,75% of the female models, in contrast to 87,50% for the male models.

We further scrutinized the composition of the general prediction models and found interaction effects with gender for each predictor category, which validates the claim that gender affects the predictive value of various predictors, not only in isolated settings, but also in the context of prediction models. More specifically, our findings support the conclusions of previous studies by demonstrating that the predictive value of measures of cognitive ability can be affected by gender (Ruffing et al., 2015). Furthermore, we showed that the predictors personality, motivation and test anxiety did not interact with gender (Mellon et al., 1980; Steinmayr & Spinath, 2008; Torrecilla-Sánchez et al., 2019). Finally, this is the first study to demonstrate interactions between gender and the predictive value of measures of prior study performance, academic self-efficacy, metacognition, and vocational interests. In sum, these results confirm our hypothesis that gender does interact with predictors of academic achievement. To what extent these interactions with gender are the result of social or cultural influences, rather than innate determinants, is not something we can establish based on our data. However, it could be an interesting avenue for future research to address this.

The present study shows partial evidence that gender differences in prediction accuracy can occur with general prediction models, biased in favor of men (12.50% of the programs), as well as women (6.25% of the programs). This gender bias can be partially explained by the gender imbalance of the study programs tested. Indeed, a fairly strong correlation was discovered between the effect sizes (indicating the size of the difference in mean prediction errors) and the gender proportions of each study program. Put differently, if one gender is predominantly present in a study program, the predictions of the corresponding general model will be less accurate for one of both genders. The direction of this effect is not straightforward. For instance, if a program has predominantly male students, the prediction errors of the model were not necessarily higher on female data.

The current study established that the correspondence between male and female prediction models of a study program is generally low (mean correspondence: $M = 0.22$). Furthermore, post-hoc analyses[4] demonstrated that the correspondence between the gender-specific and the general predictive models are low too, with a mean correspondence of $M = 0.46$ and $M = 0.42$ for men and women respectively. We speculate that the composition of these gender-specific predictive models is affected by (1) the interaction effects of gender with predictor values (Mellon et al., 1980; Steinmayr & Spinath, 2008) and (2) by the different profiles at the start of higher education, due to the gender differences in the study choice process (XXXX et al., 2021; Stoet & Geary, 2020). Furthermore, we explored how these findings relate to the gender imbalance of a study program. We found that a stronger gender imbalance in the population of a study program leads to more diverse gender-specific prediction models, and a stronger correspondence between the general model and the model of the dominant gender. The low correspondence measures between the gender-specific models could thus be attributed to the fact that the prediction model of the least present gender is overly specified, which in turn causes the correspondence measure to decrease. However, we see the same pattern in study programs where the genders are more evenly distributed (e.g., bioscience has a 50/50 gender distribution, but the gender-specific models only have a correspondence of 11%. Meanwhile, the correspondence with the general model is 86% for the male model and 27% for the female model). In sum, we can still observe a large difference between the general, male, and female predictive models of academic achievement, irrespective of

[4] The analyses on the correspondence between the gender-specific and general models, and how these correspondence measures correlate with gender proportion, are reported in Appendix D.

the gender balance in a study program.

We also compared the size of the prediction errors elicited by the two gender-specific prediction models of each program and found sizable differences in 31% of the programs tested. In an attempt to explain these differences, we explored how they related to the gender imbalance of each study program. Fairly strong, positive correlations between the gender proportions of each study program and both the effect sizes (indicating the size of the difference in mean prediction errors) and the differences in explained variance were found. In other words, if one gender is substantially more present in a study program, the difference in the accuracy of the predictions between the two gender-specific models will increase, as will the difference in explained variance. This finding could raise doubt as to whether such gender-specific models would improve the current situation (i.e., the gender discrepancy in the accuracy levels of the predictions of the general models), as we still observe a discrepancy in the prediction accuracies. Future research could address this limitation and explore whether even bigger sample sizes could resolve this problem.

To our knowledge, this was the first study that studied and compared general with gender-specific models of academic achievement. Our results showed that gender-specific prediction models tend to improve predictions within gender groups, as its prediction errors were often significantly smaller than the prediction errors of the general models. This finding is particularly striking, given the fact that the gender-specific models are based on considerably less data than the prediction models across genders. Additionally, we explored how these differences in prediction accuracy were related to the gender imbalance of each study program. We found that, despite the low amount of data points, a stronger gender imbalance in a study program still is associated with larger differences in the accuracy of the predictions of the general vs. the gender-specific model.

Finally, we explored differences in the explained variances of the general and gender-specific models, based on gender-specific data. We found that the explained variances of the gender-specific models tended to be bigger than that of the general models (in all programs for the male model, and in 75% of the programs for the female model). These differences in explained variance were once again positively related to gender imbalance. However, we found that these conclusions were different when we compared the explained variances of the gender-specific models (on gender-specific data) with those of the general models (on the data across genders). While the male model still performed better in 75% of the programs, the female model only outperformed the general model in 37.50% of the programs. Future research should look further into this finding and explore why the female models tend to explain less variance than the male models do, when compared to the explained variance of the general model (on all data).

Limitations and Recommendations for Future Research

Our findings can be used to further improve the existing predictive models of academic achievement. By discerning our models by gender, we lowered prediction errors and improved our models' predictive value. The current study also has some limitations. First, the generalizability of our findings is constrained to open access study environments: the only qualification for our subjects to enter academic study programs is that they have a degree from secondary school. An interesting avenue for future research would be to replicate this study in a closed-access study environment. A second limitation is that the gender-specific models were sometimes based on quite small samples (*n* ranging from 39 to 600). While the present study certainly took precautionary measures to limit the negative consequences of such small samples (e.g., the AICc procedure), more research based on larger samples would be beneficial. A third limitation of our study is that cognitive ability was not operationalized in the same manner for every study program, due to the nature of the available variables. Two programs did not have any data on comprehensive reading, and only a small portion of the study programs offered data on physics and/or chemistry tests. Another limitation is that we used first year grades as an indicator of academic achievement. While research has reported that first year results are a valuable predictor of overall academic achievement (de Koning et al., 2012), future studies should investigate whether our results also hold for alternative outcomes such as timely graduation. Finally, given the history of our dataset, we categorized students into male and female students, based on their self-reported gender identity. However, it is important to recognize that the concept of gender identity is a spectrum that is much broader than such a binary distinction (Bass et al., 2018; Diamond, 2020). As such, further discussion is needed on how this gender spectrum should be addressed in future prediction models of academic achievement.

We believe that these empirical observations are important knowledge for study counselors when developing study orientation and/or selection tools, as our findings provide a valuable contribution to the theoretical understanding of the impact of gender in an academic setting. Indeed, we explored how gender interacts with predictors of academic achievement, and how a distinction by gender impacts predictive models of academic achievement, using a uniquely large sample size and predictor pool. Subsequently, it could be of great theoretical importance for other scholars to examine how large the, potentially undetected, effect of gender is on certain specific predictors of academic achievement. Such effects should be studied in the context of prediction models of academic achievement, as we have demonstrated the lack of research in this particular field. We deem it valuable to determine the impact that gender, or any other demographic variable, has in these prediction models, as a fair and unbiased prediction of academic achievement is paramount in study orientation or selection devices. Furthermore, improving such study orientation or selection devices by evaluating gender differences could be especially valuable for study fields that traditionally have highly gendered narratives, such as STEM programs.

Conclusion

The current study demonstrated that gender often interacts with cognitive, affective, and demographic predictors of academic achievement. Furthermore, we found that predictive models of academic achievement do not make equally accurate predictions for males and females in some study programs, which could be related to two mechanisms. First, we showed that this accuracy difference is partly due to imbalanced gender proportions in the program populations. Second, the present study demonstrated that the correspondence between the predictors present in the composition of general, male, and female models of academic achievement is remarkably low, irrespective of the gender balance of a study program. We also found that gender-specific prediction models of academic achievement tend to explain more variance and make more accurate predictions, compared to the general model. In practice, an interesting avenue would therefore be to further explore the benefits of gender-specific predictive models of academic achievement.

References

American Psychological Association. (2022). *Gender*. Retrieved from https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/gender#:~:text=Gender%20refers%20to%20the%20attitudes,to%20people%20as%20social%20groups.

Astin, A. W., & Holland, J. L. (1961). The Environmental Assessment Technique: A way to measure college environments. *Journal of Educational Psychology*, *52*(6), 308–316. https://doi.org/10.1037/h0040137

Bass, M., Gonzalez, L. J., Colip, L., Sharon, N., & Conklin, J. (2018). Rethinking gender: The nonbinary approach. *American Journal of Health-System Pharmacy*, *75*(22), 1821–1823. https://doi.org/10.2146/ajhp180236

Binet, A. (1903). *L'etude expérimentale de l'intelligence [Experimental study of intelligence].* Paris: Schleicher.

Buchmann, C., DiPrete, T., & McDaniel, A. (2008). Gender Inequalities in Education. *Annual Review of Sociology*, *34*, 319–337. https://doi.org/10.1146/annurev.soc.34.040507.134719

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A practical information-theoretic approach* (2nd ed.). Springer-Verlag.

Callan, G. L., Marchant, G. J., Finch, W. H., & German, R. L. (2016). Metacognition, strategies, achievement, and demographics: Relationships across countries. *Educational Sciences: Theory and Practice*, *16*(5), 1485–1502. https://doi.org/10.12738/estp.2016.5.0137

Carvalho, R. G. G. (2016). Gender differences in academic achievement: The mediating role of personality. *Personality and Individual Differences*, *94*, 54–58. https://doi.org/10.1016/j.paid.2016.01.011

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics and Probability Letters*, *33*(2), 201–208. https://doi.org/10.1016/s0167-7152(96)00128-9

Connolly, P. (2006). The effects of social class and ethnicity on gender differences in GCSE attainment: A secondary analysis of the Youth Cohort Study of England and Wales 1997-2001. *British Educational Research Journal*, *32*(1), 3–21. https://doi.org/10.1080/01411920500401963

Cortright, R. N., Lujan, H. L., Blumberg, A. J., Cox, J. H., & DiCarlo, S. E. (2013). Higher levels of intrinsic motivation are related to higher levels of class performance for male but not female students. *Advances in Physiology Education*, *37*(3), 227–232. https://doi.org/10.1152/advan.00018.2013

Credé, M., & Kuncel, N. R. (2008). Study Habits, Skills, and Attitudes. *Perspectives on Psychological Science*, *3*(6), 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

de Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality*, *10*(5), 405–425. https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<405::AID-PER255>3.0.CO;2-M

de Fruyt, F., & Rolland, J. P. (2010). *PfPI. Beschrijving persoonlijkheid op het werk. Handleiding [PfPI. Description personality at work. Manual].* Amsterdam/Antwerpen: Pearson.

de Fruyt, F., van Leeuwen, K., de Bolle, M., & de Clercq, B. (2008). Sex differences in school performance as a function of conscientiousness, imagination and the mediating role of problem behaviour. *European Journal of Personality*, *22*(3), 167–184. https://doi.org/10.1002/per.675

de Koning, B. B., Loyens, S. M. M., Rikers, R. M. J. P., Smeets, G., & van der Molen, H. T. (2012). Generation Psy: Student characteristics and academic achievement in a three-year problem-based learning bachelor program. *Learning and Individual Differences*, *22*(3), 313–323. https://doi.org/10.1016/j.lindif.2012.01.003

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology*, *49*(3), 182–185. https://doi.org/10.1037/a0012801

Diamond, L. M. (2020). Gender Fluidity and Nonbinary Gender Identities Among Children and Adolescents. *Child Development Perspectives*, *14*(2), 110–115. https://doi.org/10.1111/cdep.12366

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-Control and Academic Achievement. *Annual Review of Psychology*, *70*, 373–393. https://doi.org/10.1146/annurev-psych-010418-103230

Freudenthaler, H. H., Spinath, B., & Neubauer, A. C. (2008). Predicting school achievement in boys and girls. *European Journal of Personality*, *22*(3), 231–245. https://doi.org/10.1002/per.678

Hodara, M., & Lewis, K. (2017). *How well does high school grade point average predict college performance by student urbanicity and timing of college entry?* Washington, DC: U.S. Department of Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Psychological Assessment Resources.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education*, *28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y

Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Ion, A., Nye, C. D., & Iliescu, D. (2019). Age and Gender Differences in the Variability of Vocational Interests. *Journal of Career Assessment*, *27*(1), 97–113. https://doi.org/10.1177/1069072717748646

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-regulation and ability predictors of academic success during college: A predictive validity study. *Journal of Advanced Academics*, *20*(1), 42–68. https://doi.org/10.4219/jaa-2008-867

Kling, K. C., Noftle, E. E., & Robins, R. W. (2013). Why Do Standardized Tests Underpredict Women's Academic Performance? The Role of Conscientiousness. *Social Psychological and Personality Science*, *4*(5), 600–606. https://doi.org/10.1177/1948550612469038

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: Why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, *25*, 67–72. https://doi.org/10.1016/j.lindif.2013.01.005

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review*, *25*, 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kuncel, N. R., Ones, D. S., & Hezlett, S. A. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181. https://doi.org/10.1037/0033-2909.127.1.162

Kusurkar, R. A., Ten Cate, T. J., Vos, C. M. P., Westers, P., & Croiset, G. (2013). How motivation affects academic performance: A structural equation modelling analysis. *Advances in Health Sciences Education*, *18*(1), 57–69. https://doi.org/10.1007/s10459-012-9354-3

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Marks, G. N., & O'Connell, M. (2021a). No evidence for cumulating socioeconomic advantage. Ability explains increasing SES effects with age on children's domain test scores. *Intelligence*, *88*. https://doi.org/10.1016/j.intell.2021.101582

Marks, G. N., & O'Connell, M. (2021b). Inadequacies in the SES–Achievement model: Evidence from PISA and other studies. *Review of Education*, *9*(3). https://doi.org/10.1002/rev3.3293

Mellon, P. M., Schmitt, N., & Bylenga, C. (1980). Differential predictability of females and males. *Sex Roles*, *6*(2), 173–177. https://doi.org/10.1007/BF00287340

Ministerie van Onderwijs en Vorming [Department of Education]. (2012). *Omzendbrief SO/2012/01 [Newsletter SO/2012/01].* Retrieved from http://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14370.

Naderi, H., Abdullah, R., Aizan, H. T., & Sharir, J. (2010). Intelligence and academic achievement: An investigation of gender differences. *Life Science Journal*, *7*(1), 83–87. https://doi.org/10.7537/marslsj070110.15

Nauta, M. M. (2010). The Development, Evolution, and Status of Holland's Theory of Vocational Personalities: Reflections and Future Directions for Counseling Psychology. *Journal of Counseling Psychology*, *57*(1), 11–22. https://doi.org/10.1037/a0018213

Nguyen, N. T., Allen, L. C., & Fraccastoro, K. (2005). Personality predicts academic performance: Exploring the moderating role of gender. *Journal of Higher Education Policy and Management*, *27*(1), 105–117. https://doi.org/10.1080/13600800500046313

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New Findings and Theoretical Developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Nix, S., Perez-Felkner, L., & Thomas, K. (2015). Perceived mathematical ability under challenge: a longitudinal perspective on sex segregation among STEM degree fields. *Frontiers in Psychology*, *6*, 530. https://doi.org/10.3389/fpsyg.2015.00530

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2016). Gender Differences in Test Anxiety and Their Impact on Higher Education Students' Academic Achievement. *Procedia - Social and Behavioral Sciences*, *228*, 154–160. https://doi.org/10.1016/j.sbspro.2016.07.023

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational Interests and Performance: A Quantitative Summary of Over 60 Years of Research. *Perspectives on Psychological Science*, *7*(4), 384–403. https://doi.org/10.1177/1745691612449021

OECD. (2017). *The State of Higher Education 2015-2016*. Retrieved from http://www.oecd.org/education/imhe/the-state-of-higher-education-201516.htm

OECD. (2020). *Education at a Glance 2020: OECD Indicators*. OECD Publishing, Paris. https://doi.org/10.1787/69096873-en

Owen, S., & Froman, R. (1988). *Development of a college academic self-efficacy scale. Paper presented at the National Council on Measurement in Education*. New Orleans, L.A.

Päßler, K., & Hell, B. (2012). Do Interests and Cognitive Abilities Help Explain College Major Choice Equally Well for Women and Men? *Journal of Career Assessment*, *20*(4), 479–496. https://doi.org/10.1177/1069072712450009

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, *48*(12), 1503–1510.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379.

Petrides, K. V., Chamorro-Premuzic, T., Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behaviour and achievement. *British Journal of Educational Psychology*, *75*(2), 239–255. https://doi.org/10.1348/000709904X24735

Pierre, G., Sanchez Puerta, M. L., Valerio, A., & Rajadel, T. (2014). *STEP skills measurement surveys: innovative tools for assessing skills.* Retrieved from https://openknowledge.worldbank.org/handle/10986/19985.

Pirmohamed, S. A., Debowska, A., & Boduszek, D. (2017). Gender differences in the correlates of academic achievement among university students. *Journal of Applied Research in Higher Education*, 1–25. https://doi.org/10.1108/jarhe-03-2016-0015

Poole, P., Shulruf, B., Rudland, J., & Wilkinson, T. (2012). Comparison of UMAT scores and GPA in prediction of performance in medical school: A national study. *Medical Education*, *46*(2), 163–171. https://doi.org/10.1111/j.1365-2923.2011.04078.x

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–328. https://doi.org/10.1037/a0014996

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, Controlled, and Amotivated Types of Academic Motivation: A Person-Oriented Analysis. *Journal of Educational Psychology*, *99*(4), 734–746. https://doi.org/10.1037/0022-0663.99.4.734

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Robbins, S. B., Le, H., Davis, D., Lauver, K., Langley, R., & Carlstrom, A. (2004). Do Psychosocial and Study Skill Factors Predict College Outcomes? A Meta-Analysis. *Psychological Bulletin*, *130*(2), 261–288. https://doi.org/10.1037/0033-2909.130.2.261

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*(1), 83–92. https://doi.org/10.1016/j.intell.2006.05.004

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

Rounds, J., & Su, R. (2014). The Nature and Power of Interests. *Current Directions in Psychological Science*, *23*(2), 98–103. https://doi.org/10.1177/0963721414522812

Ruban, L. M., & McCoach, D. B. (2005). Gender differences in explaining grades using structural equation modeling. *Review of Higher Education*, *28*(4), 475–502. https://doi.org/10.1353/rhe.2005.0049

Ruffing, S., Wach, F.-S., Spinath, F. M., Brünken, R., & Karbach, J. (2015). Learning strategies and general cognitive ability as predictors of gender-specific academic achievement. *Frontiers in Psychology*, *6*, 1–12. https://doi.org/10.3389/fpsyg.2015.01238

Sawilowsky, S. S. (2009). Very large and huge effect sizes. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Soenens, B., & Vansteenkiste, M. (2005). Antecedents and outcomes of self-determination in 3 life domains: The role of parents' and teachers' autonomy support. *Journal of Youth and Adolescence*, *34*(6), 589–604. https://doi.org/10.1007/s10964-005-8948-y

Spada, M. M., & Moneta, G. B. (2014). Metacognitive and motivational predictors of surface approach to studying and academic examination performance. *Educational Psychology*, *34*(4), 512–523. https://doi.org/10.1080/01443410.2013.814196

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research*, *56*(2), 230–243. https://doi.org/10.1080/00131881.2014.898917

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

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences.* (5th ed.). Routledge.

Stoet, G., & Geary, D. C. (2020). Gender differences in the pathways to higher education. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(25), 14073–14076. https://doi.org/10.1073/pnas.2002861117

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests. *Psychological Bulletin*, *135*(6), 859–884. https://doi.org/10.1037/a0017364

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Torgrimson, B. N., & Minson, C. T. (2005). Sex and gender: What is the difference? In *Journal of Applied Physiology* (Vol. 99, Issue 3, pp. 785–787). https://doi.org/10.1152/japplphysiol.00376.2005

Torrecilla-Sánchez, E. M., Olmos Miguélañez, S., & Martínez Abad, F. (2019). Explanatory factors as predictors of academic achievement in PISA tests. An analysis of the moderating effect of gender. *International Journal of Educational Research*, *96*, 111–119. https://doi.org/10.1016/j.ijer.2019.06.002

Trapmann, S., Hell, B., Hirn, J. W., & Schuler, H. (2008). Meta-Analysis of the Relationship Between the Big Five and Academic Success at University. *Zeitschrift Für Psychologie/Journal of Psychology*, *215*(2), 132–151. https://doi.org/10.1027/0044-3409.215.2.132

Tulbure, C., & Gavrilla, C. (2019). Gender differences reflected in academic achievement in higher education. *Journal of Educational Sciences & Psychology*, *9*(1), 139–145.

Turner, E. A., Chandler, M., & Heffer, R. W. (2009). The influence of parenting styles, achievement motivation, and self-efficacy on academic performance in college students. *Journal of College Student Development*, *50*(3), 337–346. https://doi.org/10.1353/csd.0.0073

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational Profiles from a Self-Determination Perspective: The Quality of Motivation Matters. *Journal of Educational Psychology*, *101*(3), 671–688. https://doi.org/10.1037/a0015083

Vantieghem, W., Vermeersch, H., & Van Houtte, M. (2014). Transcending the gender dichotomy in educational gender gap research: The association between gender identity and academic self-efficacy. *Contemporary Educational Psychology*, *39*(4), 369–378. https://doi.org/10.1016/j.cedpsych.2014.10.001

Tables

Table 1

Study Programs

Number and Acronym	Program	*n*	*n* male	*n* female	Gender proportion [a]
-	Across programs [b]	4,570	1,873	2,697	.59
1 – PSY	Psychology	697	97	600	.86
2 – MED	Medicine	277	110	167	.60
3 – COM SCI	Communication science	155	39	116	.75
4 – POL SCI	Political science	121	67	54	.55
5 – LAW	Law	231	83	148	.64
6 – CRIM	Criminology	242	50	192	.79
7 – LING	Linguistics	243	57	186	.77
8 – VET MED	Veterinary medicine	228	49	179	.79
9 – REHAB SCI	Rehabilitation science and physiotherapy	462	166	296	.64
10 – PHARMA	Pharmaceutical science	307	56	251	.82
11 – BIOSCI	Bioscience	333	171	162	.51
12 – ECON	Economy	491	311	180	.63
13 – ENG	Engineering	338	269	69	.80
14 – COMMER SCI	Commercial science	377	171	206	.55
15 – ENG TECH	Engineering technology	441	397	44	.90
16 – APL LING	Applied linguistics	183	47	136	.74

[a] Gender proportion is the amount of students of the dominant gender in a study program, divided by the total amount of students in that program. [b] The data across programs is based on all programs but excluding data from medicine and rehabilitation science and physiotherapy (this will be explained in further detail in ‘Procedures and Analyses’).

Table 2

Pool of Predictors

Category	Variables
Dependent variable	GPA
Cognitive variables	Vocabulary Comprehensive reading [a] Mathematics (baseline/advanced) [b] Chemistry [c] Physics [d] High school Dutch package [e] High school mathematics package High school GPA
Affective variables	Conscientiousness Self-control Grit Academic self-efficacy (comprehension and effort) Test anxiety Motivation (autonomous and controlled) Metacognition (knowledge and regulation) Realistic interest dimension Investigative interest dimension Artistic interest dimension Social interest dimension Enterprising interest dimension Conventional interest dimension
Demographic variables	Age at start higher education Degree mother Degree father Language Nationality (Belgian – not Belgian) Socioeconomic status Scholarship

Note. GPA = grade point average.

[a] Not used in medicine and rehabilitation science and physiotherapy. [b] The advanced mathematics test was used in the programs medicine, bioscience, economy, engineering, commercial science and engineering technology. The normal mathematics test was used in the remaining programs. [c] Only used in veterinary medicine, rehabilitation science and physiotherapy, pharmaceutical science and bioscience. [d] Only used in veterinary medicine, rehabilitation science and physiotherapy, and pharmaceutical science. [e] While no explicit measure for writing was included in our variables, we

believe this variable was implicitly included in the variable ‘high school Dutch package’. In the educational system of XXXX, where this study was performed, writing is an especially important component of the Dutch course in secondary school (XXXX, n.d.). As such, we assume that writing was, at least partially, included in our prediction models of academic achievement.

Table 3

Presence of Predictor Categories and Subsequent Interactions with Gender

	% models in which the category is present	% models with interactions between the category and gender
Cognitive predictors	100	43.75
Affective predictors	87.50	28.57
Demographic predictors	100	6.25

Note. Variance Inflation Factor (VIF) values of the predictors were all below 10, indicating that the multicollinearity levels between the predictors lay within the acceptable range (Stevens, 2012).

Table 4

Comparison of the Prediction Errors Elicited by the General Models on Male and Female Data

Program	*t*-statistic [a]	Cohen's *d* [b]
Across programs [c]	$t(4568) = 0.95$	0.03
1 – PSY	$t(695) = 4.36$***	0.46
2 – MED	$t(275) = 0.66$	0.08
3 – COM SCI	$t(153) = -0.75$	0.14
4 – POL SCI	$t(119) = 0.54$	0.01
5 – LAW	$t(229) = 0.30$	0.04
6 – CRIM	$t(240) = 1.26$	0.20
7 – LING	$t(241) = -0.47$	0.07
8 – VET MED	$t(226) = 1.16$	0.19
9 – REHAB SCI	$t(460) = 2.83$**	0.28
10 – PHARMA	$t(305) = 1.84$	0.27
11 – BIOSCI	$t(331) = 0.99$	0.11
12 – ECON	$t(489) = 1.24$	0.12
13 – ENG	$t(336) = -3.40$**	0.57
14 – COMMER SCI	$t(375) = 1.37$	0.14
15 – ENG TECH	$t(439) = 0.78$	0.12
16 – APL LING	$t(181) = -0.60$	0.10

[a] A positive *t*-value indicates a higher mean prediction error on male data, while a negative *t*-value indicates a higher mean prediction error on female data. [b] The Cohen's *d* measure indicates the size of the difference in mean prediction errors. [c] The data across programs is based on all programs but excluding data from medicine and rehabilitation science and physiotherapy.

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

Table 5

Presence of Predictor Categories in Male and Female Models of Academic Achievement

	% male models in which the category is present	% female models in which the category is present
Cognitive predictors	100	100
Affective predictors	93.75	87.50
Demographic predictors	87.50	68.75

Note. Variance Inflation Factor (VIF) values of the predictors were all below 10, indicating that the multicollinearity levels between the predictors lied within the acceptable range (Stevens, 2012).

Table 6

Correspondence Between the Male and Female Models of Academic Achievement

Program	Correspondence gender-specific models [a]
Across programs [b]	.57
1 – PSY	.27
2 – MED	.33
3 – COM SCI	.00
4 – POL SCI	.31
5 – LAW	.38
6 – CRIM	.14
7 – LING	.19
8 – VET MED	.18
9 – REHAB SCI	.44
10 – PHARMA	.10
11 – BIOSCI	.18
12 – ECON	.33
13 – ENG	.13
14 – COMMER SCI	.22
15 – ENG TECH	.08
16 – APL LING	.29

[a] The correspondence measures are calculated as the total amount of shared predictors between the two models, divided by the total amount of predictors present across both models. [b] The data across programs is based on all programs but excluding data from programs medicine and rehabilitation science and physiotherapy.

Table 7

Comparison of the Prediction Errors Elicited by the Gender Specific Models on Gender Specific Data

Program	t-statistic [a]	Cohen's d [b]	R^2 Difference [c]
Across programs [d]	t(4568) = 3.02**	0.09	.04
1 – PSY	t(695) = 1.38	0.15	.21
2 – MED	t(275) = 0.10	0.01	.06
3 – COM SCI	t(153) = -2.16*	0.40	.14
4 – POL SCI	t(119) = 0.07	0.01	-.01
5 – LAW	t(229) = -0.30	0.04	.11
6 – CRIM	t(240) = -0.09	0.01	.04
7 – LING	t(241) = -4.55***	0.51	.34
8 – VET MED	t(226) = -1.04	0.17	.10
9 – REHAB SCI	t(460) = 2.50*	0.24	-.03
10 – PHARMA	t(305) = 1.54	0.23	-.12
11 – BIOSCI	t(331) = 1.23	0.14	-.04
12 – ECON	t(489) = 1.83	0.17	.05
13 – ENG	t(336) = -1.88	0.27	.27
14 – COMMER SCI	t(375) = 1.05	0.11	.06
15 – ENG TECH	t(439) = 3.33**	0.42	-.12
16 – APL LING	t(181) = -3.81***	0.51	.38

[a] A positive t-value indicates a higher mean prediction error for the female specific model, while a negative t-value indicates a higher mean prediction error for the male specific model. [b] The Cohen's d measure indicates the size of the difference in mean prediction errors. [c] The explained sample variance is explained by Nagelkerke's adjusted R^2. 'R^2 Difference' is the R^2 of the male model minus the R^2 of the female model. [d] The data across programs is based on all programs but excluding data from medicine and rehabilitation science and physiotherapy.

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

Table 8

Comparison of Prediction Errors of the General and Gender-specific Models

Programs	Male data: general vs. gender-specific model			Female data: general vs. gender-specific model		
	t-statistic [a]	Cohen's *d* [b]	R^2 Difference [c]	*t*-statistic [a]	Cohen's *d* [b]	R^2 Difference [c]
Across programs [d]	*t*(1872) = 5.77***	0.13	-.02	*t*(2696) = 14.01***	0.27	-.02
1 – PSY	*t*(96) = 3.29**	0.33	-.27	*t*(599) = 0.78	0.03	-.01
2 – MED	*t*(109) = 0.79	0.08	-.03	*t*(166) = -1.62	0.13	.01
3 – COM SCI	*t*(38) = 1.42	0.23	-.06	*t*(115) = -1.29	0.12	.00
4 – POL SCI	*t*(66) = 2.18*	0.27	-.14	*t*(53) = 1.12	0.15	-.12
5 – LAW	*t*(82) = 1.12	0.12	-.06	*t*(147) = 0.19	0.02	-.03
6 – CRIM	*t*(49) = 2.03*	0.29	-.24	*t*(191) = 2.21*	0.16	-.05
7 – LING	*t*(56) = 3.89***	0.52	-.26	*t*(185) = 1.24	0.09	-.04
8 – VET MED	*t*(48) = 2.57*	0.37	-.27	*t*(178) = 0.63	0.05	-.01
9 – REHAB SCI	*t*(165) = 1.76	0.14	-.05	*t*(295) = 1.84	0.11	-.01
10 – PHARMA	*t*(55) = 0.61	0.08	-.22	*t*(250) = 0.38	0.02	.00
11 – BIOSCI	*t*(170) = 0.84	0.06	-.01	*t*(161) = 1.12	0.09	-.05
12 – ECON	*t*(310) = 0.79	0.05	-.02	*t*(179) = 1.25	0.09	-.08
13 – ENG	*t*(268) = 12.24***	0.75	-.03	*t*(68) = 5.57***	0.67	-.05
14 – COMMER SCI	*t*(170) = 1.39	0.11	-.07	*t*(205) = 0.58	0.04	-.04
15 – ENG TECH	*t*(396) = 0.07	0.00	-.10	*t*(43) = 2.34*	0.35	-.22
16 – APL LING	*t*(46) = 3.27**	0.48	-.14	*t*(135) = -2.03*	0.18	.03

[a] A positive *t*-value indicates a higher mean prediction error for the general model, while a negative *t*-value indicates a higher mean prediction error for the gender-specific model. [b] The Cohen's *d* measure indicates the size of the difference in mean prediction errors. [c] The explained sample variance is explained by Nagelkerke's adjusted R^2. 'R^2 Difference' is the R^2 of the general model on the gender-specific data minus the R^2 of the gender-specific model on the

gender-specific data. [d] The data across programs is based on all programs but excluding data from medicine and rehabilitation science and physiotherapy.

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

Table 9

Comparison of the R^2 of the General and Gender-specific Model

Program	R^2 Difference [a]	
	General (data across genders) vs. male model (male data)	General (data across genders) vs. female model (female data)
Across programs [b]	-.02	.03
1 – PSY	-.14	.07
2 – MED	-.04	.03
3 – COM SCI	-.09	.05
4 – POL SCI	-.07	-.08
5 – LAW	-.07	.04
6 – CRIM	-.05	-.01
7 – LING	-.32	.02
8 – VET MED	-.12	-.02
9 – REHAB SCI	.02	-.01
10 – PHARMA	.08	-.03
11 – BIOSCI	.01	-.03
12 – ECON	-.01	.04
13 – ENG	-.05	.22
14 – COMMER SCI	-.05	.01
15 – ENG TECH	.12	-.01
16 – APL LING	-.29	.08

[a] The explained sample variance is explained by Nagelkerke's adjusted R^2. 'R^2 Difference' is the R^2 of the general model on the data across genders minus the R^2 of the gender-specific model on the gender-specific data. [b] The data across programs is based on all programs but excluding data from medicine and rehabilitation science and physiotherapy.

Appendix A

Measures: Description and Reliability

GPA was used as the dependent variable (*M* = 489.72, *SD* = 205.53). This variable ranges from 0 to 1000 and is used to provide an indication of a student's academic achievement across courses. Because all students of a specific study program have identical curricula and that each program is modeled independently, possible bias due to program-specific factors is excluded from the present study.

Seven demographic variables were included as predictors. First, we incorporated the students' age at the start of higher education (M = 18.28, SD = 1.70). Second, we included both language (1 = Dutch/2 = French/3 = other (EU)/4 = other (non-EU)) and nationality (0 = not Belgian/1 = Belgian). Third, for socio-economic status (SES) we included four variables, as SES is a multi-dimensional construct. If only a single component of SES had been included, its effect could have been severely overestimated (Sirin, 2005). We included a dichotomous measure that quantified whether a student belongs to a group of low SES or not. The categorization into a low SES-group depended on whether the student met any other following criteria: receiving a scholarship or having a mother who did not obtain a degree from secondary education. This procedure is based on the practices of the Flemish Department of Education (Ministerie van Onderwijs en Vorming, 2012). Furthermore, to allow for more specificity, we also included the measures that made up the SES-variable separately: we considered whether the student was given a scholarship or not, and the *degree of the mother and father*. For these last variables, a score between 0 (no degree) and 9 (PhD) was assigned. There is no need for concern regarding the multicollinearity of these predictors, as the Variance Inflation Factor (VIF) values of each predictor included in the models were always below 10.

Three variables were used to comprise study background. We opted to include self-reported *high school GPA* (*M* = 72.07, *SD* = 6.96), *high school Dutch package* (*M* = 4.01, *SD* = 1.41) and *high school mathematics package* (*M* = 5.03, *SD* = 2.42).

For cognitive ability, five factors were included in the present study which were all scored on a scale of 20. The tests used are all valid measures of academic achievement (XXXX et al., 2017; XXXX et al., 2015; XXXX et al., 2022). The first test is *vocabulary* (M = 17.50, SD = 1.89, Chronbach's α = .79), which we tested using the LexTALE test (Lemhöfer & Broersma, 2012). This is a short, objective test that consists of 60 items where students are asked to indicate whether the stimulus on screen is an actual word or not. The second factor is *comprehensive reading* (M = 14.81, SD = 4.707, Chronbach's α = .78). For this, students had to read an English text of medium length about a social psychological project and were asked five multiple choice (MC) questions about it afterwards. This test was not given to students in medicine or rehabilitation science and physiotherapy. The third factor is *mathematics*, which was tested with a normal (M = 16.57, SD = 3.30, Chronbach's α = .70) and an advanced test (M = 11.18, SD = 4.30, Chronbach's α = .96). The normal test included 20 questions on basic mathematics, both in MC-format and as open questions. An example of an item is "A book that is on a 40% discount costs €18. How much did it cost prior to the discount?". The advanced mathematics test also included 20 items, included both MC and open questions. These items were more advanced than those of the normal test, including questions such as "Present the general equation of a circle with center (-2, 1) and radius 3". The advanced mathematics test was used in the programs medicine, bioscience, economy, engineering, commercial science and engineering technology. The normal mathematics test was used in the remaining programs. The fourth factor was a *physics* test (M = 11.76, SD = 3.71, Chronbach's α = .96), where students had to solve 20 MC questions such as "What is Newton's first law?". This test was filled out by students in veterinary medicine, rehabilitation science and physiotherapy, and pharmaceutical science. Finally, the fifth factor was a *chemistry* test, given to students in veterinary medicine, rehabilitation science and physiotherapy, pharmaceutical science and bioscience (M = 14.72, SD = 3.56, Chronbach's α = .98). 20 questions in the MC format were administered, with items like "What is the total number of valence electrons of a sulfur atom?".

The Self-Regulation Questionnaire was used to estimate a score of 16 items for both *autonomous* and *controlled motivation* on a scale from 0 to 20 (Vansteenkiste et al., 2009). Eight items such as "I'm motivated to study this program because it interests me" measured autonomous motivation (*M* = 15.03, *SD* = 2.47, Chronbach's α = .86). Controlled motivation (*M* = 8.28, *SD* = 3.15, Chronbach's α = .88) was also measured using eight items, for example: "I'm motivated to study this program because I'm supposed to do this".

Test anxiety (*M* = 9.99, *SD* = 2.50, Chronbach's α = .92) was assessed using the Cognitive Test Anxiety Scale (Cassady & Johnson, 2002), which includes 25 items like "I do not perform well on exams". Students had to indicate how characteristic they found these statements on a scale from one to four (totally not characteristic for me - totally characteristic for me).

Conscientiousness (*M* = 150.22, *SD* = 20.65, Chronbach's α = .86) was assessed with the Personality for Professionals Inventory (de Fruyt & Rolland, 2010). 48 items were administered, on which students had respond on a 1 to 5 scale (not characteristic at all - very characteristic). An example of an item is "I easily procrastinate".

The Metacognitive Awareness Inventory (Schraw & Dennison, 1994) was used to estimate *metacognition knowledge* (*M* = 13.65, *SD* = 2.10, Chronbach's α = .86) and *regulation* (*M* = 13.01, *SD* = 2.01, Chronbach's α = .93). 52 items were presented to the students, for which they had to indicate to which degree they agreed with them on a 1 to 6 scale (completely disagree- completely agree). Afterwards, scores were rescaled to a final score between 0 and 20. Items such as "I know my intellectual strengths and weaknesses" were included.

Grit (*M* = 13.33, *SD* = 1.85, Chronbach's α = .73) was assessed through the Grit Scale (Duckworth et al., 2007). The scale includes 20 items such as "I finish whatever I begin", that need to be rated on a scale from 1 to 5 (not at all like me - very much like me).

The Brief Self-control Scale (Tangney et al., 2004) was used to assess *self-control* (*M* = 13.01, *SD* = 1.88, Chronbach's α = .74). 13 items were presented to the students, for which they had to indicate on a 1 to 5 scale (totally not agree - totally agree) how much they agreed with the statements. An example of an item is: “I have difficulty concentrating”.

To assess *academic self-efficacy comprehension* (M = 14.77, SD = 1.68, Chronbach’s α = .80) and *academic self-efficacy effort* (M = 15.16, SD = 1.96, Chronbach’s α = .77), an adaptation of the College Academic Self-Efficacy Scale was used (Owen & Froman, 1988). For this, students had to appraise their competence in coping with situations or tasks (such as “Tutor another student”) on a 1 to 5 scale (not capable - fully capable). Fourteen items were administered to estimate comprehension, and eight to estimate effort.

The XXXX[5] questionnaire (XXXX et al., 2017) was used to assess the vocational interests of students. 172 yes-or-no items were presented to the students, that each loaded on one of the RIASEC scales. The items in the questionnaire were either occupations like “lawyer” (loading on the E-scale) or tasks such as “composing a piece of music” (loading on the A-scale). Scores ranging from 0 to 100 were assigned to students on all the RIASEC scales: the R (M = 19.07, SD = 24.16, Chronbach’s α = .92), I (M =34.09, SD = 21.78, Chronbach’s α = .88), A (M = 29.65, SD = 25.60, Chronbach’s α = .92), S (M = 35, SD =26.34, Chronbach’s α = .92), E (M = 33.51, SD = 28.44, Chronbach’s α = .93) and C (M = 21.30, SD = 23.29, Chronbach’s α = .91) dimension.

[5] The name of this questionnaire also needed to be masked for blinded review.

Appendix B:

Correlation Tables

Table B**1**

Correlations for All Variables, Excluding Chemistry and Physics

Variable	1	2	3	4	5	6	7	8
1. Comprehensive reading	-							
2. Scholarship [a]	-0.06***	-						
3. Diploma parent 1 [b]	0.08***	-0.27***	-					
4. Diploma parent 2 [b]	0.10***	-0.27***	0.54***	-				
5. Grit	-0.03	0.00	-0.06***	-0.07***	-			
6. Gender [c]	0.01	0.03	-0.07***	-0.07***	0.14***	-		
7. SES [d]	-0.06***	0.84***	-0.33***	-0.44***	0.01	0.03*	-	
8. GPA	0.16***	-0.10***	0.14***	0.17***	0.02	0.09***	-0.11***	-
9. Age at start HE	-0.03*	0.14***	-0.10***	-0.13***	0.04*	0.00	0.14***	-0.16***
10. Autonomous motivation	0.06***	0.01	-0.01	-0.01	0.33***	0.22***	0.01	0.10***
11. Controlled motivation	-0.03*	0.01	0.03*	0.02	-0.18***	-0.08***	0.01	0.03*
12. Nationality [e]	-0.01	-0.07***	0.05***	0.09***	-0.04**	-0.04*	-0.05***	0.07***
13. Conscientiousness	0.01	-0.01	-0.02	-0.04*	0.62***	0.10***	-0.01	0.16***
14. HS GPA	0.15***	-0.06***	0.06***	0.07***	0.13***	0.11***	-0.06***	0.41***
15. Metacognition (knowledge)	0.09***	-0.01	0.01	0.03	0.37***	0.04**	-0.02	0.09***
16. Metacognition (regulation)	0.06***	-0.00	0.00	0.01	0.37***	0.11***	-0.00	0.08***
17. Language [f]	-0.07***	0.18***	-0.15***	-0.20***	-0.01	0.03*	0.19***	-0.13***
18. Test anxiety	-0.12***	0.05**	-0.06***	-0.09***	-0.24***	0.15***	0.06***	-0.15***
19. HS Dutch package	0.01	-0.02	0.01	0.00	0.01	0.05***	-0.02	0.05***
20. HS mathematics package	0.06***	-0.07***	0.16***	0.16***	-0.05***	-0.29***	-0.08***	0.15***
21. Vocabulary	0.15***	-0.07***	0.11***	0.10***	0.05***	-0.03*	-0.05***	0.16***
22. Self-control	0.01	-0.03	-0.03	-0.03*	0.60***	0.15***	-0.03	0.13***
23. Academic SE (comprehension)	0.11***	-0.01	0.06***	0.09***	0.23***	-0.17***	-0.02	0.08***
24. Academic SE (effort)	0.01	-0.02	-0.04*	-0.02	0.45***	0.16***	-0.02	0.16***

	1	2	3	4	5	6	7	8
25. R	0.00	-0.02	0.09***	0.09***	-0.09***	-0.50***	-0.02	-0.05***
26. I	0.10***	0.01	0.06***	0.05***	0.03*	-0.04**	0.02	0.03*
27. A	0.08***	-0.01	0.02	0.05**	-0.11***	0.18***	0.01	0.01
28. S	-0.00	0.01	-0.09***	-0.08***	0.06***	0.40***	0.03*	0.06***
29. E	-0.01	-0.03	0.01	0.01	0.03*	-0.15***	-0.02	0.01
30. C	-0.07***	0.01	-0.03*	-0.03	0.05**	-0.18***	0.00	-0.00
31. Mathematics [g]	0.16**/ 0.20***	-0.05**/ -0.07**	0.10***/ 0.13***	0.09***/ 0.15**	-0.03/ 0.03	-0.08 ***/ -0.14***	-0.07***/ -0.08***	0.22***/ 0.23***
	9	10	11	12	13	14	15	16
9. Age at start HE	-							
10. Autonomous motivation	0.06***	-						
11. Controlled motivation	-0.02	0.04*	-					
12. Nationality [e]	-0.18***	-0.04**	0.02	-				
13. Conscientiousness	0.00	0.45***	-0.08***	-0.02	-			
14. HS GPA	-0.09***	0.18***	0.01	0.05**	0.31***	-		
15. Metacognition (knowledge)	-0.00	0.44***	-0.03*	-0.04**	0.55***	0.20***	-	
16. Metacognition (regulation)	0.03	0.46***	-0.01	-0.04**	0.56***	0.17***	0.76***	-
17. Language [f]	0.14***	0.02	0.05**	-0.22***	-0.02	-0.05***	0.02	0.02
18. Test anxiety	0.04**	-0.11***	0.22***	-0.03	-0.30***	-0.24***	-0.31***	-0.14***
19. HS Dutch package	-0.05**	-0.00	-0.01	0.06***	0.02	-0.01	0.01	0.00
20. HS mathematics package	-0.14***	-0.05***	0.06***	0.06***	0.02	0.12***	0.02	-0.02
21. Vocabulary	-0.05***	0.05**	-0.05**	0.08***	0.08***	0.12***	0.08***	0.06***
22. Self-control	-0.03	0.34***	-0.15***	-0.01	0.73***	0.24***	0.39***	0.41***
23. Academic SE (comprehension)	-0.02	0.32***	-0.00	-0.01	0.36***	0.17***	0.52***	0.42***
24. Academic SE (effort)	-0.02	0.45***	-0.07***	-0.02	0.59***	0.25***	0.48***	0.49***
25. R	-0.05***	-0.08***	0.05***	0.04**	-0.03	-0.01	-0.00	-0.02
26. I	-0.01	0.21***	0.01	-0.01	0.10***	0.11***	0.13***	0.13***
27. A	0.02	0.07***	0.00	-0.01	-0.05***	0.02	0.03	0.05**
28. S	0.06***	0.16***	-0.03*	-0.02	0.04**	-0.00	0.06***	0.10***
29. E	-0.02	0.05***	0.11***	0.02	0.10***	-0.04**	0.05***	0.04**
30. C	-0.02	0.03*	0.12***	0.03	0.11***	-0.06***	0.03	0.03

	9	10	11	12	13	14	15	16
31. Mathematics [g]	-0.07**/ -0.11***	-0.01/ 0.12***	0.01/ 0.02	0.02/ 0.06**	0.03/ 0.09***	0.09***/ 0.30***	0.02/ 0.14***	0.03/ 0.07**
	17	18	19	20	21	22	23	24
17	-							
18	0.05**	-						
19	-0.01	-0.00	-					
20	-0.06***	-0.10***	-0.04**	-				
21. Vocabulary	-0.12***	-0.14***	0.04*	0.06***	-		-	
22. Self-control	0.01	-0.28***	0.01	-0.02	0.09***	-		
23. Academic SE (comprehension)	-0.01	-0.37***	-0.00	0.26***	0.11***	0.21***	-	
24. Academic SE (effort)	-0.01	-0.22***	0.02	-0.02	0.04**	0.51***	0.44***	-
25. R	-0.02	-0.08***	-0.06***	0.48***	0.00	-0.08***	0.21***	-0.07***
26. I	0.00	-0.04*	-0.00	0.28***	0.02	0.04**	0.22***	0.08***
27. A	0.04**	0.06***	0.03	-0.22***	0.02	-0.04**	-0.01	-0.04*
28. S	0.02	0.13***	0.04**	-0.34***	0.03*	0.04**	-0.06***	0.07***
29. E	0.01	-0.03	0.04*	-0.05***	-0.02	0.02	0.07***	-0.01
30. C	0.02	-0.00	0.03	0.09***	0.01	0.04*	0.09***	0.04*
31. Mathematics [g]	-0.05*/ -0.08***	-0.06**/ -0.16***	0.04*/ 0.00	0.34***/ 0.53***	0.21*** /0.16***	0.02/ 0.05*	0.20***/ 0.30***	0.02/ 0.06**
	25	26	27	28	29	30	31	
25. R	-							
26. I	0.37***	-						
27. A	0.02	0.14***	-					
28. S	-0.29***	0.16***	0.43***	-				
29. E	0.09***	-0.04*	0.18***	0.16***	-			
30. C	0.20***	0.09***	-0.03*	0.03	0.70***	-		
31. Mathematics [g]	0.10***/ 0.25***	0.16***/ 0.28***	-0.01/ -0.06*	0.03/ -0.14***	-0.04/ -0.24***	0.05*/ -0.19***	-	

Note. This correlation matrix was based on data from all programs, but medicine and rehabilitation science and physiotherapy (n = 4,393). We excluded data from medicine and rehabilitation science and physiotherapy , as these programs do not include data on comprehensive reading. SES =

socioeconomic status, GPA = Grade Point Average, HE = higher education, HS = high school, Academic SE = academic self-efficacy, R = realistic interest dimension, I = investigative interest dimension, A = artistic interest dimension, S = social interest dimension, E = enterprising interest dimension, C = conventional interest dimension.

[a] 0 = no scholarship and 1 = scholarship. [b] Score between 0 (no degree) and 9 (PhD). [c] 1 = male and 2 = female. [d] 0 = student does not belong to low SES-group and 1 = student belongs to low SES-group. [e] 0 = not Belgian and 1 = Belgian. [f] 1 = Dutch, 2 = French, 3 = other (EU) and 4 = other (non-EU). [g] Correlations reported before the slash are based on the normal mathematics test (n = 2,413). Correlations after the slash are based on the advanced mathematics test (n = 1,980).

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

Table B2

Correlations for Chemistry and Physics

Variable	1	2
1.Chemistry	-	0.49***
2. Physics	0.49***	-
3. Scholarship [a]	-0.05	-0.11***
4. Diploma parent 1 [b]	0.06	0.12***
5. Diploma parent 2 [b]	0.08**	0.17***
6. Grit	0.01	-0.06*
7. Gender [c]	-0.01	-0.11***
8. SES [d]	-0.06	-0.12***
9. GPA	0.28***	0.38***
10. Age at start HE	-0.08**	-0.14***
11. Autonomous motivation	0.11***	0.01
12. Controlled motivation	0.06	0.07*
13. Nationality [e]	-0.01	0.01
14. Conscientiousness	0.12***	0.03
15. HS GPA	0.21***	0.21***
16. Metacognition (knowledge)	0.10**	0.04
17. Metacognition (regulation)	0.11***	0.03
18. Language [f]	0.02	-0.04
19. Test anxiety	-0.05	-0.13***
20. HS Dutch package	0.02	0.07*
21. HS mathematics package	0.23***	0.27***
22. Vocabulary	0.14***	0.19***
23. Self-control	0.06	-0.01
24. Academic SE (comprehension)	0.15***	0.11***
25. Academic SE (effort)	0.12***	0.03
26. R	0.06*	0.12***
27. I	0.22***	0.23***
28. A	0.05	0.06*
29. S	-0.01	0.04
30. E	0.04	0.05
31. C	0.05	0.03
32. Mathematics	0.33***	0.34***

Note. This correlation matrix was based on data from veterinary medicine, rehabilitation science and physiotherapy, and pharmaceutical science (n = 1,003). We excluded data from the other programs, as these programs do not include data on both chemistry and physics. SES = socioeconomic status, GPA = Grade point average, HE = higher education, HS = high school, Academic SE = academic self-efficacy, R = realistic interest dimension, I = investigative interest dimension, A = artistic interest dimension, S = social interest dimension, E = enterprising interest dimension, C = conventional interest dimension.

[a] 0 = no scholarship and 1 = scholarship. [b] Score between 0 (no degree) and 9 (PhD). [c] 1 = male and 2 = female. [d] 0 = student does not belong to low SES-group and 1 = student belongs to low SES-group. [e] 0 = not Belgian and 1 = Belgian. [f] 1 = Dutch, 2 = French, 3 = other (EU) and 4 = other (non-EU).

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

Appendix C:

Gender- and Program-specific Models Predicting GPA.

Table C1

Gender- and Program-specific Models Predicting GPA

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
Across programs	Across genders	-301.02 + 2.66 x Comprehensive reading + 4.68 x Diploma parent 1 + 8.99 x Diploma parent 2 - 8.98 x Grit – 27.21 x Age at start HE + 3.51 x Mathematics + 2.49 x Controlled motivation + 0.74 x Conscientiousness + 8.61 x HS GPA + 6.60 x Metacognition knowledge – 20.91 x Language - 5.46 x Test anxiety + 56.58 x HS Dutch package + 16.75 x HS mathematics package + 7.15 x Vocabulary – 7.65 x Academic SE comprehension + 8.43 x Academic SE effort + 0.36 x R - 0.53 x I - 0.61 x S + 1.25 x C – 203.79 x Gender - 6.28 x Metacognition knowledge:Gender – 25.12 x HS Dutch package:Gender - 0.63 x R:Gender + 0.80 x S:Gender - 0.60 x C:Gender	.27	153.67
	Male	94.48 + 2.30 x Comprehensive reading + 6.84 x Diploma parent 2 - 5.80 x Grit - 37.80 x Age at start HE + 3.14 x Mathematics + 3.72 x Autonomous motivation + 3.93 x Controlled motivation + 8.81 x HS GPA – 31.50 x Language– 4.93 x Test anxiety + 28.31 x HS Dutch package + 17.86 x HS mathematics package + 7.61 x Vocabulary – 7.81 x Academic SE comprehension + 11.05 x Academic SE effort – 0.41 x R + 0.61 x E	.29	157.19
	Female	30.70 + 2.75 x Comprehensive reading + 4.43 x Diploma parent 1 + 11.91 x Diploma parent 2 - 12.22 x Grit - 18.81 x Age at start HE + 3.23 x Mathematics + 1.09 x Conscientiousness + 8.37 x HS GPA - 5.72 x Metacognition knowledge - 15.93 x Language - 5.82 x Test anxiety + 16.35 x HS	.24	150.65

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
		mathematics package + 7.67 x Vocabulary - 8.04 x Academic SE comprehension + 6.04 x Academic SE effort - 0.71 x R - 0.62 x I - 0.27 x A + 1.10 x S		
1 – Psychology	Across genders	-2467.80 + 24.76 x Comprehensive reading + 14.57 x Diploma parent 1 + 20.66 x HS GPA - 10.06 x Metacognition knowledge + 11.20 x Metacognition regulation - 29.47 x Language - 12.60 x Test anxiety + 228.97 x HS Dutch package + 25.73 x HS mathematics package + 11.52 x Mathematics + 8.39 x Vocabulary - 20.34 x Academic SE comprehension + 1173.09 x Gender - 10.71 x Comprehensive reading:Gender - 7.24 x HS GPA:Gender - 99.61 x HS Dutch package:Gender	.34	153.77
	Male	223.42 + 22.03 x Comprehensive reading + 27.88 x Grit - 92.54 x SES - 74.57 x Age at start HE + 26.85 x Autonomous motivation + 14.87 x HS GPA + 36.82 x Vocabulary - 54.26 x Self-control - 30.26 x Academic SE comprehension - 2.26 x C	.47	127.41
	Female	-193.63 + 3.17 x Comprehensive reading + 17.30 x Diploma parent 2 + 6.07 x HS GPA - 10.91 x Test anxiety + 26.85 x HS mathematics package + 13.84 x mathematics + 8.95 x Vocabulary - 22.64 x Academic SE comprehension + 8.18 x Academic SE effort	.27	112.12
2 – Medicine	Across genders	221.84 + 8.17 x HS GPA - 6.66 x Academic SE comprehension - 2.20 x A - 20.91 x Gender + 0.99 x A:Gender	.28	78.72
	Male	150.89 + 7.44 x HS GPA - 1.16 x A	.32	58.67
	Female	266.85 + 8.20 x HS GPA - 10.90 x Vocabulary	.25	58
3 – Communication science	Across genders	923.89 + 6.51 x Comprehensive reading + 18.19 x Diploma parent 1 + 12.57 x Autonomous motivation + 12.47 x HS GPA - 19.90 x Metacognition regulation + 102.79 x HS mathematics package + 155.29 x Gender - 43.05 x HS mathematics package:Gender	.28	140.44

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
	Male	-82.39 + 10.87 x Comprehensive reading - 55.28 x SES + 24.53 x Autonomous motivation + 1.14 x Test anxiety + 66.56 x HS mathematics package - 17.53 x Self-control - 1.80 x I + 0.66 x E	.36	81.51
	Female	-483.73 + 23.58 x Diploma parent 1 + 14.62 x HS GPA - 17.54 x Metacognition regulation	.22	117.53
4 – Political science	Across genders	774.26 + 34.58 x Diploma parent 2 - 29.64 x Grit - 101.41 x Age at start HE + 16.64 x Autonomous motivation + 2.11 x Conscientiousness + 14.98 x HS GPA - 77.64 x Language + 47.28 x HS mathematics package - 1.54 x C	.52	139.96
	Male	413.29 - 22.82 x Diploma parent 1 + 44.81 x Diploma parent 2 - 25.45 x Grit - 124.84 x Age at start HE + 37.31 x Autonomous motivation + 22.41 x HS GPA - 164.67 x Language + 198.16 x HS Dutch package + 37.83 x HS mathematics package - 28.34 x Academic SE comprehension	.59	101.38
	Female	-1342.29 + 33.20 x Diploma parent 1 + 41.16 x Diploma parent 2 + 32.25 x Autonomous motivation + 6.10 x Conscientiousness + 15.19 x HS GPA - 23.85 x Metacognition knowledge - 56.91 x Academic SE effort	.60	100.57
5 – Law	Across genders	170.36 - 11.91 x Grit - 92.47 x Age at start HE + 15.96 x HS GPA + 25 x Metacognition knowledge + 22.38 x HS mathematics package + 7.72 x Mathematics + 12.88 x Self-control + 16.76 x Academic SE effort + 292.34 x Gender - 22.57 x Metacognition knowledge:Gender	.49	111.71
	Male	1302.86 - 136.79 x Age at start HE + 8.47 x Controlled motivation + 15.08 x HS GPA + 24.92 x HS mathematics package + 24.14 x Academic SE effort	.56	88.05
	Female	-753.57 + 15.21 x HS GPA - 21.81 x Metacognition knowledge - 9.85 x Test anxiety + 25.93 x HS mathematics package + 8.75 x Mathematics + 16.15 x Academic SE effort	.46	90.68
6 – Criminology	Across genders	796.99 - 14.55 x Grit + 2.65 x Conscientiousness - 4.66 x HS GPA - 63.37 x Language + 185.63 x HS Dutch package + 26.21 x HS mathematics package - 72.49 x Academic SE comprehension + 0.81 x S	.34	128.02

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
		- 589.68 x Gender + 6.44 x HS GPA:Gender - 72.18 x HS Dutch package:Gender + 31.44 x Academic SE comprehension:Gender		
	Male	324.22 + 37.79 x Metacognition knowledge - 36.69 x Metacognition regulation + 102.45 x HS Dutch package + 31.49 x HS mathematics package - 36.68 x Academic SE comprehension + 2.45 x E	.40	96.03
	Female	683.48 + 4.65 x Comprehensive reading - 19.08 x Grit - 48.41 x Age at start HE + 3.49 x Conscientiousness + 6.56 x HS GPA - 11.16 x Metacognition knowledge - 50.66 x Language + 18.27 x HS mathematics package - 2.02 x R + 0.88 x S	.36	97.05
7 – Linguistics	Across genders	418.73 + 19.97 x Diploma parent 2 + 15.28 x Autonomous motivation + 14.94 x HS GPA – 317.89 x HS Dutch package – 936.15 x Gender + 223.79 x HS Dutch package:Gender	.44	131.03
	Male	-468.13 - 39.48 x Grit + 25.84 x Autonomous motivation + 14.66 x HS GPA - 83.49 x Language - 79.57 x HS Dutch package + 20.51 x Mathematics - 27.04 x Academic SE comprehension + 35.38 x Academic SE effort + 3.85 x R - 2.17 x S + 2.69 x E - 3.16 x C	.75	61.76
	Female	-1342.37 + 6.24 x Comprehensive reading + 121.93 x Scholarship + 16.16 x Diploma parent 2 - 133.16 x SES + 15.57 x Autonomous motivation + 13.39 x HS GPA + 111 x HS Dutch package	.42	101.81
8 – Veterinary medicine	Across genders	-653.05 + 13.22 x Diploma parent 1 + 17.83 x Fysica_score_20 + 8.28 x HS GPA + 14.75 x Metacognition regulation + 29.66 x HS mathematics package + 13.51 x Self-control - 21.55 x Academic SE comprehension	.36	151.83
	Male	-580.04 + 23.02 x Physics + 29.94 x Controlled motivation - 291.02 x Nationality + 45.64 x HS mathematics package + 46.57 x Self-control	.48	107.33

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
	Female	-789.38 + 15.67 x Diploma parent 1 + 14.34 x Physics + 9.39 x HS GPA + 20.06 x Metacognition regulation + 30.68 x HS mathematics package + 15.79 x Vocabulary - 25.71 x Academic SE comprehension	.38	121.27
9 – Rehabilitation sciences and physiotherapy	Across genders	301.15 + 12.75 x Physics – 73.37 x Age at start HE + 0.96 x Conscientiousness + 10.78 x HS GPA + 125.37 x HS Dutch package + 24.09 x HS mathematics package – 17.04 x Academic SE comprehension + 321.88 x Gender – 69.65 x HS Dutch package:Gender	.35	145.02
	Male	-559.44 + 16.56 x Diploma parent 2 + 11.92 x Physics + 13.23 x HS GPA - 74.21 x Language + 28.65 x HS mathematics package - 16.62 x Academic SE comprehension	.32	127.46
	Female	1938.39 + 13.48 x Physics - 113.87 x Age at start HE + 9.96 x HS GPA - 7.35 x Test anxiety - 12.51 x HS Dutch package + 17.81 x HS mathematics package - 14.38 x Academic SE comprehension	.35	106.59
10 – Pharmaceutical science	Across genders	-207.70 + 11.44 x Diploma parent 2 - 28.83 x Grit + 10.35 x Fysica_score_20 + 2.22 x Conscientiousness + 8.15 x HS GPA - 43.39 x Language + 27.02 x Test anxiety + 15.96 x HS mathematics package - 17.40 x Academic SE comprehension + 205.74 x Gender - 22.05 x Test anxiety:Gender	.32	146.83
	Male	2241.99 - 147.40 x Age at start HE + 13.74 x Controlled motivation + 12.06 x HS GPA	.24	128.16
	Female	143.11 + 8.43 x Chemistry - 34.34 x Grit + 8.69 x Physics + 1.66 x Conscientiousness + 8.58 x HS GPA - 53.27 x Language - 15.56 x Test anxiety - 0.99 x A	.35	106.76
11 – Bioscience	Across genders	-1241.09 + 17.24 x Diploma parent 2 + 12.26 x HS GPA - 10.51 x Metacognition knowledge + 54.64 x HS Dutch package + 9.53 x Mathematics + 15.89 x Vocabulary + 15.21 x Academic SE effort	.32	143.15

Program(s)	Gender	Model	$R^{2\,a}$	GPA error [b]
	Male	-1049.88 + 24 x Diploma parent 2 + 13.28 x HS GPA - 17.76 x Metacognition knowledge + 62.61 x HS Dutch package + 8.55 x Mathematics + 18.62 x Academic SE effort	.31	115.15
	Female	-1067.78 + 13.54 x Diploma parent 1 + 12.52 x HS GPA + 20.74 x Vocabulary + 8.34 x Mathematics + 1.44 x I - 1.64 x S + 1.40 x E	.34	103.22
12 – Economy	Across genders	367.55 + 3.03 x Comprehensive reading + 10.02 x Diploma parent 2 – 59.96 x Age at start HE + 13.52 x HS GPA + 12.53 x HS mathematics package + 8.40 x Mathematics	.36	137.07
	Male	669.33 - 73.78 x Age at start HE + 14.29 x HS GPA + 11.92 x HS mathematics package + 9.09 x Mathematics	0.36	111.56
	Female	-75.54 -15.13 x Grit + 11.16 x HS GPA - 11.60 x Test anxiety + 12.42 x Mathematics	.32	96.69
13 – Engineering	Across genders	-78.29 + 16.80 x Diploma parent 2 - 61.03 x Age at start HE + 7.86 x Autonomous motivation + 10.85 x HS GPA + 21.04 x HS mathematics package + 36.66 x Mathematics + 212.12 x Gender - 17.04 x Mathematics:Gender	0.33	140.82
	Male	116.66 + 16.31 x Diploma parent 2 - 64.17 x Age at start HE + 10.47 x Autonomous motivation + 10.81 x HS GPA + 23.96 x HS mathematics package + 20.49 x Mathematics + 0.96 x E - 1.21 x C	.39	103.86
	Female	-340.32 + 11.21 x HS GPA	.11	127.37
14 – Commercial science	Across genders	284.50 - 15.24 x Diploma parent 2 - 78.43 x Age at start HE + 12.29 x HS GPA + 18.25 x HS mathematics package + 5.69 x Mathematics + 44.17 x Vocabulary + 278.13 x Gender + 18.08 x Diploma parent 2:Gender - 21.43 x Vocabulary:Gender	.30	129.37
	Male	-14.93 - 65.13 x Age at start HE + 17.09 x HS GPA - 64.15 x Language + 34.25 x HS mathematics package + 25.60 x Vocabulary + 1.49 x R	.35	103.51

Program(s)	Gender	Model	$R^{2\,a}$	GPA error[b]
	Female	1373.31 + 20.23 x Diploma parent 2 - 87.53 x Age at start HE + 11.03 x HS GPA - 13.20 x Metacognition knowledge + 5.20 x Mathematics	.30	95.02
15 – Engineering technology	Across genders	809.19 - 12.33 x Grit - 52.19 x Age at start HE + 4.87 x HS GPA - 12.6 x Test anxiety + 22.03 x HS mathematics package + 8.54 x Mathematics + 15.86 x Academic SE effort + 63.41 x Gender	.30	129.37
	Male	900.90 - 11.67 x Grit - 49.81 x Age at start HE + 4.71 x HS GPA - 70.87 x Language - 12.88 x Test anxiety + 21.43 x HS mathematics package + 8.46 x Mathematics + 16.75 x Academic SE effort	.19	125.66
	Female	-775.56 - 77.04 x Scholarship + 136.35 x SES + 24.08 x Autonomous motivation + 14.24 x HS GPA - 47.76 x Metacognition regulation + 33.87 x Academic SE comprehension	.31	88.56
16 – Applied linguistics	Across genders	-839.26 + 9.01 x Comprehensive reading + 19.66 x Diploma parent 1 + 16.14 x HS GPA + 1.42 x S - 1.4092 x E - 71.11 x Gender	0.34	147.56
	Male	-1300.55 + 9.56 x Comprehensive reading + 29.19 x Diploma parent 1 + 17.86 x HS GPA + 57.24 x HS mathematics package - 26.95 x Academic SE comprehension + 23.18 x Academic SE effort + 1.67 x S	.63	80.56
	Female	-813.71 + 7.05 x Comprehensive reading + 16.28 x HS GPA	.25	126.29

Note. Variance Inflation Factor (VIF) values of the predictors were all below 10, indicating that the multicollinearity levels between the predictors lied within the acceptable range (Stevens, 2012). HS = high school, HE = higher education, Academic SE = Academic self-efficacy, R = realistic interest dimension, I = investigative interest dimension, A = artistic interest dimension, S = social interest dimension, E = enterprising interest dimension, C = conventional interest dimension.

[a] The explained population variance is represented by Nagelkerke's R^2. [b] The GPA (grade point average) error is used to reflect the accurateness of the

model. It is the average absolute prediction error of the prediction model (i.e., the average of the difference between the actual GPA of each student and the GPA predicted by the model for that student).

APPENDIX D

Correspondence between the general and gender-specific prediction models

We explored the correspondence between the general model and the gender-specific models (see Table D1). For the male models, a mean correspondence of $M = 0.46$, with a standard deviation of $SD = 0.23$ was found. The correlation between these correspondence measures and the gender proportion of each study program (calculated as the percentage of male students, excluding the model across programs) was $r(14) = .89$, $p < .001$. The same was done for the female models, which resulted in a mean correspondence of $M = 0.42$ with a standard deviation of $SD = 0.17$. The correlation between these correspondence measures and the gender proportion of each study program (calculated as the percentage of female students, excluding the model across programs) was $r(14) = .77$, $p < .001$. In sum, if one gender is predominantly present in a study program, the composition of the prediction model of the dominant gender aligns better with the composition of the general model.

Table D1

Correspondence Between the General and Gender-specific Models of Academic Achievement

Program	Correspondence male and general model [a]	Correspondence female and general model [a]
Across programs [b]	.68	.77
1 – PSY	.26	.35
2 – MED	.40	.40
3 – COM SCI	.27	.50
4 – POL SCI	.58	.33
5 – LAW	.36	.45
6 – CRIM	.27	.50
7 – LING	.23	.57
8 – VET MED	.33	.67
9 – REHAB SCI	.40	.67
10 – PHARMA	.08	.50
11 – BIOSCI	.86	.27
12 – ECON	.67	.25
13 – ENG	.75	.17
14 – COMMER SCI	.50	.57

15 – ENG TECH	.88	.08
16 – APL LING	.50	.40

[a] The correspondence measures are calculated as the total amount of shared predictors between the two models, divided by the total amount of predictors present across both models. [b] The data across programs is based on all programs but excluding data from programs medicine and rehabilitation science and physiotherapy.