



Disentangling the Association Between the Big Five Personality Traits and Student Achievement: Meta-Analytic Evidence on the Role of Domain Specificity and Achievement Measures

Jennifer Meyer¹ · Thorben Jansen¹ · Nicolas Hübner² · Oliver Lüdtke¹

Accepted: 7 October 2022 / Published online: 31 January 2023
© The Author(s) 2023

Abstract

Students' academic achievement is a central predictor of a long list of important educational outcomes, such as access to higher education and socioeconomic success. Prior studies have extensively focused on identifying variables that are related to academic achievement and an important variable in this context appears to be students' personality. Notably, although findings from more recent studies suggested that the association between student achievement and personality varies by the subject domain (language vs. STEM) and the type of achievement measure (grades vs. test scores), systematic meta-analytical evidence is still lacking. To address this gap in the educational research literature, we conducted a meta-analysis based on 78 studies, with 1491 effect sizes representing data from 500,218 students and 110 samples from elementary to high school. We used a random-effects model with robust variance estimation to calculate mean effect sizes and standard deviations. We found moderating effects of measure or domain for all five personality traits, with differences in the direction of the effects. Our results highlight the importance of the domain and measure when examining how personality traits relate to academic achievement in school. The combination of subject domain and achievement was also found to be relevant for some of the traits. These findings emphasize that subject domains and types of achievement measures should be explicitly considered when investigating the personality saturation of student achievement. We discuss implications for future research, highlighting that there is no “best” or “more objective” achievement measure but, instead, that achievement measures should be chosen based on the research question of interest.

Keywords Meta-analysis · Academic achievement · Personality · Grades · Test scores · Subject domains

✉ Jennifer Meyer
jmeyer@leibniz-ipn.de

Extended author information available on the last page of the article

Students' academic achievement is a central predictor of many important life outcomes such as educational attainment, retention, post-school choices, work and life satisfaction, long-run earnings, and work performance (Melin et al., 2003; Mottaz, 1984; Spengler et al., 2018; Wilmot & Ones, 2019). Over the past decades, educational researchers have invested considerable effort in identifying important predictors of students' academic achievement. As a result of this, studies have repeatedly documented the relevance of students' personality in predicting academic achievement (Borghans et al., 2016; Hübner et al., 2022; Lechner et al., 2017; Meyer et al., 2019). Two extensive meta-analyses have provided important support for these findings (Mammadov, 2022; Poropat, 2009). Specifically, Poropat (2009) operationalized student achievement in terms of grade point averages (GPA) and Mammadov (2022) considered a blend of all available different achievement measures. However, the moderating role of achievement measures or subject domains was not considered in either prior meta-analysis.

This gap in the literature is somewhat surprising because prior research has emphasized that the correlation of different types of achievement measures (e.g., grades and test scores) is far from perfect (Borghans et al., 2016; Willingham et al., 2002) and that the association between student achievement and personality can substantially vary by subject domain and achievement measure (e.g., Brandt et al., 2020; Hübner et al., 2022; Lechner et al., 2017; Meyer et al., 2019; Tetzner et al., 2020). For example, several studies found larger associations between openness and standardized test scores compared to grades, whereas associations between conscientiousness and grades were larger compared to standardized test scores (e.g., Meyer et al., 2019; Spengler et al., 2013).

Furthermore, recent educational psychological work has developed and successfully tested a comprehensive framework to better explain differential associations between personality traits and different types of achievement measures. Based on this, meta-analytically investigating the potentially moderating influence of subject domains and types of achievement measures on the personality saturation of student achievement seems highly important in order to extend our understanding of how students' characteristics might differentially shape different measures of their achievement (Hübner et al., 2022).

Personality Traits and Academic Achievement

Personality traits describe “relatively enduring patterns of thoughts, feelings, and behaviors that reflect the tendency to respond in certain ways under certain circumstances” (Roberts, 2009, p. 140). The most frequently applied and studied taxonomies of personality traits are the five-factor model (FFM) and the Big Five, both of which contain neuroticism, extraversion, openness, conscientiousness, and agreeableness (John et al., 2008; McCrae & Costa, 1987; McCrae & John, 1992); these are the personality traits that we consider in this article. Please note that throughout the article, we use the terms FFM and Big Five synonymously.

The personality traits described in the FFM have been found to play a key role in student achievement (Borghans et al., 2016; De Raad & Schouwenburg, 1996; Israel et al., 2022). The personality saturation of achievement measures is typically

explained by different types of behavior that are beneficial for student learning and achievement. As one example, students who score high on conscientiousness more actively engage in course work and in doing their homework (Gray & Watson, 2002). In addition, the role of personality traits in achievement situations has been found to accumulate across the life span; researchers have hypothesized several (causal) pathways for these cumulative effects, explaining why achievement and personality traits are related (Caspi et al., 2005; see also De Raad & Schouwenburg, 1996; Roberts et al., 2007). What has been found to be most relevant for the developmental transition during adolescence is the process of active niche picking (Caspi et al., 2005; Roberts et al., 2007). It has been suggested that students choose educational experiences and environments whose qualities match their personalities (Lüdtke et al., 2011). Differential mechanisms can be assumed for each of the five traits to understand how personality traits and academic achievement are related (for detailed descriptions of why each individual trait can be related to achievement, see De Raad & Schouwenburg, 1996; Poropat, 2009). In the current study, we focused on how personality is reflected in different achievement measures and subject domains, as described in the next section.

Taking a Closer Look: The Relevance of Academic Subject Domains and Different Achievement Measures

Recent research has suggested that the role that personality traits play in student achievement largely depends on the academic subject domain and varies for different achievement measures (Brandt et al., 2020; Hübner et al., 2022; Lechner et al., 2017; Meyer et al., 2019).

Subject Domains

Considering subject domains when investigating the association between personality traits and achievement in educational research is important for at least two reasons. First, learning in school takes place in different subject domains, such as students' first and second languages, mathematics, and sciences, and domain-specific achievement shapes student career decisions even beyond subject-specific expectancy beliefs and interests (e.g., Guo et al., 2017; Nagy et al., 2008) and, thus, can influence the decision to pursue careers in different fields (e.g., STEM; Hübner et al., 2017; Jansen et al., 2015; Perez et al., 2014; Schoon & Eccles, 2014). In addition, a person's belief that they are either a "math person" or a "language person" can be relevant to student achievement and engagement (Wan et al., 2021). Using an overall GPA when investigating the association between students' personality and their achievement, as was done in prior meta-analyses on this topic, does not adequately reflect such domain specificities, as such an achievement measure is based on a blend of different subject grades that might count differently toward the GPA (Brookhart et al., 2016; Hübner et al., 2020). Second, taking a closer look at the specific patterns of the personality saturation of achievement measures on the level of subject domains was found to constitute a promising step

towards achieving a better understanding of the differential associations and underlying mechanisms. This is reflected in many empirical studies that have hypothesized and tested such domain-specific associations (e.g., Meyer et al., 2019; Spengler et al., 2013; Spinath et al., 2010). Vedel (2016) investigated the subject domain as a moderator of the association between personality traits and achievement in higher education. In her study, Vedel (2016) found that different traits were related more or less strongly to different academic majors in college, thus providing important evidence to support the idea that subject domain might play an important role in understanding how personality relates to student achievement.

Achievement Measures

Currently, a long list of different achievement measures exists, which typically distinguishes between grades and standardized test achievement (Willingham et al., 2002). As outlined in previous research, grades and standardized test scores are only moderately correlated, which means that they, at least in part, might depend on different student characteristics (Borghans et al., 2016; Hübner et al., 2022). A major difference between the two achievement measures lies in the process of assigning grades, which is based on students' work in class and is less objective/standardized than the process of assigning scores directly from standardized tests. This means that grades more strongly depend on subjective teacher evaluations than standardized test results do.

PASH (personality–achievement saturation hypothesis; Hübner et al., 2022) provides a comprehensive framework to describe differences between different achievement measures. In this framework, it is argued that personality manifests itself in a more pronounced way (“higher personality saturation”) under certain circumstances that are reflected in different achievement measures. PASH distinguishes between five different features of achievement measures that should be considered when studying the personality–achievement association in the field of education (i.e., standardization, relevance, curricular validity, instructional sensitivity, and cognitive ability saturation). Grades are typically characterized by a low level of standardization, a moderate to a high level of relevance, high curricular validity, high instructional sensitivity, and moderate to high cognitive ability saturation. Standardized tests, in contrast, are described as being highly standardized, lower in relevance, curricular validity, and instructional sensitivity, and high in cognitive ability saturation. In sum, the framework and the empirical findings of Hübner et al. (2022) underline the great importance of making a more fine-grained distinction between different types of achievement measures in order to better understand the differential associations between personality and student achievement.

Empirical Findings: An Overview of Each Personality Trait

Openness to Experience

Openness describes “the breadth, depth, originality, and complexity of an individual’s mental and experiential life” (John et al., 2008; p. 120). Meta-analytical results on GPA or composite scores have shown positive associations between openness

and student achievement ($r=0.10$; Poropat, 2009; $r=0.13$; Mammadov, 2022). However, empirical evidence suggests that there is a domain-specific pattern: openness can be assumed to reflect a more verbal and less of a numerical orientation (e.g., Marsh et al., 2006; Nofhle & Robins, 2007) and, thus, might be less beneficial for subject domains requiring the ability to revise default procedures or emphasize rules and regulations (see Gatzka & Hell, 2017) and might be more beneficial for subject domains that include rewards for critical thinking and creative ideas (i.e., language subject domains; Gatzka & Hell, 2017; Lipnevich et al., 2016). Some studies were able to support these assumptions, showing larger associations between openness and achievement in language compared to STEM subject domains (e.g., Hübner et al., 2022; Meyer et al., 2019; Spengler et al., 2013). However, there are some contrasting results, showing negative associations with language grades (e.g., Brandt et al., 2020; Hendriks et al., 2011; Westphal et al., 2020b).

Considering the role of achievement measures, standardized tests might include unfamiliar tasks that are less closely associated with the curriculum and are less instructionally sensitive. It could be assumed that students scoring high on openness might be able to better apply their cognitive prerequisites in such foreign assessment situations as they tend to seek out intellectually stimulating situations (Schwaba et al., 2017). Knowledge acquired during these intellectual free-time activities can enhance achievement in standardized tests (Willingham et al., 2002). Some prior findings support these assumptions, indicating a higher relevance of openness for standardized tests (e.g., Hübner et al., 2022; Meyer et al., 2019; Spengler et al., 2013). However, some studies found evidence that openness had negative effects on STEM achievement that was measured with tests (Meyer et al., 2019) and grades (Hendriks et al., 2011; Spengler et al., 2013).

Conscientiousness

Conscientiousness describes “socially prescribed impulse control that facilitates task- and goal-directed behavior, such as thinking before acting, delaying gratification, following norms and rules, and planning, organizing, and prioritizing tasks” (John et al., 2008, p. 120). Conscientiousness is the personality trait with the most substantial association with academic achievement, with a mean correlation of $r=0.19$ (Poropat, 2009; $r=0.20$; Mammadov, 2022). Associations of conscientiousness with grades are consistent for both language and STEM subject domains across studies (e.g., De Fruyt et al., 2008; Rosander et al., 2011). Its associations with academic effort and hard work make conscientiousness a beneficial trait for students in all areas of academic achievement (Nofhle & Robins, 2007; Trautwein et al., 2009). Nonetheless, even though generally positive effects were found across studies, some studies observed stronger associations for STEM compared to language subject domains (e.g., Brandt et al., 2020; Meyer et al., 2019). This could be due to the fact that these particularly challenging subject domains (i.e., STEM) require persistent learning behavior and analytical thinking to understand equations and solve problems (see Duckworth & Seligman, 2005; MacCann et al., 2009). However, some studies also found effects in the opposite direction, with higher effect sizes found for language than for STEM subject domains (e.g., Israel et al., 2019; Spengler et al., 2013).

PASH (Hübner et al., 2022) suggests that for achievement measures where teachers' personal preferences have less influence on the evaluation process, the role of conscientiousness should decrease as conscientiousness-related behavior that students show in class (e.g., doing homework, actively engaging in class) is not as relevant for highly standardized achievement measures as for achievement measures with low standardization (e.g., grades). Thus, conscientiousness is likely to be more closely related to measures that are low in standardization, as is typical of course grades (e.g., Brandt et al., 2020; Lechner et al., 2017; Nofle & Robins, 2007; Spengler et al., 2013). However, previous findings for standardized test scores are not consistent, as some studies obtained nonsignificant results and also found negative associations in both language and STEM subject domains (Meyer et al., 2019; Spengler et al., 2013; Westphal et al., 2020b).

Extraversion

Extraversion is “an energetic approach toward the social and material world and it includes traits such as sociability, activity, assertiveness, and positive emotionality” (John et al., 2008, p. 120). It can be assumed that associations of extraversion with achievement might depend on the social features of the learning contexts (De Raad & Schouwenbourg, 1996; Mammadov, 2022), which can vary across subject domains (Brandt et al., 2021; Meyer et al., 2019). For example, extraverted behavior in the classroom can be beneficial in language subject domains because oral language competencies can be a part of the assessments, but extraverted behavior could be distracting in STEM subject domains. Such effects might also depend on the achievement measures, as extraverted behavior is easily observable and might therefore influence grading but be less relevant for standardized tests. Some empirical studies were able to support these assumptions, showing higher associations of extraversion with grades (Brandt et al., 2021; Meyer et al., 2019) and in language subject domains (Brandt et al., 2020, 2021; Israel et al., 2019).

Agreeableness

Agreeableness is a factor that “contrasts a prosocial and communal orientation toward others with antagonism and includes traits such as altruism, tender-mindedness, trust, and modesty” (John et al., 2008, p. 120). It can be assumed that students' agreeable behavior in the classroom can be beneficial in influencing teachers' evaluations and, thus, grading. For example, agreeable students enjoy classroom discussions (Chamorro-Premuzic et al., 2007), and agreeableness relates to a preference for structure, cooperation, and social participation in the classroom (Pawlowska et al., 2014). Such effects may be less relevant for standardized tests. Prior studies have supported this assumption by finding larger effect sizes for grades than for test scores (e.g., Brandt et al., 2021). However, some studies also showed small positive associations with test scores (Brandt et al., 2021), whereas others found negative correlations (Meyer et al., 2019). With regard to a possible moderating effect of the subject domain, some studies have suggested positive associations between agreeableness and achievement in both

language (e.g., Brandt et al., 2021; Israel et al., 2019; Steinmayr & Spinath, 2008) and STEM (Dumfart & Neubauer, 2016) domains; others found negative correlations (Meyer et al., 2019; Spengler et al., 2013).

Neuroticism

Neuroticism contrasts “emotional stability and even-temperedness with negative emotionalities, such as feeling anxious, nervous, sad, and tense” (John et al., 2008, p. 120). Differential effects can be hypothesized regarding the role of achievement measures. Students who score higher on neuroticism take their schoolwork seriously to prevent making mistakes and they strive for perfectionism (Smith et al., 2019). However, students with higher scores on neuroticism perceive their environment as more stressful compared to more emotionally stable students; they are more vulnerable to stress (Ebstrup et al., 2011; McCrae, 1990; Murberg & Bru, 2007; Szabó, 2011; Uliaszek et al., 2010) and disengagement in coping (Carver & Connor-Smith, 2010), which is related to test anxiety (Hoferichter et al., 2014). Thus, neuroticism can be observed by teachers and incorporated into grading.

However, the direction of the associations is not fully clear. On the one hand, high neuroticism can lead to careful learning behavior given its association with perfectionism (Smith et al., 2019). On the other hand, nervousness can also lead to less accepted social behavior. For example, high neuroticism, which can be associated with higher anxiety, can lead to less involvement in classroom discussions and this can have detrimental effects on grades. Matching these contrasting hypotheses, empirical studies found that, for STEM grades, effect sizes tended to be negative (e.g., Israel et al., 2019) or nonsignificant (Meyer et al., 2019). However, some findings also indicated positive associations of neuroticism with language grades (e.g., Brandt et al., 2020; Meyer et al., 2019; Rosander et al., 2011). For standardized test scores, negative associations with neuroticism are assumed as anxiety can be triggered by highly stressful (testing) situations (Byrne et al., 2015). Previous studies have shown associations with test scores to be largely negative, with some nonsignificant findings found in language subject domains (e.g., Furnham et al., 2009; Israel et al., 2019; Meyer et al., 2019). Accordingly, domain-specific effects can be hypothesized, for example, anxiety often occurs in math, as reflected by a large amount of research on mathematics anxiety and evidence of negative associations with achievement (Barroso et al., 2021).

Summary of Empirical Findings

Overall, the empirical literature suggests that considering both the academic measure and the domain is important. However, the results of the studies conducted up to now do not make it possible to answer the question of how important these features are for the overall associations of personality traits with achievement. To illustrate the need for a meta-analysis that takes achievement measures and subject domains into account, Table 1 shows how the pattern of results varies depending on the

achievement measure and the domain (see also Supplementary Material, Table S3 for more details on the individual studies).

For example, for openness, previous findings have been positive in language subject domains for both tests and grades, but largely nonsignificant for STEM grades. For conscientiousness, most correlations have been positive for grades across domains but nonsignificant for test scores across domains. For extraversion, more studies have shown positive associations in language subject domains, whereas the results in STEM subject domains have been nonsignificant. For agreeableness, the pattern suggests more positive findings in language subject domains and nonsignificant or negative findings in STEM subject domains. However, some studies have also shown negative associations in language subject domains. For neuroticism, most findings have been nonsignificant for language grades, negative for STEM grades, and negative for test scores in both language and STEM subject domains.

The Present Study

Previous meta-analyses that considered students in elementary and secondary education (e.g., Mammadov, 2022; Poropat, 2009) did not consider potential domain-specific association patterns. This is surprising because a rich set of recent studies has emphasized the importance of a closer consideration of subject domains and achievement measures as potential moderators of the personality–achievement association (e.g., see Table 1). At the same time, a fine-grained distinction between the association patterns of different subject domains (i.e., language vs. STEM) and achievement measures (i.e., grades vs. standardized tests) is required for a better theoretical and practical integration of findings on the personality saturation of achievement measures. To the best of our knowledge, the current study is the first to make this distinction while meta-analytically investigating this research topic.

To do this, we examined three main research questions. In Research Question (RQ) 1, we investigated whether the subject domain moderates the association between personality traits and student achievement. For openness, we expected to find higher effect sizes for language subject domains, on the basis of the current literature (e.g., Hübner et al., 2022; Meyer et al., 2019; Spengler et al., 2013). For conscientiousness, we expected to find positive associations across domains. For extraversion, agreeableness, and neuroticism, we conducted exploratory analyses given the largely inconsistent pattern of results found in empirical studies up to now.

In RQ2, we investigated whether the type of achievement measure moderates the association between personality traits and student achievement. For openness, we expected to find larger associations with standardized tests compared to grades (e.g., Hübner et al., 2022; Meyer et al., 2019; Spengler et al., 2013). For conscientiousness, we expected to find larger associations with grades compared to test scores (e.g., Brandt et al., 2019; Lechner et al., 2017; Nofle & Robins, 2007; Spengler et al., 2013). For extraversion and agreeableness, it seems plausible that the personality saturation would be greater for grades compared to standardized tests. However, we conducted exploratory analyses given the largely inconclusive pattern of

Table 1 Distribution of effect sizes in the studies depending on measure and domain

k	Openness						Conscientiousness						Extraversion						Agreeableness						Neuroticism																																																																																																																																				
	CG		LG		ST		CG		LG		ST		CG		LG		ST		CG		LG		ST		CG		LG		ST																																																																																																																																
	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG	LT	SG																																																																																																																													
	24	19	21	11	15	35	26	26	13	15	26	22	24	13	16	28	20	21	13	14	26	21	23	15	16	18	14	8	10	10	26	21	23	4	5	6	12	4	5	1	14	8	6	4	2	3	2	0	0	1	3	1	0	2	1	2	2	2	1	1	0	2	2	2	1	2	2	2	1	2	2	3	0	2	0	1	0	Close to zero/n.s	4	2	9	1	2	6	3	1	5	5	15	6	11	3	6	12	7	12	1	3	17	10	9	4	4	0	0	3	0	0	0	0	0	0	0	0	0	2	0	2	0	1	0	1	0	1	0	1	2	0	3	1	2	0	1	0	0	0	1	2	0	0	2	3	4	3	7	3	5	0	3	1	5	4	6	4	13	8	12

Note. Note that findings were sometimes heterogeneous even within studies: If a study reported multiple effect sizes, we refer to the overall directions the effect had for the sake of conciseness, that is, this table aims to illustrate how different studies found differential effects across measures and subject domains. Note that, in some cases, studies did not report the significance of the results, *positive*, all effect sizes reported in the study indicated positive associations; *positive/n.s.*, effect sizes indicated both positive and nonsignificant associations; *close to zero/n.s.*, the majority of effect sizes indicated nonsignificant results or small effect sizes but the statistical significance was not reported; *negative/n.s.*, effect sizes indicated both negative and nonsignificant associations; *negative*, all effect sizes reported in the study indicated negative associations; *k*, number of effect sizes; *CG*, composite grades; *LG*, language grade; *SG*, STEM grade; *ST*, language test scores; *ST*, STEM test scores. More information on the studies can be found in Table S3 in the Supplementary Material

wempirical results this far. For neuroticism, we had reasons to assume effects in both directions, as described above, and therefore conducted exploratory analyses.

Finally, in RQ3, we investigated whether there is evidence for a two-way moderation effect of academic subject domain and type of achievement measure (i.e., subject domain \times achievement measure) on the effect sizes (i.e., the correlation between personality trait and achievement). In other words, we investigated whether the effects found between subject domains are similar across measures or whether the effects found between subject domains are enhanced or reduced in grades or standardized test scores, respectively. As no prior studies were identified that investigated this interaction effect, we addressed this research question in an exploratory way.

Methods

Literature Search

We conducted a search of the following databases to identify relevant articles for this meta-analysis: PsycINFO, Web of Science, PubMed, ERIC, and ProQuest. The database searches used the following terms and Boolean operators: (academic OR education OR school) AND (grade OR GPA OR performance OR achievement) AND (personality OR temperament). We limited our search to articles and dissertations that were made available before August 10, 2022. We excluded correlations with measures comparable to the FFM traits, such as the Extraversion and Neuroticism Scales of the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) or the HEXACO (Ashton & Lee, 2008). Thus, in all the studies that we selected, measures had been used that had been validated as assessing the FFM dimensions. This ensured comparability between the findings. The abstracts were screened by the first author. A subsample of 617 studies was screened by a second rater (student assistant); this resulted in an absolute agreement of 95.7% regarding the reports sought for retrieval and 99.5% regarding the reports ultimately included in the meta-analysis.

In addition to the literature search in the databases, all of the studies reported in the meta-analyses of Poropat (2009) and Mammadov (2022) were considered for inclusion in this meta-analysis (see Supplementary Material, Table S2). We included all of the studies included in prior meta-analyses that reported estimates differentiating between subject domains and/or measures (Mammadov: $n = 13$; Poropat: $n = 39$), resulting in 41 studies that were included from previous reviews. This means that our review includes 37 additional studies that were not considered in the previous meta-analyses. The literature search is illustrated in Fig. 1.

Please note that we excluded studies that did not provide individual effect sizes for the respective measures or subject domains, that is, studies that reported only composites averaged across both measures and subject domains (see “[Eligibility Criteria](#)” below). Because of this, we did not include as many studies as Mammadov (2022) did. In the Supplementary Material, Table S2, we provide an overview of the studies that were included in the previous meta-analyses by Poropat (2009) and Mammadov (2022) on how personality traits relate to academic

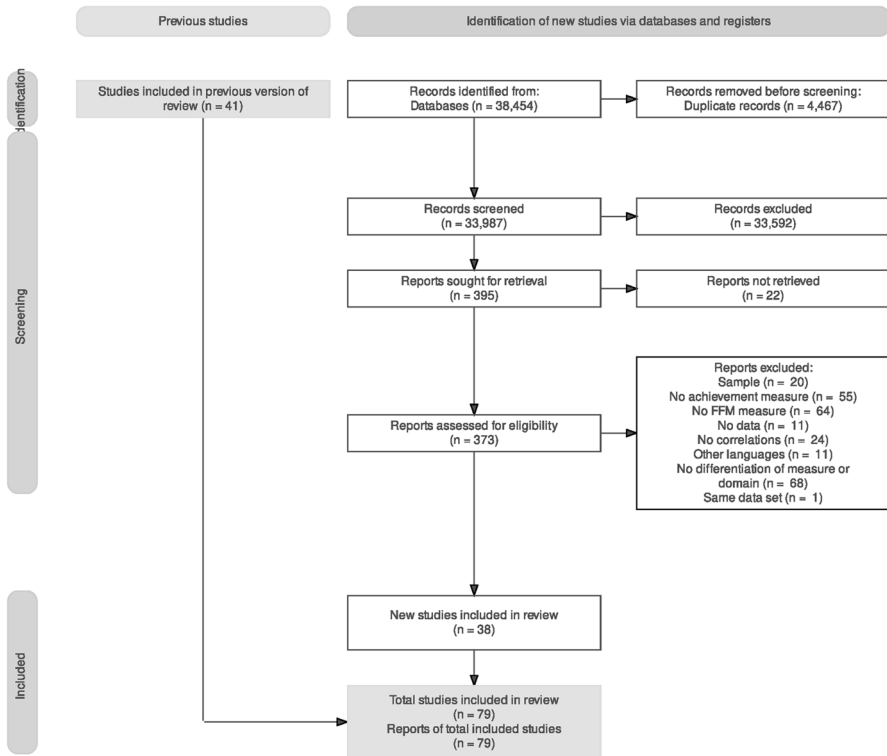


Fig. 1 Flow diagram illustrating the literature search. *Note.* “Studies included in the previous version of review” refers to studies that were included in the previous meta-analyses by Poropat (2009) and Mammadov (2022). More information on which studies were included in the two prior reviews and which were additionally included in our review can be found in Table S2 in the Supplementary Material

achievement. Neither of those prior meta-analyses investigated the role of different types of achievement measures or subject domains. We also list the studies that were included in the current meta-analysis and we provide reasons for why we did not include some of the studies that were included in the prior meta-analyses.

Eligibility Criteria

Inclusion

We included (a) studies reporting at least one bivariate correlation between one of the five-factor personality traits and at least one achievement measure, (b) studies reporting separate effect sizes for achievement measures (e.g., grades, test scores) or subject domains (e.g., STEM, language), (c) studies conducted in populations from elementary and secondary education, (d) studies written in the English language, and (e) studies using self-reported grades as the achievement measure.

Exclusion

We excluded (a) studies reporting self-ratings, academic self-concepts, or other measures of students' academic self-beliefs as the achievement measure (e.g., Cao & Meng, 2020; Gatzka, 2021; Ghapanchi et al., 2011), and (b) studies that did not provide bivariate correlation coefficients. This ensured comparability between studies. If the studies only reported betas (i.e., standardized coefficients) of a regression ($n = 19$), we contacted the corresponding author; eight authors (42%) replied and sent us the correlations. We also excluded (c) studies whose effect sizes were based on the same sample as another study included in our meta-analysis: when two studies analyzed the same samples, we included the study that provided more data relevant to our research question (e.g., we included Kappe & van der Flier, 2010, and Westphal et al., 2020b, but excluded Kappe & van der Flier, 2012, and Westphal et al., 2020a).

Similarly, some studies provided longitudinal analyses with measures of personality and achievement at multiple time points. To avoid using the same data basis more than once, we coded only the first measurement because, usually, the first data collection contains the most data points due to dropouts in later measurements.

Further, we excluded (d) studies that focused on other personality frameworks (e.g., Avram et al., 2019). Although there is some overlap between the dimensions of the FFM and those of the HEXACO (Ashton & Lee, 2008) and Eysenck and Eysenck's Big Three (Eysenck & Eysenck, 1975), we followed Poropat (2009) and included only studies with measures validated to assess the FFM personality traits. This allowed for a high internal validity of our results. Finally, we excluded (e) studies that reported only composite measures of achievement, such as GPA, or other composites that were averaged across both measures and subject domains. We did not exclude studies that averaged across subject domains but differentiated between test scores and grades (i.e., if individual grades in math and English were averaged, we included this effect size but coded it as mixed for the subject domain).

Coding¹

Two independent raters coded each study that fulfilled the eligibility criteria. Overall, three raters were involved in the coding procedure: the first author and two trained student research assistants. Discrepancies were solved by discussion. Interrater reliability, measured with Cohen's kappa, was 0.95. We coded all studies on the following variables for each reported effect size: personality trait, subject domain, achievement measure, effect size (correlation), publication status, sample size, first variable of the correlation (i.e., FFM traits or intelligence), personality reliability, and achievement reliability. We coded reliability to correct measurement errors (see "Statistical

¹ For more detailed information on coding, our codebook can be found at <https://osf.io/8h6wd/>

Corrections”). For descriptive reasons, we also coded mean age, percentage of female students, and region of data collection (see Supplementary Material, Table S1). We also coded educational level but we were unable to perform more specific analyses due to the small number of effect sizes in elementary education (see Supplementary Material, Table S1). We coded the subject domain, classifying science and mathematics as STEM and subject domains focusing on languages as language. We coded subjects that could not be classified according to this distinction, such as history, sports, and social sciences, as “other domains”. These subjects included religious studies, geography, history, practical (a composite of art, music, home and consumer studies, and crafts), social science/studies, sports, humanities, and music. However, each of these subjects appeared in no more than two studies, which is why a detailed, subject-specific investigation was not possible in this category. Details on the effect sizes that we coded as “other domains” can be found in Table S6 in the Supplementary Material.

One important coding issue related to FFM measures was the use of deviating labels for the same scales. This was most important for the neuroticism dimension, which is sometimes called emotional stability, reflecting the opposite pole of the same measure. In this case, the arithmetic signs for all correlations between neuroticism and achievement were reversed (e.g., minus to plus) when we coded the correlations in order to ensure comparability. If necessary, we also recoded data on achievement so that higher values indicated higher levels of achievement. We also coded whether studies reported correlations with measures of intelligence to investigate the association of personality traits and achievement while controlling for cognitive abilities.

Statistical Analyses

We summarized 1491 effect sizes, representing data from 500,218 students and 110 samples from elementary to high school. We used the Fisher r -to- z transformed correlation coefficient as the outcome measure. Effect sizes were weighted by the inverse sampling variance. We calculated the sampling variance using Formula 12.27 from Borenstein et al. (2009). We applied a random-effects model to calculate the mean effect size for each personality trait and estimated the amount of heterogeneity using the restricted maximum-likelihood estimator (Viechtbauer, 2005). To calculate standard errors within the subgroup analyses, we used the pooled subgroup variance instead of the variance of each subgroup, following recommendations from Rubio-Aparicio et al. (2020). We estimated τ (i.e., the standard deviation of the true effects across studies) as a measure of the between-study variation.

Dependency of Effect Sizes

Some studies investigated the relation between multiple measures and subject domains and thus reported more than one effect size. Effect sizes from the same meta-analytic sample are dependent because they share the same method and sample. In this meta-analysis, the number of effect sizes within studies ranged from 1 to 94. Averaging

effect sizes from the same study without correction would underestimate the amount of between-study heterogeneity (Schmidt & Hunter, 2015). Taking into account the dependency in our analyses, we corrected the sampling error using (cluster) robust variance estimation (Hedges et al., 2010). We carried out the analysis using R (version 4.1.0, R Core Team, 2020) and the metafor package (version 3.0.2, Viechtbauer, 2010).

Computing Partial Correlations ρ^1

To investigate partial correlations while correcting for intelligence, we used the sample estimates for each of the effect sizes and computed the partial correlations using the correlations we had coded from the original studies between intelligence and personality traits. The average correlations between personality and intelligence, as well as between achievement and intelligence can be found in Table S5 in the Supplementary Material.

Statistical Corrections

Measurement error can affect the magnitude of the correlations. Following the approach of Poropat (2009), we corrected the correlations on the basis of the reliability of the instruments before we averaged them. For these corrections, we used estimates of Cronbach's alpha provided by each study wherever possible. If studies did not report an estimate of alpha (which was the case in 7% of the effect sizes included), estimates were obtained from the original validating studies for the relevant personality scale (Poropat, 2009). If no validating study was available, we used an estimate of alpha derived from Viswesvaran and Ones's (2000) meta-analysis of FFM reliabilities. For the achievement measures, only 19% of the studies reported estimates of alpha for academic achievement. Following a procedure reported by Poropat (2009), we corrected self-reported grades for their unreliability using the estimated reliability of self-report grades of 0.86 as used by Mammadov (2022). When studies did not report reliabilities for standardized tests, we used the mean reliability of all achievement measures as an estimate (0.90). We chose this estimation method because the studies did not provide the information necessary to obtain more accurate reliability estimates for the achievement tests.

For the achievement measures, reliability, as measured with Cronbach's alpha, ranged from 0.68 (Rosander et al., 2011) to 0.95 (Westphal et al., 2020b), with a mean of 0.80 based on all reliabilities that were reported in the original studies. For the personality measures, reliability ranged from 0.32 (Brandt et al., 2020) to 0.85 (Bergold & Steinmayr, 2018), with a mean of 0.69 across all personality traits. The greatest difference between reported and corrected effect size was found in Hübner et al. (2022), with 0.44 (reported) and 0.64 (corrected). Across all studies, the absolute difference between reported and corrected effect size was small (absolute mean difference = 0.03). As a robustness check, we compared the results for corrected and uncorrected effect sizes and found a similar pattern (see Table 3, columns for ρ and ρ^{raw}). We corrected correlations for scale reliability prior to the combination of the correlations into overall estimates because corrected correlations are more directly comparable than raw correlations (Schmidt & Hunter, 1996).

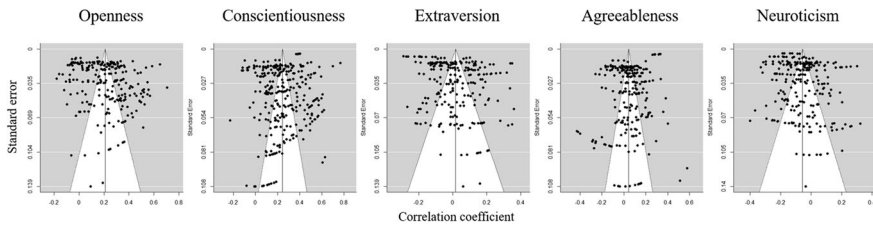


Fig. 2 Funnel plots to inspect the data regarding potential bias

Analyses of Potential Bias

Publication Bias

First, to assess potential publication bias, we aimed to compare the effect sizes of published and unpublished studies but we found only one unpublished study. Second, we conducted a funnel plot analysis for each of the five traits (see Fig. 2) to examine whether studies with significant correlations were more likely to be published than studies with nonsignificant results. Egger's test (Egger et al., 1997) showed potential funnel plot asymmetry for openness ($z=3.46$, $p<0.001$), extraversion ($z=2.31$, $p=0.021$), and neuroticism ($z=2.57$, $p=0.010$). No potential asymmetry was found for conscientiousness ($z=-0.17$, $p=0.866$) or agreeableness ($z=-1.14$, $p=0.252$). Additionally, we conducted trim-and-fill analyses (Duval & Tweedie, 2004) for openness, extraversion, and neuroticism. The trim-and-fill funnel plots (see Figure S1 in the Supplementary Material) showed that the filled studies would have had significantly lower effect sizes than zero, suggesting that the asymmetry found in the funnel plots might be due to reasons other than publication bias (Peters et al., 2008).

Outliers

We conducted an outlier analysis (Viechtbauer & Cheung, 2010) and identified four influential effect sizes within the 1491 effect sizes included in the data set. These cases comprised one correlation between conscientiousness and grades in the language domain (Brandt et al., 2021), two correlations between agreeableness and test scores in the STEM domain (Mammadov et al., 2021), and one correlation between neuroticism and test scores in the language domain (Brandt et al., 2022). We performed robustness checks without the outlier effect sizes, which showed a similar result pattern (see Supplementary Material, Table S6). The influential cases were kept in the data set because they seemed plausible and may reveal patterns in future research. For example, the particularly high effect size found in Brandt et al. (2021) might be an influential case because it refers to elementary school students. As our analyses on the educational level as a moderator showed, samples from elementary schools had higher effect sizes than those from secondary schools or high schools (see moderator analyses regarding educational level as shown in Table S4 in the Supplementary Material). However, for

Table 2 Overview of mean effect sizes per trait

Trait	ρ	95% CI	r	Poropat (2009)	Mammadov (2022)
Openness	0.21 ^B	[0.17, 0.25]	0.17 ^C	0.10	0.13
Conscientiousness	0.24 ^B	[0.21, 0.28]	0.20 ^B	0.19	0.20
Extraversion	0.02 ^C	[-0.02, 0.05]	0.02 ^D	-0.01	0.01
Agreeableness	0.04 ^C	[0.01, 0.08]	0.04 ^D	0.07	0.07
Neuroticism	-0.05 ^D	[-0.09, -0.02]	-0.04 ^E	-0.01	-0.02
Intelligence	0.42 ^A	[0.38, 0.47]	0.36 ^A	0.25	0.42

Note. ρ , mean correlation corrected for measurement error; r , mean correlation not corrected for measurement error. In this table, we also report results uncorrected for measurement error for reasons of comparability. Correlations marked with an A were significantly larger than correlations marked with a B. Correlations marked with a C were significantly smaller than correlations marked with a B and significantly larger than correlations marked with a D

conscientiousness, we found only 50 effect sizes from elementary schools, compared to 303 effect sizes from secondary and 95 from high school samples (see Table S1 in the Supplementary Material). If the effect sizes were balanced across educational levels, the case would most likely not be an influential case.

Results

We provide the characteristics of the individual studies and their effect sizes in the raw data spreadsheet, which is available as an Excel file.² Overall, the pattern of results for the mean associations of achievement with the FFM traits (see Table 2) was comparable to the pattern found in the previous meta-analyses of Poropat (2009) and Mammadov (2022; see Table 2). The true outcomes appeared to be heterogeneous ($p < 0.001$) for every personality trait according to the Q-test ($Q_O[241] = 13,888$, $Q_C[299] = 26,072$, $Q_E[244] = 13,874$, $Q_A[239] = 22,158$, $Q_N[258] = 10,774$) and according to the values of τ (see Table 3).

In the following, we report the specific results, addressing the three research questions for each of the traits, respectively. First, we report results on the moderating effects of the subject domain (composite across achievement measures). Second, we report the moderating effects of the achievement measure (composite across subject domains). Third, we report results on the effect sizes for the subject domain considering measures and vice versa. Here, we only included studies reporting both measure and subject domain for each effect size. All results reported in the following sections of the manuscript are based on coefficients corrected for measurement error. As robustness checks, we also report results for uncorrected effects in Table 3 (r^{raw}). The pattern of findings was largely replicated in these robustness checks, with slightly smaller effect sizes overall. Further, we report our findings after we controlled for cognitive abilities (ρ^I , Table 3).

² Data and syntax for our analyses can be found at <https://osf.io/8h6wd/>

Table 3 Results showing mean correlations depending on subject domain, measure, and interaction of domain*measure for the five traits

	<i>k</i>	<i>N</i>	Studies	ρ	τ	ci.lb	ci.ub	r^{raw}	ρ^I
Openness	247	229,504	60	0.21	0.160	0.17	0.25	0.17	0.12
Domain									
Language	87	217,956	30	0.25	0.143	0.20	0.29	0.20	0.16
STEM	85	207,804	34	0.13	0.143	0.08	0.18	0.10	0.04
Measure									
Grades	174	327,445	54	0.21	0.186	0.16	0.26	0.17	0.13
Test scores	64	165,224	23	0.22	0.128	0.18	0.26	0.17	0.11
Domain*measure									
Language grades	57	137,660	27	0.25	0.155	0.19	0.30	0.19	0.16
Language test scores	28	79,796	16	0.27	0.103	0.23	0.31	0.21	0.16
STEM grades	55	127,475	31	0.12	0.182	0.05	0.18	0.10	0.04
STEM test scores	28	79,829	18	0.16	0.129	0.10	0.21	0.12	0.04
Conscientiousness	305	513,300	74	0.24	0.155	0.21	0.28	0.20	0.21
Domain									
Language	110	502,298	37	0.23	0.168	0.19	0.28	0.19	0.20
STEM	100	215,590	40	0.22	0.168	0.17	0.26	0.18	0.19
Measure									
Grades	220	349,668	66	0.28	0.155	0.25	0.32	0.23	0.25
Test Scores	76	442,400	27	0.13	0.151	0.08	0.19	0.11	0.12
Domain*measure									
Language grades	73	145,895	37	0.27	0.161	0.22	0.32	0.22	0.22
Language test scores	35	355,903	19	0.16	0.165	0.09	0.23	0.13	0.14
STEM grades	68	134,987	41	0.26	0.135	0.22	0.31	0.22	0.23
STEM test scores	30	80,103	21	0.11	0.151	0.03	0.19	0.09	0.11
Extraversion	250	258,461	61	0.02	0.117	-0.02	0.05	0.02	0.02
Domain									
Language	93	306,909	29	0.06	0.134	0.00	0.11	0.05	0.05
STEM	87	237,316	35	-0.02	0.134	-0.06	0.02	-0.02	-0.01
Measure									
Grades	169	339,662	55	0.03	0.112	0.00	0.06	0.03	0.03
Test scores	72	313,530	23	0.00	0.149	-0.08	0.08	0.00	0.01
Domain*measure									
Language grades	59	137,799	26	0.08	0.106	0.05	0.12	0.07	0.08
Language test scores	32	168,610	15	0.02	0.158	-0.08	0.13	0.02	0.03
STEM grades	55	127,245	30	-0.01	0.099	-0.05	0.02	-0.01	-0.01
STEM test scores	30	109,571	19	-0.02	0.147	-0.09	0.05	-0.01	-0.01
Agreeableness	245	503,261	60	0.04	0.116	0.01	0.08	0.04	0.02
Domain									
Language	94	493,920	30	0.06	0.144	0.01	0.10	0.05	0.04
STEM	82	206,661	32	0.01	0.144	-0.02	0.05	0.01	-0.01

Table 3 (continued)

	<i>k</i>	<i>N</i>	Studies	ρ	τ	ci.lb	ci.ub	r^{raw}	ρ^I
Measure									
Grades	167	323,595	53	0.06	0.117	0.03	0.09	0.05	0.04
Test scores	69	441,166	23	0.01	0.163	-0.06	0.07	0.01	-0.02
Domain*Measure									
Language grades	59	137,799	26	0.06	0.130	0.01	0.10	0.05	0.04
Language test scores	33	355,621	16	0.06	0.172	-0.01	0.13	0.04	0.03
STEM grades	54	126,859	29	0.04	0.091	0.01	0.08	0.03	0.02
STEM test scores	26	79,302	16	-0.05	0.129	-0.12	0.02	-0.03	-0.07
Neuroticism	264	488,862	61	-0.05	0.109	-0.09	0.02	-0.04	-0.01
Domain									
Language	104	583,581	31	-0.03	0.111	-0.07	0.01	-0.02	0.01
STEM	88	236,975	34	-0.09	0.111	-0.14	-0.05	-0.07	-0.04
Measure									
Grades	175	340,166	55	-0.03	0.109	-0.06	-0.01	-0.03	0.00
Test scores	80	589,542	23	-0.10	0.129	-0.15	-0.05	-0.08	-0.04
Domain*measure									
Language grades	62	138,222	27	0.00	0.108	-0.04	0.04	0.00	0.02
Language test scores	40	444,859	17	-0.07	0.104	-0.12	-0.01	-0.06	-0.01
STEM grades	56	127,141	30	-0.06	0.095	-0.10	-0.03	-0.05	-0.02
STEM test scores	30	109,334	18	-0.14	0.155	-0.21	-0.06	-0.10	-0.06

Note. *k*, number of effect sizes; *N*, number of students within primary studies; *Studies*, number of primary studies that included multiple effect sizes; ρ , mean correlation corrected for measurement error; τ , index of heterogeneity (i.e., standard deviation of the true effects across studies); *ci.lb*, 95% confidence interval of ρ , lower bound; *ci.ub*, 95% confidence interval of ρ , upper bound; r^{raw} , mean correlation not corrected for measurement error; ρ^I , partial correlation controlling for cognitive ability

Openness Addressing RQ1, for openness, we found a domain-specific effect, with larger correlations in language compared to STEM subject domains ($\rho = 0.25$ vs. 0.13 ; $z = -7.80$; $p < 0.001$). Investigating the role of achievement measure (RQ2), we did not find a significant difference between the effect sizes for grades and for standardized test scores ($\rho = 0.21$ vs. 0.22 ; $z = 0.39$; $p = 0.696$). Addressing RQ3, and looking at grades more specifically, we found a significant difference between the correlations in language and in STEM subject domains ($\rho = 0.25$ vs. 0.12 ; $z = -7.70$; $p < 0.001$); the correlations of openness with language grades were larger than the correlations with STEM grades. Also, we found a significant difference between test scores in language and STEM subject domains ($\rho = 0.27$ vs. 0.16 ; $z = -5.25$; $p < 0.001$), with larger correlations found for languages compared to STEM test scores. The overall interaction between domain and measure was not significant ($t[43] = -0.84$; $p = 0.405$), indicating that the difference between language and STEM achievement was similar for grades and test scores. For an illustration, see Fig. 3.

Conscientiousness Addressing RQ1, for conscientiousness, we did not find a significant moderating effect of domain (language: $\rho = 0.23$; STEM: $\rho = 0.22$; $z = -1.20$;

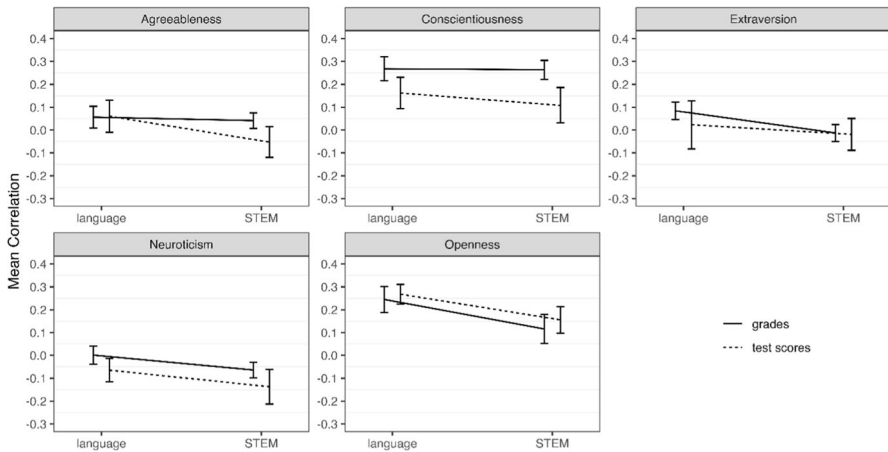


Fig. 3 Means and confidence intervals of how the five traits relate to academic achievement considering the role of measure and domain

$p = 0.230$), but we did find a moderating effect of measure (RQ2), with larger correlations found for grades ($\rho = 0.28$) compared to test scores ($\rho = 0.13$; $z = 5.17$; $p < 0.001$). When looking at the individual comparisons, that is, the interaction between subject domain and measure, the pattern remained stable: we found larger correlations of grades compared to test scores with both STEM ($\rho = 0.26$ vs. 0.11 ; $z = 4.54$; $p < 0.001$) and language subject domains ($\rho = 0.27$ vs. 0.16 ; $z = 3.27$; $p = 0.001$). We did not find any domain-specific differences for test scores in language compared to STEM subject domains ($\rho = 0.16$ vs. 0.11 ; $z = 1.89$; $p = 0.058$), or for grades ($\rho = 0.27$ vs. 0.26 ; $z = -0.29$; $p = 0.772$). There was no statistically significant overall interaction between domain and measure ($t[60] = 1.58$; $p = 0.119$), indicating that the difference between language and STEM achievement was similar for grades and test scores. For an illustration, see Fig. 3.

Extraversion For extraversion, addressing RQ1, we found a domain-specific pattern, with larger correlations in language compared to STEM subject domains ($\rho = 0.06$ vs. -0.02 ; $z = 4.94$; $p < 0.001$). In RQ2, we did not find any significant differences depending on the achievement measure ($\rho = 0.03$ vs. 0.00 ; $z = -0.88$; $p = 0.379$). For the individual comparisons in RQ3, the domain-specific pattern was not consistent across measures: we found larger associations for grades in language compared to STEM subject domains ($\rho = 0.08$ vs. -0.01 ; $z = 6.33$; $p < 0.001$) but not for test scores ($\rho = 0.02$ vs. -0.02 ; $z = 1.45$; $p = 0.149$). Furthermore, we did not find any significant differences between grades and test scores in either STEM ($\rho = -0.01$ vs. -0.02 ; $z = -0.19$; $p = 0.848$) or language ($\rho = 0.08$ vs. 0.02 ; $z = 1.27$; $p = 0.204$) domains. The interaction between measure and subject domain was not significant ($t[44] = -1.71$; $p = 0.095$; see Fig. 3 for an illustration).

Agreeableness For agreeableness, we found evidence of a moderating effect of domain (RQ1), with larger correlations in language ($\rho = 0.06$) compared to STEM

($\rho=0.01$) domains ($z=3.172$; $p=0.002$). The correlations of agreeableness with achievement in STEM subject domains were nonsignificant. We did not find measure-specific effects for grades compared to standardized test scores ($\rho=0.06$ vs. 0.01 ; RQ2; $z=1.65$; $p=0.098$). The correlations of agreeableness with test scores did not differ from zero. Addressing RQ3, we found that the pattern for grades and test scores differed depending on domain. For grades, the moderating effects of the domain were nonsignificant (language: $\rho=0.06$ vs. STEM $\rho=0.04$; $z=0.90$; $p=0.369$). For test scores, we found a domain-specific effect, with larger correlations found with language ($\rho=0.06$) compared to STEM subject domains ($\rho=-0.05$; $z=5.47$; $p<0.001$). We also found a significant interaction ($t[48]=3.04$; $p=0.004$) between domain and measure (see Fig. 3). The interaction effect indicates that the role of the domain differed between grades and test scores: associations in language and STEM subject domains differed more for test scores ($\rho=0.06$ vs. -0.05) than for grades ($\rho=0.06$ vs. 0.04).

Neuroticism For neuroticism, overall, we found negative effects on achievement. We found significant differences when comparing language ($\rho=-.03$) and STEM ($\rho=-.09$) domains (RQ1; $z=4.47$; $p<0.001$) and also when comparing grades and standardized test scores ($\rho=-.03$ vs. -0.10 ; RQ2; $z=2.86$; $p=0.004$). Looking at the more specific analyses in RQ3, we found more negative correlations of neuroticism with test scores compared to grades in both STEM subject domains ($\rho=-.14$ vs. -0.07 ; $z=-2.48$; $p=0.013$) and language subject domains ($\rho=-0.06$ vs. $.00$; $z=-2.81$; $p=0.005$). We also found larger negative correlations of neuroticism with STEM grades ($\rho=-0.06$) compared to languages grades ($\rho=0.00$; $z=-4.71$; $p<0.001$), and the same domain-specific pattern for test scores, with larger negative correlations of neuroticism with STEM test scores ($\rho=-0.14$) compared to language test scores ($\rho=-0.07$; $z=-2.75$; $p=0.006$). We did not find any significant interaction effect between domain and measure ($t[49]=0.26$; $p=0.793$; see Fig. 3).

Discussion

The results of our meta-analysis extend the findings of previous meta-analyses (e.g., Mammadov, 2022; Poropat, 2009), providing a systematic approach to understanding the association of personality traits and achievement in school in more detail. Several empirical studies have noted the importance of academic domain and achievement measure when examining the association between personality traits and academic achievement (e.g., Brandt et al., 2020; Hübner et al., 2022; Israel et al., 2019; Meyer et al., 2019; Spengler et al., 2013), but no meta-analysis so far has specifically addressed these associations.

Our study has three main findings. First, our results highlight the importance of the domain and the measure for understanding how personality traits relate to academic achievement in school, showing that associations differ to a large degree depending on the academic domain and the achievement measure. This highlights the fact that computing the mean across domains and achievement measures might

not be the most appropriate estimate to describe how personality traits and academic achievement are related. Accordingly, our findings allow for a new understanding of previous results (e.g., Mammadov, 2022; Poropat, 2009) and may help to explain why effect sizes for personality–achievement associations substantially vary across different studies. Second, however, considering subject domains and achievement measures alone might not be sufficient, as domain-specific patterns differed depending on whether grades or standardized tests were used as the outcome measure. As such, the combination of both subject domain and achievement was found to be relevant for some of the traits. Third, the pattern of findings differed across the five traits, suggesting that personality becomes observable in student achievement in different ways.

Openness

Regarding the domain specificity of openness, our findings highlight what has been noted in prior research (e.g., Meyer et al., 2019; Spengler et al., 2013), namely, openness is more relevant to academic achievement in language subject domains. Future research is needed to investigate the reasons for this finding. Different explanations for why openness relates to achievement have been provided in the literature and the role of learning approaches (Komarraju et al., 2011) as domain-general motivational aspects has been discussed. In view of our findings, it might be worthwhile to take a more domain-specific approach to investigate the role of openness, for example, by considering domain-specific conceptualizations of motivation (e.g., Marsh et al., 2006; Zhang & Ziegler, 2016). Another explanation could be found in the characteristics of the domain itself, as some of the demands can differ between subject domains; for example, research could investigate the role of creativity as a strong correlate of openness (see Puente-Díaz et al., 2022; Chamorro-Premuzic, 2006), which might be more relevant in language than in STEM subject domains.

Furthermore, for openness, we did not find any evidence of measure-specific associations. One could argue that the association of openness with achievement is less influenced by the features of the outcome measures, as described in the PASH framework (Hübner et al., 2022). Openness seems to stand out in this regard as the principles of the PASH framework might apply to a smaller degree, at least in language subject domains. Moreover, openness is the personality trait most strongly related to intelligence, which—in turn—is more closely related to achievement test scores than it is to grades. Whereas intelligence refers to students' cognitive abilities, openness refers to students' interest in and willingness to explore new tasks (see Ziegler et al., 2012), which, in line with our findings, might be observed especially in testing situations and less so in the classroom context.

Conscientiousness

For conscientiousness, overall, we found larger effect sizes for the association with grades compared to test scores. This matches prior research that has highlighted the

role of conscientiousness in grading (e.g., Westphal et al., 2021). Teachers include their observations of a students' study and classroom behavior (e.g., homework assignments) in their grading, irrespective of subject domain. This highlights the role of teacher expectations and evaluations of student behavior such as self-fulfilling prophecies and perceptual biases that are based on conscientious study behavior (Jussim & Harber, 2005). Such processes can influence grades, which are typically less standardized, but more relevant, curricularly valid, and instructionally sensitive than standardized tests (Hübner et al., 2022).

From this perspective, our findings of larger correlations between grades and conscientiousness, compared to test scores and conscientiousness, in both STEM and language subject domains are in line with the suggestions made in PASH. For students with high scores on conscientiousness, their engaged study behavior might influence teacher perceptions (Gray & Watson, 2002; Roberts et al., 2014) across subject domains. Our results did not show any domain-specific differences in conscientiousness.

Extraversion

Our results show that extraversion was relevant to students' achievement mainly in language grades, indicating that being more extraverted might lead to beneficial social behavior that enhances teacher evaluations, especially in language subject domains. Given their preference for social interaction (Chamorro-Premuzic et al., 2006), extraverted students are more likely to speak more frequently, which is especially important in language subject domains (Dewaele et al., 2008). Accordingly, extraverted students are more likely to be judged positively by teachers, with effects reflecting their oral participation in class as a crucial part of grading, especially in language subject domains. These positive associations with grading might also be related to extraverted students being well-liked among peers (e.g., van der Linden et al., 2010; Wolters et al., 2014) and this, in turn, could possibly facilitate students' engagement in classroom discussions. In contrast, some previous research has suggested that extraversion can be associated with disruptive classroom behavior (i.e., talking to peers during lessons; De Raad & Schouwenburg, 1996), which might be detrimental in some learning situations. However, these assumptions cannot be supported by the results of our meta-analyses. Overall, we did not find any evidence of negative associations of extraversion.

Agreeableness

Our results show that considering the role of measure and subject domain is of particular relevance for agreeableness. We found a rather small overall association of agreeableness with achievement and, overall, no moderating role of measure, but, when looking at the measures more specifically, we found a more differential pattern: We found positive associations with language test scores, whereas associations with STEM test scores were nonsignificant. However, we did not find any

domain-specific differences in grades. Students scoring higher on agreeableness might display more socially adaptive behavior in the classroom, resulting in positive perceptions by teachers (van der Linden et al., 2010), which are reflected in grades regardless of the domain. Agreeableness can become observable for teachers through learning activities and social behavior in the classroom and, thus, can be included in grading (e.g., Ehrler et al., 1999) or can affect learning itself (e.g., by choice of seating position, see Hemyari et al., 2013). However, it should be noted that the effect sizes found for agreeableness and grades in both subject domains were small.

We found a significant difference between associations of agreeableness with language test scores compared to STEM test scores. Previous research on teams has found that agreeableness is one of the strongest personality predictors of team performance. Bradley et al. (2013) found that the associations of agreeableness with performance might be explained by communication. The role of communication in language subject domains could explain why agreeableness is more beneficial in these subject domains as the demands and/or contextual features of these subject domains may amplify the ways in which agreeableness relates to achievement (e.g., working in a team and less emphasis on results; see Wilmot & Ones, 2022).

However, it remains unclear why agreeableness might be associated with lower achievement in STEM testing situations. One explanation could be the competitiveness of the subject domain and testing situation; previous research has suggested that agreeableness is less beneficial to job performance in competitive working environments (Judge & Zapata, 2015; see also Brandt et al., 2020). If students feel that testing situations in STEM are competitive, this might decrease their level of achievement. For grades, there may be a buffering effect; teachers might grade students with agreeable behavior in the classroom more favorably (e.g., van der Linden et al., 2010), resulting in a discrepancy between grades and test scores for STEM subject domains. Even though learning in STEM is not fostered by agreeable traits, such agreeable traits may be beneficial for grading regardless of the subject domain. However, these are just preliminary hypotheses to explain these differences. Moreover, we interpreted our findings on the basis of the significant domain-specific difference we found for agreeableness regarding the associations with test scores. Please note that the coefficients themselves were small and nonsignificant. Future research on these processes is needed to explain the interaction fully.

Neuroticism

For neuroticism, in contrast to the other four personality traits, we found larger negative correlations in STEM compared to language subject domains. This finding underlines the role that anxiety can play in learning situations, which is prevalent mainly in STEM subject domains (Hoferichter et al., 2014). Students' emotionality, which is an important aspect of neuroticism, can be detrimental to academic success, especially in more anxiety-provoking subject domains (e.g., mathematics as an important STEM subject domain, see Barroso et al., 2021). Notably, prior studies (e.g., Mammadov, 2022; Poropat, 2009) found the main associations of neuroticism

to be close to zero. However, our results showed that this could be due to differential findings depending on the measure and subject domain, with larger negative effect sizes found for STEM subject domains and test scores. Notably, the effect sizes of the correlation between neuroticism and STEM test scores come close to effect sizes usually found for conscientiousness, highlighting the importance of this result.

Implications for Future Research

Our findings show that the associations of personality traits with achievement cannot be generalized across measures and subject domains. Future researchers investigating the role of personality traits in association with learning outcomes should specifically consider which aspects of achievement are relevant to their investigation and should choose the appropriate achievement measure accordingly. Our findings highlight that there is no “best” or “more objective” achievement measure for investigating the personality saturation of achievement measures but, instead, that achievement measures should be chosen based on the research question of interest.

Finally, Soto et al. (2022) recently emphasized that students can differ regarding their social, emotional, and behavioral (SEB) skills. Soto et al. (2022) argued that these noncognitive skills resemble the Big Five personality traits in terms of their SEB content. While the Big Five are on a different level of analysis (e.g., traits vs. skills; see also Soto et al., 2022), it will be an interesting question for future research to investigate how these SEB skills relate to different achievement measures and subject domains. It can be hypothesized, based on our findings and the similarities between the taxonomies, that these aspects of achievement could also moderate SEB skill–achievement associations. Investigating these questions will be interesting especially given that these skills might be easier to target in interventions than personality traits and are thus of great interest to policymakers and practitioners (Duckworth & Yeager, 2015; Kautz et al., 2014; Napolitano et al., 2021; OECD, 2015).

Limitations

As described above, our meta-analysis is limited to the FFM. We did not include more specific views on personality (see Möttus et al., 2020) or different conceptual frameworks (e.g., HEXACO; Ashton & Lee, 2020). This limits the scope of our investigation; however, this limitation was necessary in order not only to link our findings to previous research (e.g., Mammadov, 2022; Poropat, 2009) but also to systematically analyze the moderating role of subject domains and measures. Accordingly, including different models of personality was beyond the scope of our investigation. However, future meta-analytic research could focus on other models of personality and how they relate to academic achievement, potentially comparing their results with what we and others found regarding the FFM.

Our analysis was limited to comparing language and STEM subject domains as well as grades and standardized achievement tests. This approach was inspired by

previous studies on the association between achievement measures and personality that have often contrasted STEM and language domains as well as grades and standardized achievement tests. Although these are commonly used distinctions and are relevant, well-researched academic outcomes, we acknowledge that this does not cover the entirety of academic outcomes and subject domains. Some previous studies considered different outcomes (e.g., teacher appraisals) or subject domains (e.g., sports or social subject domains); however, at the moment, there are not enough studies of this kind to meaningfully investigate them in a meta-analytic approach. Further research is needed to more closely investigate these aspects for different outcomes and domains. Because of this, we cannot exclude the possibility of other potential moderators that might be confounded with the subject domain. For example, we did not consider the role of the personality instrument used even though prior research has indicated that substantial differences can depend on which personality facet is assessed with the questionnaire (see Möttus, 2016).

Similarly, there might be differential effects depending on which personality instrument was used in the respective studies. We coded this information, which can be reviewed in our Supplementary Material. For example, Andersen et al. (2020) used two items to assess agreeableness; this resulted in a correlation with a conscientiousness that was larger than usual. Further, the reliability of the two-item instrument was low, which resulted in larger correlations when correcting for measurement error. However, when considering the raw coefficients, the pattern of results remained stable. Thus, our findings can be viewed as robust despite these measurement problems. Still, these issues need to be kept in mind when interpreting the effects.

It would have been interesting to investigate the role of other moderators in the association between personality traits and achievement across measures and domains, such as educational level, gender, and age, but, unfortunately, we did not identify a sufficient number of studies within each combination of domain and measure to meaningfully investigate these questions (see Supplementary Material, Table S1). Accordingly, we performed moderator analyses for educational level, gender, and age for the entire database as additional analyses; the results of these analyses can be found in Table S4 in the Supplementary Material. Future studies are needed to investigate whether the effect of subject domain and achievement indicator on the association of personality traits and achievement changes with educational level, gender, or age (e.g., De Raad & Schouwenburg, 1996).

The limited number of effect sizes found within the different combinations of domains and measures might have resulted in low power for some of our analyses, which is perhaps why we were not able to reject the null hypothesis of no statistical interaction effect in some cases. However, the main goal of our meta-analysis was to summarize the available empirical evidence on the moderating role of domain and measure when investigating associations between the Big Five and achievement. Thus, we were more concerned with the size (i.e., point estimates) and precision of our estimates (of the domain- and measure-specific associations). The precision, which is a function of the number of studies, is reflected in the width of the confidence intervals reported (see Table 3). A low precision results in a decreased power

to reject a false null hypothesis, which should be kept in mind when interpreting our results.

On a related note, it would be interesting to consider the role of subject domain and measure in tertiary education. However, tertiary education differs from school education in important ways. First, samples in tertiary education are range-restricted regarding cognitive ability and academic aspirations, given the selectivity of applications and admissions. Similarly, considering domain-specific differences in achievement is challenging because students choose courses depending on their interests and prior achievement to a greater extent compared to in-school education, further increasing the selectivity of the samples within subject domains. Furthermore, a thorough distinction between standardized test scores and grades seems very complicated at the tertiary level and the meaning of these measures differs substantially, compared to the meaning of the measures in elementary and secondary schools. For these reasons, we had doubts about the relevance of our research questions in tertiary education and therefore focused on school students only.

Furthermore, we aimed to interpret findings with regard to the aspects of the learning and assessment situations in the classroom, trying to disentangle how personality relates to different aspects of achievement. However, we did not examine which aspect is most important in explaining the differential findings. This means that we aimed to only describe the meta-analytic pattern of results; with our analyses, we were not able to test the assumptions about why, for example, personality traits relate more closely to language subject domains or why we found differential effects for achievement measure depending on the subject domain. Future research needs to conduct studies to disentangle these processes more systematically (e.g., using the PASH framework, Hübner et al., 2022).

We only included studies reported in English to enable a sufficient understanding of the studies when coding. However, this limits the generalizability of our results as we only can speak of the literature in English. Furthermore, we coded the region of data collection to assess where the investigated samples came from. We found that, of the effect sizes included, 87% were from European countries (see Supplementary Material, Table S1). Accordingly, our results can only be generalized across the European context. Clearly, more research is needed on populations in other continents.

The abstract screening was performed solely by the first author, which could have led to biased screening and thus might constitute a potential limitation of the current study. In this context, it is important to remember that we considered prior meta-analyses by Propat (2009) and Mammadov (2022) to make sure we did not miss any eligible studies published until 2022. In addition, to further evaluate the quality of our initial screening process, a second rater screened a subsample of abstracts from 617 studies. The interrater reliability for the selected studies was very high and amounted to 99.5%. On the basis of this, we believe that the threat of biased study screening was small.

Finally, funnel plot analyses showed asymmetry for three personality traits, similar to findings in previous meta-analyses (Mammadov, 2022). Asymmetry could be a consequence of publication bias or heterogeneity. We conducted trim-and-fill analyses (Duval & Tweedie, 2004) for openness, extraversion, and neuroticism to

get a more detailed picture of the potential bias. The trim-and-fill funnel plots (see Figure S1 in the Supplementary Material) showed that the filled studies would have had significantly lower effect sizes than zero, suggesting that the asymmetry found in the funnel plots might be due to reasons other than publication bias (Peters et al., 2008). However, we could not test the asymmetry within subject domains and measures because of the limited number of studies. Further, the sample sizes were not equally distributed between our levels of moderators. Accordingly, the funnel plots' potential asymmetry could be an indicator of publication bias or a result of heterogeneity in sample sizes between the levels of the moderators. We cannot disentangle whether these differences led to asymmetry in the funnel plot because of publication bias or because of true differences in the associations. However, we argue that publication bias should be less of a problem when considering the associations of the five traits with achievement as, even if one of the effects differed from expectations, it is unlikely that the study would not be published because of this. In this regard, studies investigating the five traits differ from other studies, such as intervention studies, thereby potentially making publication bias less of an issue.

Conclusion

This study closely investigated the role of subject domain and achievement measure in the personality saturation of student achievement. It thus extends prior meta-analytical work by considering two moderators, which were theorized and found to be relevant in prior educational psychological work. We found that this more detailed perspective provided important new insights regarding the correlation of specific aspects of achievement and personality, which, as expected, differed substantially from prior findings that were based on composite measures. Accordingly, for all traits, in order to understand how personality is related to academic achievement, considering subject domain and measure can be considered highly beneficial. However, our findings also highlight that considering subject domains and achievement measures alone might not be sufficient, as we found that domain-specific patterns differed depending on whether grades or standardized tests were used as the outcome measure. As such, the combination of both subject domain and achievement was found to be relevant for some of the traits. On the basis of these findings, we conclude that there is no one measure that best reflects personality but that the measure should be chosen carefully in connection with each research (or practical) question and context. Overall, considering different achievement measures and subject domains was found to be an important step toward gaining a better understanding of how personality traits differentially shape students' academic achievement.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-023-09736-2>.

Acknowledgements We would like to thank James E. Pustejovsky for helpful comments on an earlier version of this manuscript, our student assistants Marlene Steinbach, Mira Tanz, and Antonia Möck for supporting us in this work, and Gráinne Newcombe for editorial support during the preparation of this article.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability Data are available at: <https://osf.io/8h6wd/>.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

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

- *Andersen, S. C., Gensowski, M., Ludeke, S. G., & John, O. P. (2020). A stable relationship between personality and academic performance from childhood through adolescence: An original study and replication in hundred-thousand-person samples. *Journal of Personality*, 88(5), 925–939. <https://doi.org/10.1111/jopy.12538>
- Ashton, M. C., & Lee, K. (2008). The HEXACO model of personality structure and the importance of the H factor. *Social and Personality Psychology Compass*, 2(5), 1952–1962. <https://doi.org/10.1111/j.1751-9004.2008.00134.x>
- Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, 34(4), 492–510. <https://doi.org/10.1002/per.2242>
- Avram, E., Burtaverde, V., & Zanafirescu, A. Ş. (2019). The incremental validity of career adaptability in predicting academic performance. *Social Psychology of Education*, 22(4), 867–882. <https://doi.org/10.1007/s11218-019-09505-6>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2021). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*, 147(2), 134–168. <https://doi.org/10.1037/bul0000307>
- *Bergold, S., & Steinmayr, R. (2018). Personality and Intelligence Interact in the Prediction of Academic Achievement. *Journal of Intelligence*, 6(2), 1–18. <https://doi.org/10.3390/jintelligence6020027>
- Borenstein, M., Cooper, H [H.], Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*, 2, 221–235.
- *Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *PNAS*, 113(47), 13354–13359. <https://doi.org/10.1073/pnas.1601135113>
- Bradley, B. H., Baur, J. E., Banford, C. G., & Postlethwaite, B. E. (2013). Team players and collective performance: How agreeableness affects team performance over time. *Small Group Research*, 44(6), 680–711. <https://doi.org/10.1177/1046496413507609>
- *Brandt, N. D., Lechner, C. M., Tetzner, J., & Rammstedt, B. (2020). Personality, cognitive ability, and academic performance. Differential associations across school subjects and school tracks. *Journal of Personality*, 88(2), 249–265. <https://doi.org/10.1111/jopy.12482>

- *Brandt, N. D., Becker, M., Tetzner, J., Brunner, M., & Kuhl, P. (2021). What teachers and parents can add to personality ratings of children: Unique associations with academic performance in elementary school. *European Journal of Personality*, 35(6), 814–832. <https://doi.org/10.1177/0890207020988436>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- Byrne, K. A., Silasi-Mansat, C. D., & Worthy, D. A. (2015). Who chokes under pressure? The Big Five personality traits and decision-making under pressure. *Personality and Individual Differences*, 74, 22–28. <https://doi.org/10.1016/j.paid.2014.10.009>
- Cao, C., & Meng, Q. (2020). Exploring personality traits as predictors of English achievement and global competence among Chinese university students: English learning motivation as the moderator. *Learning and Individual Differences*, 77. <https://doi.org/10.1016/j.lindif.2019.101814>
- Carver, C. S., & Connor-Smith, J. (2010). Personality and coping. *Annual Review of Psychology*, 61, 679–704. <https://doi.org/10.1146/annurev.psych.093008.100352>
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, 56, 453–484. <https://doi.org/10.1146/annurev.psych.55.090902.141913>
- Chamorro-Premuzic, T. (2006). Creativity versus conscientiousness: Which is a better predictor of student performance? *Applied Cognitive Psychology*, 20(4), 521–531. <https://doi.org/10.1002/acp.1196>
- Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. L. (2006). Incremental validity of the typical intellectual engagement scale as predictor of different academic performance measures. *Journal of Personality Assessment*, 87(3), 261–268. https://doi.org/10.1207/s15327752jpa8703_07
- Chamorro-Premuzic, T., Furnham, A., & Lewis, M. (2007). Personality and approaches to learning predict preference for different teaching methods. *Learning and Individual Differences*, 17, 241–250. <https://doi.org/10.1016/j.lindif.2006.12.001>
- De Raad, B., & Schouwenburg, H. C. (1996). Personality in learning and education: A review. *European Journal of Personality*, 10, 303–336. [https://doi.org/10.1002/\(SICI\)1099-0984\(199612\)10:5<303::AID-PER262>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-0984(199612)10:5<303::AID-PER262>3.0.CO;2-2)
- *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>
- Dewaele, J. M., Petrides, K. V., & Furnham, A. (2008). Effects of trait emotional intelligence and socio-biographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, 58(4), 911–960. <https://doi.org/10.1111/j.1467-9922.2008.00482.x>
- Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12), 939–944. <https://doi.org/10.1111/j.1467-9280.2005.01641.x>
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237–251. <https://doi.org/10.3102/0013189X15584327>
- *Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences*, 37(1), 8–15. <https://doi.org/10.1027/1614-0001/a000182>
- Duval, S., & Tweedie, R. (2004). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Ebstrup, J. F., Eplov, L. F., Pisinger, C., & Jørgensen, T. (2011). Association between the Five Factor personality traits and perceived stress: Is the effect mediated by general self-efficacy? *Anxiety, Stress & Coping*, 24(4), 407–419. <https://doi.org/10.1080/10615806.2010.540012>
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Ehrler, D. J., Evans, G. J., & McGhee, R. L. (1999). Extending Big-Five theory into childhood: A preliminary investigation into the relationship between Big-Five personality traits and behavior

- problems in children. *Psychology in the Schools*, 36(6), 451–458. [https://doi.org/10.1002/\(SICI\)1520-6807\(199911\)36:6%3C451::AID-PITS1%3E3.0.CO;2-E](https://doi.org/10.1002/(SICI)1520-6807(199911)36:6%3C451::AID-PITS1%3E3.0.CO;2-E)
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. Hodder and Stoughton.
- *Furnham, A., Mونسen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, Big Five personality traits, approaches to learning and cognitive ability predictors of academic performance. *The British Journal of Educational Psychology*, 79(4), 769–782. <https://doi.org/10.1348/978185409X412147>
- Gatzka, T., & Hell, B. (2017). Openness and postsecondary academic performance: A meta-analysis of facet-, aspect-, and dimension-level correlations. *Journal of Educational Psychology*, 110, 355–377. <https://psycnet.apa.org/doi/10.1037/edu0000194>
- Gatzka, T. (2021). Aspects of openness as predictors of academic achievement. *Personality and Individual Differences*, 170. <https://doi.org/10.1016/j.paid.2020.110422>
- Ghapanchi, Z., Khajavy, G. H., & Asadpour, S. F. (2011). L2 motivation and personality as predictors of the second language proficiency: Role of the Big Five traits and L2 motivational self system. *Canadian Social Science*, 7(6), 148–155. <https://doi.org/10.3968/j.css.1923669720110706.030>
- Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, 70(2), 177–206. <https://doi.org/10.1111/1467-6494.05002>
- Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: Dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction*, 49, 81–91. <https://doi.org/10.1016/j.learninstruc.2016.12.007>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Hemyari, C., Zomorodian, K., Ahrari, I., Tavana, S., Parva, M., Pakshir, K., ... & Sahraian, A. (2013). The mutual impact of personality traits on seating preference and educational achievement. *European Journal of Psychology of Education*, 28(3), 863–877. <https://doi.org/10.1007/s10212-012-0144-3>
- *Hendriks, A. A. J., Kuyper, H., Lubbers, M. J., & van der Werf, M. P. C. (2011). Personality as a moderator of context effects on academic achievement. *Journal of School Psychology*, 49(2), 217–248. <https://doi.org/10.1016/j.jsp.2010.12.001>
- Hofrichter, F., Raufelder, D., & Eid, M. (2014). The mediating role of socio-motivational relationships in the interplay of perceived stress, neuroticism, and test anxiety among adolescent students. *Psychology in the Schools*, 51, 736–752. <https://doi.org/10.1002/pits>
- Hübner, N., Wille, E., Cambria, J., Oschatz, K., Nagengast, B., & Trautwein, U. (2017). Maximizing gender equality by minimizing course choice options? Effects of obligatory coursework in math on gender differences in STEM. *Journal of Educational Psychology*, 109, 993–1009. <https://doi.org/10.1037/edu0000183>
- Hübner, N., Wagner, W., Hochweber, J., Neumann, M., & Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *Journal of Educational Psychology*, 112(1), 204–220. <https://doi.org/10.1037/edu0000351>
- *Hübner, N., Spengler, M., Nagengast, B., Borghans, L., Schils, T., & Trautwein, U. (2022). When academic achievement (also) reflects personality: Using the personality-achievement saturation hypothesis (PASH) to explain differential associations between achievement measures and personality traits. *Journal of Educational Psychology*, 114(2), 326–345. <https://doi.org/10.1037/edu0000571>
- *Israel, A., Lüdtke, O., & Wagner, J. (2019). The longitudinal association between personality and achievement in adolescence. Differential effects across all Big Five traits and four achievement indicators. *Learning and Individual Differences*, 72, 80–91. <https://doi.org/10.1016/j.lindif.2019.03.001>
- *Israel, A., Brandt, N. D., Grund, S., Köller, O., Lüdtke, O., & Wagner, J. (2022). Personality and psychosocial functioning in early adolescence: Age-differential associations from the self- and parent perspective. *European Journal of Personality*, 36(2), 218–244. <https://doi.org/10.1177/08902070211005636>
- Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences. Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, 41(2), 13–24. <https://doi.org/10.1016/j.cedpsych.2014.11.002>

- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (pp. 114–158). The Guilford Press.
- Judge, T. A., & Zapata, C. P. (2015). The person–situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal*, 58(4), 1149–1179. <https://doi.org/10.5465/amj.2010.0837>
- Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3
- Kappe, R., & van der Flier, H. (2010). Using multiple and specific criteria to assess the predictive validity of the Big Five personality factors on academic performance. *Journal of Research in Personality*, 44(1), 142–145. <https://doi.org/10.1016/j.jrp.2009.11.002>
- Kappe, R., & Van Der Flier, H. (2012). Predicting academic success in higher education: What's more important than being smart? *European Journal of Psychology of Education*, 27(4), 605–619. <https://doi.org/10.1007/s10212-011-0099-9>
- Kautz, T., Heckman, J. J., Diris, R., Weel, B. T., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (No. w20749). National Bureau of Economic Research. <https://doi.org/10.3386/w20749>
- Komaraju, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51(4), 472–477. <https://doi.org/10.1016/j.paid.2011.04.019>
- Lechner, C., Danner, D., & Rammstedt, B. (2017). How is personality related to intelligence and achievement? A replication and extension of Borghans et al. and Salkever. *Personality and Individual Differences*, 111, 86–91. <https://doi.org/10.1016/j.paid.2017.01.040>
- Lipnevich, A. A., Preckel, F., & Krumm, S. (2016). Mathematics attitudes and their unique contribution to achievement: Going over and above cognitive ability and personality. *Learning and Individual Differences*, 47, 70–79. <https://doi.org/10.1016/j.lindif.2015.12.027>
- Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down University Avenue: Life paths, life events, and personality trait change at the transition to university life. *Journal of Personality and Social Psychology*, 101(3), 620.
- MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, 19(4), 451–458. <https://doi.org/10.1016/j.lindif.2009.03.007>
- Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, 90(2), 222–255. <https://doi.org/10.1111/jopy.12663>
- *Mammadov, S., Cross, T. L., & Olszewski-Kubilius, P. (2021). A look beyond aptitude: The relationship between personality traits, autonomous motivation, and academic achievement in gifted students. *Roepers Review*, 43(3), 161–172.
- *Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, 74(2), 403–456. <https://doi.org/10.1111/j.1467-6494.2005.00380.x>
- McCrae, R. R. (1990). How well is openness represented in natural languages? *European Journal of Personality*, 4, 119–129.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90. <https://doi.org/10.1037//0022-3514.52.1.81>
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Melin, R., Fugl-Meyer, K. S., & Fugl-Meyer, A. R. (2003). Life satisfaction in 18-to 64-year-old Swedes: In relation to education, employment situation, health and physical activity. *Journal of Rehabilitation Medicine*, 35(2), 84–90.
- *Meyer, J., Fleckenstein, J., Retelsdorf, J., & Köller, O. (2019). The relationship of personality traits and different measures of domain-specific achievement in upper secondary education. *Learning and Individual Differences*, 69, 45–59. <https://doi.org/10.1016/j.lindif.2018.11.005>
- Mottaz, C. (1984). Education and work satisfaction. *Human Relations*, 37(11), 985–1004.
- Möttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, 30(4), 292–303.

- Möttus, R., Wood, D., Condon, D. M., et al. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the big few traits. *European Journal of Personality*, 34(6), 1175–1201. <https://doi.org/10.1002/per.2311>
- Murberg, T. A., & Bru, E. (2007). The role of neuroticism and perceived school-related stress in somatic symptoms among students in Norwegian junior high schools. *Journal of Adolescence*, 30, 203–212. <https://doi.org/10.1016/j.adolescence.2006.02.001>
- Nagy, G., Garrett, J., Trautwein, U., Cortina, K. S., Baumert, J., & Eccles, J. (2008). Gendered high school course selection as a precursor of gendered occupational careers: The mediating role of self-concept and intrinsic value. In H. M. G. Watt & J. S. Eccles (Eds.), *Gendered occupational outcomes: Longitudinal assessments of individual, social, and cultural influences* (pp. 115–143). American Psychological Association.
- Napolitano, C. M., Sewell, M. N., Yoon, H. J., Soto, C. J., & Roberts, B. W. (2021). Social, emotional, and behavioral skills: An integrative model of the skills associated with success during adolescence and across the life span. *Frontiers in Education*, 6, 679561. <https://doi.org/10.3389/educ.2021.679561>
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93(1), 116. <https://doi.org/10.1037/0022-3514.93.1.116>
- Organisation for Economic Co-Operation and Development. (2015). *Skills for social progress: The power of social and emotional skills*. Latest retrieval: Jan, 21, 2023 from <http://www.oecd.org/education/skills-for-social-progress-9789264226159-en.htm>
- Pawlowska, D. K., Westerman, J. W., Bergman, S. M., & Huelsman, T. J. (2014). Student personality, classroom environment, and student outcomes: A person–environment fit analysis. *Learning and Individual Differences*, 36(2), 180–193. <https://doi.org/10.1016/j.lindif.2014.10.005>
- Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology*, 106(1), 315–329.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322–338. <https://doi.org/10.1037/a0014996>
- Puente-Díaz, R., Cavazos-Arroyo, J., Puerta-Sierra, L., & Vargas-Barrera, F. (2022). The contribution Openness to Experience and its two aspects to the explanation of idea generation, evaluation and selection: A metacognitive perspective. *Personality and Individual Differences*, 185, 111240. <https://doi.org/10.1016/j.paid.2021.111240>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, 43(2), 137–145. <https://doi.org/10.1016/j.jrp.2008.12.015>
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345. <https://doi.org/10.1111/j.1745-6916.2007.00047.x>
- Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology*, 50, 1315–1330. <https://doi.org/10.1037/a0031109>
- *Rosander, P., Bäckström, M., & Stenberg, G. (2011). Personality traits and general intelligence as predictors of academic performance. A structural equation modelling approach. *Learning and Individual Differences*, 21(5), 590–596. <https://doi.org/10.1016/j.lindif.2011.04.004>
- Rubio-Aparicio, M., López-López, J. A., Viechtbauer, W., Marín-Martínez, F., Botella, J., & Sánchez-Meca, J. (2020). Testing categorical moderators in mixed-effects meta-analysis in the presence of heteroscedasticity. *The Journal of Experimental Education*, 88(2), 288–310.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223. <https://psycnet.apa.org/doi/10.1037/1082-989X.1.2.199>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

- Schoon, I., & Eccles, J. S. (2014). *Gender differences in aspirations and attainment*. Cambridge University Press.
- Schwaba, T., Luhmann, M., Denissen, J. J. A., Chung, J. M., & Bleidorn, W. (2017). Openness to experience and culture-openness transactions across the lifespan. *Journal of Personality and Social Psychology*, 115(1), 118–136. <https://doi.org/10.1037/pspp0000150>
- Smith, M. M., Sherry, S. B., Vidovic, V., Saklofske, D. H., Stoeber, J., & Benoit, A. (2019). Perfectionism and the five-factor model of personality: A meta-analytic review. *Personality and Social Psychology Review*, 23(4), 367–390. <https://doi.org/10.1177/1088868318814973>
- Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology*. Advance online publication. <https://doi.org/10.1037/pspp0000401>
- Spengler, M., Damian, R. I., & Roberts, B. W. (2018). How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability. *Journal of Personality and Social Psychology*, 114(4), 620–636. <https://doi.org/10.1037/pspp0000185>
- *Spengler, M., Lüdtke, O., Martin, R., & Brunner, M. (2013). Personality is related to educational outcomes in late adolescence. Evidence from two large-scale achievement studies. *Journal of Research in Personality*, 47(5), 613–625. <https://doi.org/10.1016/j.jrp.2013.05.008>
- Spinath, B., Freudenthaler, H. H., & Neubauer, A. C. (2010). Domain-specific school achievement in boys and girls as predicted by intelligence, personality and motivation. *Personality and Individual Differences*, 48(4), 481–486. <https://doi.org/10.1016/j.paid.2009.11.028>
- *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>
- Szabó, M. (2011). The emotional experience associated with worrying: Anxiety, depression, or stress? *Anxiety, Stress & Coping*, 24(1), 91–105. <https://doi.org/10.1080/10615801003653430>
- *Tetzner, J., Becker, M., & Brandt, N. D. (2020). Personality-achievement associations in adolescence-examining associations across grade levels and learning environments. *Journal of Personality*, 88(2), 356–372. <https://doi.org/10.1111/jopy.12495>
- *Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology*, 97(6), 1115–1128. <https://doi.org/10.1037/a0017048>
- Uliaszek, A. A., Zinbarg, R. E., Mineka, S., Craske, M. G., & Sutton, J. M. (2010). The role of neuroticism and extraversion in the stress-anxiety and stress-depression relationships. *Anxiety, Stress & Coping*, 23, 363–381. <https://doi.org/10.1080/10615800903377264>
- van der Linden, D., Scholte, R. H., Cillessen, A. H., te Nijenhuis, J., & Segers, E. (2010). Classroom ratings of likeability and popularity are related to the Big Five and the general factor of personality. *Journal of Research in Personality*, 44(5), 669–672. <https://doi.org/10.1016/j.jrp.2010.08.007>
- Vedel, A. (2016). Big Five personality group differences across academic majors: A systematic review. *Personality and Individual Differences*, 92, 1–10. <https://doi.org/10.1016/j.paid.2015.12.011>
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3), 261–293. <https://doi.org/10.3102/10769986030003261>
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>.
- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in “Big Five Factors” personality assessment: Reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60(2), 224–235. <https://doi.org/10.1177/00131640021970475>
- Wan, S., Lauermaun, F., Bailey, D. H., & Eccles, J. S. (2021). When do students begin to think that one has to be either a “math person” or a “language person”? A meta-analytic review. *Psychological Bulletin*, 147(9), 867. <https://doi.org/10.1037/bul0000340>
- Westphal, A., Vock, M., & Kretschmann, J. (2021). Unraveling the relationship between teacher-assigned grades, student personality, and standardized test scores. *Frontiers in Psychology*, 12, 530.

- Westphal, A., Lazarides, R., & Vock, M. (2020a). Are some students graded more adequately than others? Student characteristics as moderators of the relationships between teacher-assigned grades and test scores. *British Journal of Educational Psychology*. <https://doi.org/10.1111/bjep.12397>
- *Westphal, A., Vock, M., & Lazarides, R. (2020b). Are more conscientious seventh- and ninth-graders less likely to be retained? Effects of Big Five personality traits on grade retention in two different age cohorts. *Journal of Applied Developmental Psychology*, 66. <https://doi.org/10.1016/j.appdev.2019.101088>
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1–37. <https://doi.org/10.1111/j.1745-3984.2002.tb01133.x>
- Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *Proceedings of the National Academy of Sciences*, 116(46), 23004–23010.
- Wilmot, M. P., & Ones, D. S. (2022). Agreeableness and its consequences: A quantitative review of meta-analytic findings. *Personality and Social Psychology Review*, Advanced Online Publication. <https://doi.org/10.1177/10888683211073007>
- Wolters, N., Knors, H., Cillessen, A. H., & Verhoeven, L. (2014). Behavioral, personality, and communicative predictors of acceptance and popularity in early adolescence. *The Journal of Early Adolescence*, 34(5), 585–605.
- *Zhang, J., & Ziegler, M. (2016). How do the big five influence scholastic performance? A big five-narrow traits model or a double mediation model. *Learning and Individual Differences*, 50(4), 93–102. <https://doi.org/10.1016/j.lindif.2016.07.011>
- Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality*, 46(2), 173–183. <https://doi.org/10.1016/j.jrp.2012.01.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jennifer Meyer¹  · Thorben Jansen¹  · Nicolas Hübner²  · Oliver Lüdtke¹

Thorben Jansen
tjansen@leibniz-ipn.de

Nicolas Hübner
nicolas.huebner@uni-tuebingen.de

Oliver Lüdtke
oluedtke@leibniz-ipn.de

¹ Leibniz Institute for Science and Mathematics Education (IPN), Olshausenstr 62, 24118 Kiel, Germany

² Institute of Education, University of Tübingen, Münzgasse 30, 72070, Tübingen, Germany