Character-Based Admissions Criteria in the United States and in Europe
Niessen, Susan; Meijer, Rob R.

Published in:
Higher Education Admission Practices: An International Perspective

DOI:
[10.1017/9781108559607.005](https://doi.org/10.1017/9781108559607.005)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Character-Based Admissions Criteria in the United States and in Europe: Rationale, Evidence, and Some Critical Remarks

A. Susan M. Niessen and Rob R. Meijer

Globally, the importance of effective and fair college admissions procedures, on both an individual and a societal level, should not be underestimated. College admissions decisions are often based on high school grades (as, for example, in Europe), scores on standardized tests (as, for example, in China and India), or a combination of both (as, for example, in the United States). Although both grades and cognitively oriented standardized test scores are good predictors of academic performance in higher education (Berry, 2015; Westrick et al., 2015), colleges and other stakeholders seem increasingly interested in including character-based criteria in admissions procedures (Hoover, 2013). In this chapter, we discuss the rationale, practice, evidence, and effects of character-based college admissions from US and European perspectives.

Character-based admissions criteria include indications of personality, motivation, study skills and habits, and other behavioral tendencies. Other commonly used terms are noncognitive, intra- and interpersonal, and non-academic traits or skills. Because many of the different traits and skills that are defined within these catch-all terms are not entirely unrelated to cognitive or academic skills (Borghans et al., 2011; von Stumm & Ackerman, 2013), we chose to use the term "character-based" in this chapter.

The higher education admissions literature is dominated by US studies and perspectives. The academic discipline of psychology, especially the field of individual differences, has had a large influence on educational selection and admissions procedures in the United States. The influence of cognitive testing is predominant in large-scale admissions tests such as the SAT® and the ACT® tests (Lemann, 1999). In addition, the US literature on character-based traits and skills in college admissions relies on the study of individual differences in terms of intra- and interpersonal skills and traits. Examples include the "Big Five" personality traits: motivation, study skills, self-efficacy, integrity, and leadership (Credé & Kuncel, 2008; Kyllonen et al., 2014; Le et al., 2005; Oswald et al., 2004; Schmitt,

2012; Sedlacek, 2004). Kuncel, Tran, and Zhang (in this volume) provide more detail on character-based tests and some of the challenges associated with them.

In Europe, admissions procedures are generally less oriented toward the psychology of individual differences and related psychological constructs, although there are exceptions, such as the Cito test (see Bartels et al., 2002) for admissions to secondary education in the Netherlands, and the Swe-SAT (Swedish Scholastic Aptitude Test)in Sweden (e.g., Lyren, 2008). This more practice-driven (rather than theory-driven) approach to admissions adopted in Europe may be one of the reasons why there is less scientific literature on the rationale and effectiveness of educational selection outside the United States.

There is limited European-based literature that answers questions such as: Which institutional goals provide the basis on which to select students? Which instruments and criteria are used to select students? How are admissions decisions reached? Nevertheless, literature oriented in US admissions practices also influences the public debate and admissions practices in Europe, including the trend to use alternatives to the traditional admissions criteria such as previous educational performance. Examples are holistic assessment (Allman, 2009; Witzburg & Sondheimer, 2013) and the use of character-based admissions criteria (Stemler, 2012; Sternberg, 2016), assessed through interviews, motivation letters, or questionnaires. However, the research, conclusions, and policies provided in the literature based in the United States do not always provide a good fit to education systems in Europe.

In this chapter, we provide: (a) a brief description of admissions practices in the United States and Europe; (b) different rationales for implementing character-based admissions criteria in admissions testing; (c) the empirical evidence of the validity and fairness of character-based admissions tools; and (d) the process of how academic and character-based admissions criteria are combined.

5.1   A Brief Overview of Admissions Practices

5.1.1   US Practices

When applying to undergraduate programs in the United States, students typically apply to colleges or universities without declaring their major right away, although there are exceptions. The degree of selectivity varies depending on the college or university. Admissions procedures in the

United States are typically based on a meritocracy model (see Wikström & Wikström, in this volume, for a description). Academic merit, usually assessed through scores on standardized tests such as the SAT® or ACT® and high school grades, are the most important determinants of admissions decisions. However, merit in domains such as the arts, music, sports, community service, and personal development are also often considered (National Association for College Admission Counseling, 2017; Zwick, 2017a). Those non-academic merits are often assessed through personal statements, recommendations, essays, résumés, and interviews (for excellent discussions of college admissions in the United States, see Camara, 2009; Zwick, 2017b).

5.1.2   *European Practices*

Providing a complete description of European admissions practices is an almost impossible task, so we do not attempt to do so here. Instead, we try to provide some typical practices and procedures, all of which include many exceptions. We also give examples, and we mostly focus on Western Europe.

European secondary education is often stratified into different levels and completed through centrally organized final exams. The first admissions requirement is usually a high school diploma at the highest level of secondary education, where students complete the most academically demanding track. This leads to a strong pre-selection on scholastic achievement before applying to college, which results in relatively homogenous applicant pools with respect to general cognitive or scholastic skills. Therefore, it is more difficult to differentiate between applicants based on standardized tests that measure such skills (Crombag, Gaff, & Chang, 1975; Resing & Drenth, 2007). That is probably why large-scale standardized tests are rarely used in European admissions (Sweden is the exception with the use of the SweSAT).

When this minimum-level requirement is fulfilled, high school grades play a major role in selective admissions in Europe. However, the degree of selectivity differs widely per country and mostly depends on the study program rather than on the university (England is an exception). In countries such as Denmark, Belgium, France, Luxembourg, and the Netherlands, most programs are not selective and admit all students that have the required high school diploma (Orr et al., 2017).[1] Sometimes,

---

[1] The Grand écoles, the most prestigious higher education institutes in France, are an exception and are highly selective.

additional requirements are in place, such as students having taken specific courses relevant to the course of study, like physics, chemistry, and math for science programs. Medicine and other health-related disciplines are generally selective and are among the most selective programs in many countries. Others that are often (more) selective are psychology programs (in the Netherlands, Germany, Denmark, Finland, and Sweden) and the arts, music, and sport programs, with their own distinct admissions requirements.

In addition to high school grades, discipline-related admissions examinations assessing existing subject-matter knowledge are used in England, Belgium, and the Netherlands. Also, curriculum samples that provide a small simulation of the study program and require substantial preparation are used in the Netherlands, Germany, and Finland (de Visser et al., 2017; Kunina et al., 2007; Lievens & Coetsier, 2002; Niessen, Meijer, & Tendeiro, 2018; Valli & Johnson, 2007).

Admissions procedures in Europe are also mostly merit-based (see Wikström & Wikström, in this volume). Explicit character-based criteria are typically not included in admissions procedures in Europe. When they are included, they are mostly focused on motivation and fit and assessed through motivation letters, personal statements, and interviews. Examples can be found in the Netherlands, Denmark, Germany, Finland, and Sweden and in top universities in England (Cremonini et al., 2011; Steenman, 2018). The use of tests or questionnaires to assess character-based traits and skills such as motivation and personality is also rare; they are most commonly used in the Netherlands (Steenman, 2018). However, character-based admissions is gaining ground in Europe. For example, the German constitutional court has recently decided that the strong emphasis on high school grades in admissions to medical school is not fair and should not be the only criterion. They recommended the use of at least one other non-grade-based admissions criterion (Tagesschau, 2017). In addition, Dutch higher education programs are now required to adopt at least two distinct criteria in selective admissions procedures, one of which is preferably character-based (Ministry of Education, Culture, and Science, 2014).

5.2 The Rationale of Character-Based College Admissions

The aims of college admissions procedures can vary across colleges and countries, ranging from selecting candidates who will perform well academically, selecting candidates to optimize ethnic and social background diversity, or crafting a class of students with a wide variety of special

talents. The three most common arguments to include character-based assessment align with those aims (see Niessen & Meijer, 2017).

5.2.1 *Predictive and Incremental Validity*

The first argument is that they have incremental validity above traditional admissions criteria, such as high school grade point average (GPA) and standardized test scores for predicting academic achievement (Credé & Kuncel, 2008; Oswald et al., 2004; Richardson, Abrahams, & Bond, 2012; Robbins et al., 2004). This seems to be the dominant argument for considering character-based admissions in Europe, where the main aim of admissions procedures seems to be to select those students that will perform best academically (Steenman, 2018). As a result, character-based admissions criteria in Europe are more closely linked to the demands of the specific study program or the future profession. Examples are integrity-based assessments for medical school (de Leng et al., 2018) and motivation for the study program of interest (Busato et al., 2000; Wouters et al., 2017). Such measures are thus mostly aimed at predicting domain-specific academic performance or future job performance in the profession of interest.

5.2.2 *Predicting Broader Outcomes*

A second argument is that character-based admissions criteria are more suitable to predict outcomes beyond academic achievement as defined by first-year GPA. Examples of such broader outcomes are educating future leaders, promoting active citizenship, critical thinking, creativity, and innovation (Oswald et al., 2004; Stemler, 2012; Sternberg, 2016). This is commonly argued in the United States, where the more prestigious universities largely select students to promote institutional and societal goals such as leadership, active citizenship, and athletic performance (Sternberg, 2016; Zwick, 2017a). Therefore, admissions officers may also consider "legacies [i.e., children of alumni]; leaders for school publications, student government, and other areas of student life; children of influential families; those with special talents such as musicians, athletes, and public speakers; and students from under-represented ethnic groups and geographical areas" (Zwick, 2013, p. 15).

5.2.3 *Reducing Adverse Impact*

A third argument is that character-based criteria may reduce adverse impact, increase diversity (Schmitt et al., 2009; Sedlacek, 2005; Sinha

et al., 2011), and reduce selection system bias (Keiser et al., 2016; Mattern, Sanchez, & Ndum, 2017). This argument is also most commonly encountered in the US-based literature. Character-based admissions is typically not implemented to reduce adverse impact in Europe (Orr et al., 2017). One reason may be that achievement in high school is mostly used as an admissions criterion in Europe as opposed to standardized admission-test scores in the United States. Thus, adverse impact starts to play a significant role well before admission to higher education due to the stratification in secondary education in many European countries. For example, in the Netherlands, pupils are placed in one of three main educational tracks (which may have a number of sub-tracks or combination tracks), mostly depending on scholastic performance at the age of 11 or 12. And in Finland pupils are placed in an academic or a vocational track at age 15. Underrepresented minorities are less likely to complete higher-level secondary education or to apply to a university (Lamb et al., 2011; Organisation for Economic Co-operation and Development, 2010). However, there is a growing interest in this perspective, especially in admissions to medical school (Lievens & Coetsier, 2002; Stegers-Jager, 2018). Another example can be found in Denmark, where a small proportion of applicants is admitted based on work experience and motivation in addition to educational achievement. This provides applicants who did not start college directly after finishing high school better chances of admission (Cremonini et al., 2011). The SweSAT was introduced in Sweden for this reason (Cliffordson, 2008), which is interesting given the common adverse-impact-related criticism on cognition-oriented tests.

Broadly speaking, there are two common reasons to include character-based criteria in college admissions. The first reason is improved prediction, mostly of academic performance in Europe, and also of broader outcomes, such as leadership and citizenship, in the United States. The second reason is reducing adverse impact and increasing diversity in colleges.

5.3 Validity and Fairness of Character-Based Admissions Procedures

5.3.1 *Predictive Validity*

The predictive validity of the most common admissions criteria such as high school grades and scores on standardized tests in the United States is well documented (Kuncel & Hezlet, 2010; Westrick et al., 2015; Zwick, 2017b). It is, however, very difficult to obtain empirical studies on the

validity of other frequently used admissions instruments to assess character-based admissions criteria. Common instruments used to assess character-based criteria in admissions are personal statements, interviews, motivation letters, letters of recommendation, and questionnaires (see Kuncel, Tran, & Zhang, in this volume). In general, there is a scarcity of studies that provide empirical evidence for the predictive validity of these kinds of instruments. The studies that are available show that most of these instruments tend to have little predictive and incremental validity for academic achievement (Dana, Dawes, & Peterson, 2013; Goho & Blackman, 2006; Kuncel, Kochevar, & Ones, 2014; Murphy et al., 2009; Patterson et al., 2016).

For assessing character-based admissions criteria, there seems to be too much faith in procedures that seem to be valid (Highhouse, 2008; Jones, 2011). For example, the idea that unstructured in-depth interviews can have high predictive power seems ineradicable among admissions officers and applicants. In a recent study, Niessen, Meijer, and Tendeiro (2017a) investigated the preferences of students for admissions to a psychology program. Interviews were rated most favorably, whereas lottery admissions and high school GPA were rated least favorably. Similarly, Kelly et al. (2018) found favorable stakeholder reactions to interviews and situational judgment tests (SJTs), but less favorable perceptions of cognitive-ability tests and academic records. Dana et al. (2013) showed that information obtained from an unstructured interview, when added to more reliable and valid instruments, could even reduce the predictive power of the assessment procedure. Structured interviews may have value in selection procedures, although most validity studies on the use of structured interviews were conducted in the context of personnel selection (Cortina et al., 2000; Schmidt & Hunter, 1998).

In addition, the use of motivation letters, personal statements, and interviews is generally not theory-driven. However, in the admissions literature, there are studies on carefully designed SJTs, biodata scales, and questionnaires to measure character-based traits and skills, both in the United States (Schmitt, 2012; Shultz & Zedeck, 2012; Wagerman & Funder, 2007) and in Europe (Busato et al., 2000; de Leng et al., 2018; Patterson et al., 2017; Schwager et al., 2014). These studies show that character-based admissions criteria can have incremental validity over traditional admissions criteria (Credé & Kuncel, 2008; Oswald et al., 2004; Richardson et al., 2012; Robbins et al., 2004) and can have predictive validity for broader outcomes like job performance and active citizenship (Oswald et al., 2004; Stemler, 2012; Sternberg, 2016).

However, most of these studies were conducted in low-stakes settings. The generalization of these predictive validity results to high-stakes settings is not straightforward.

5.3.2 Measuring Character in High-Stakes Contexts

One of the main issues in the generalization of research findings to high-stakes admissions procedures is the possibility of faking, due to the self-report nature of most character-based instruments. Griffin and Wilson (2012) showed that applicants scored much more favorably on self-report questionnaires compared to completing the same questionnaire for research purposes. In addition, Niessen, Meijer, and Tendeiro (2017b) found that the predictive and incremental validities of scales measuring personality, study skills, and study habits were substantially lower when they were administered in an admissions context. Anglim et al. (2018), also found lower predictive validity of conscientiousness scores when obtained in a high-stakes admissions context.

The forced-choice format was recently revived as a solution to the faking problem (e.g., Markle et al., 2013; Salgado & Táuriz, 2014), after a scoring method that results in non-ipsative data was designed (Brown & Maydeu-Olivares, 2013). However, the results on whether forced-choice items are indeed substantially more resistant to faking are mixed, and some studies found that faking ability on forced-choice questionnaires depends on cognitive ability. This high cognitive saturation of forced-choice character-based assessments can even increase their predictive validity, but likely hinders their incremental validity over more cognitively loaded criteria (Christiansen, Burns, & Montgomery, 2005; Vasilopoulos et al., 2006).

One of the few studies using a character-based instrument that was conducted in high-stakes admissions testing was the use of SJTs measuring interpersonal skills in admissions to medical school in Belgium (Lievens, 2013; Lievens & Sackett, 2012). The SJT scores were statistically significant but showed low predictive and incremental validity for interpersonal GPA, internship performance, and job-performance. Thus, they would add little utility in terms of increased doctor performance in practice (Niessen & Meijer, 2016).

An alternative method that is not based on self-reports is the multiple mini-interview (MMI). MMIs consist of several highly structured interviews or role plays, typically assessed by multiple examiners or raters (Eva et al., 2004). Moderate-to-high predictive validities were obtained using MMIs in high-stakes admissions to medical school (Husbands & Dowell,

2013). However, as with all observation-based assessments, close attention should be paid to minimizing rater errors, bias, and subjectivity (Till, Myford, & Dowell, 2013) – for example, by using behaviorally anchored rating scales (Lee et al., 2017).

While results obtained in low-stakes contexts are promising, the predictive and incremental validity results based on self-report instruments have thus far not generalized to actual high-stakes admissions procedures (Thomas, Kuncel, & Credé, 2007). Instruments based on actual behavior rather than self-reports show some promising results but are more time-consuming to develop and administer.

5.3.3   Increasing Diversity through Character-Based Admissions Criteria

Another reason to include character-based criteria in admissions procedures is their alleged lower-adverse impact as compared to traditional admissions criteria such as standardized tests and high school GPA (Schmitt et al., 2009; Sedlacek, 2005; Sinha et al., 2011). However, faking and coaching (Ramsay et al., 2006) could pose a threat to realizing this promise as well, due to inequality in resources in support, practice, and preparation (Kyllonen, Walters, & Kaufman, 2005; Zwick, 2017a). In addition, measures with higher cognitive saturation (that is, a strong correlation with cognitive ability) also tend to show more adverse impact (Dahlke & Sackett, 2017). Therefore, the higher cognitive saturation of the more fake-resistant forced-choice items probably also yield more adverse impact (Christiansen et al., 2005). Furthermore, an often-overlooked alternative explanation may be that the lower adverse impact of character-based admissions criteria is an artifact caused by the lower reliability of the instruments used to measure them (see Zwick, 2017b, p. 153).

While adding character-based admissions criteria with smaller mean subgroup differences can have some merit, it probably does not yield such drastic reductions of adverse impact as is often implied (Sackett & Ellingson, 1997). For example, the scores on the SJT developed by Oswald et al. (2004) showed virtually no subgroup differences between Black and White students, and showed no relationship with standardized test scores, which are ideal results in terms of minimizing adverse impact and maximizing incremental validity. Using these results as an example, let us assume that the standardized mean difference (as indicated by Cohen's d) on the SJT for the Black and White applicants equals $d = 0$ on the SJT, and that the correlation between the SJT scores and standardized-test scores equals

$r = 0$. Standardized admissions tests often yield substantial score differences of, say, $d = 1$ between Black and White students (Sinha et al., 2011). Sackett and Ellingson (1997) developed tables to find the resulting d-value of a unit-weighted composite of the two tests (in this case, the SJT and standardized admissions test scores) based on the sum of the d-values of both tests and the correlation between the scores on both tests. Based on Sackett and Ellingson's (1997) tables, we can find that a unit-weight composite based on this ideal example would yield a standardized mean difference of $d = 0.71$. This is only a modest reduction compared to using only the standardized test scores with $d = 1$. So, while adverse impact would be reduced by adding this SJT to the admissions procedure, the effect of adding an instrument that showed no adverse impact at all would be surprisingly small. Furthermore, the resulting d-value of composite measures is lower when the correlation between the scores on both tests is higher (Sackett & Ellingson, 1997). This shows a trade-off between reducing adverse impact and maximizing incremental validity. Adding character-based criteria to assessment procedures that also contain traditional criteria such as standardized tests or high school grades will thus likely have only modest effects on adverse impact.

5.4 Combining Academic and Character-Based Admissions Criteria

Using several different admissions criteria requires the integration of different sources of information to make predictions and decisions. This integration procedure also deserves attention. A popular method to integrate a variety of information about an applicant's abilities, skills, background, and character is often referred to as "holistic assessment" (Horn, 2005; Witzburg, & Sondheimer, 2013; Wouters, 2017). Holistic assessment is based on the idea that by considering all the interactions between relevant information through expert judgment, a good impression of the person as a whole can be obtained, as compared to the limited information provided by standardized test scores (Highhouse & Kostek, 2013). However, as is known from Meehl (1954; also see Dawes, 1979; Highhouse & Kostek, 2013; Kuncel et al., 2013), statistical prediction according to a predefined decision rule is almost always superior to clinical, or holistic, prediction. In clinical prediction, information from different sources is combined (in the mind) to form a hypothesis about a candidate, and then based on this hypothesis: "we arrive at a prediction [as to] what is going to happen" (Meehl, 1954, p. 4). Although the superiority of statistical

prediction is a very solid finding in the psychological decision-making literature, admissions officers from some colleges seem to be proud not to use statistical prediction.

In our view, it is indeed ironic that many stakeholders (admissions officers, candidates, and parents) can be very critical about admissions criteria such as high school grades and standardized test scores, and at the same time unquestioningly embrace alternatives such as an unreliable, unstructured interview (e.g., Allman, 2009) and opaque holistic procedures. As Dana, Dawes, and Peterson (2013) discussed:

> The ability to sensemake combined with the tendency for biased testing allows unstructured interviewers to feel they understand an interviewee almost regardless of the information they receive. Unfortunately, a feeling of understanding, while reassuring and confidence-inspiring, is neither sufficient nor necessary for making accurate assessments. (p. 514)

In addition, we should realize that although traditional admissions criteria like high school grades and standardized tests are often defined as cognitive, they are not pure measures of cognitive or academic ability. A substantial amount of the variance in high school grades can be explained by variables that we would refer to as character-based, such as conscientiousness, grit, and self-efficacy (Borghans et al., 2016; Deary et al., 2007; Dumfart & Neubauer, 2016). Because of "their apparent value in measuring students' tenacity and commitment," Zwick (2017b, p. 193) recommended that high school grades should play a key role in admissions. Even scores on standardized tests have been shown to be related to such character-based traits (Borghans et al., 2011; von Stumm & Ackerman, 2013).

5.5   Conclusion and Discussion

Educational selection should be valid and unbiased. In our view, admissions officers, psychologists, and others who are involved in selective admissions should be very careful when including character-based criteria through methods like interviews, questionnaires, assignments, and holistic evaluations, because they can easily be misused. As Zwick (2017b) stated: "the less clear the admission criteria, the more likely they are to benefit the wealthier, more savvy candidates" (p. 193).

Recently, several medical institutions and federations adopted lists of lower-value services – that is, healthcare that is considered to be of no or limited value. These ineffective medical activities include some medical surgeries, tests, and procedures (Wammes et al., 2016). These activities had no scientific basis but were based on what was considered common

sense and tradition. It was advised not to use these procedures at all or not to use them routinely. As Wammes et al. (2016) indicated, the quality of healthcare is reflected by "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (p. 2). There is a parallel between this medical practice and the practice of college admissions. In our view, admissions procedures are guided too much by intuition-based decisions, and there is a large gap between what we know about optimal decision-making and decisions made in practice, especially when it comes to character-based assessment. This scientist–practitioner gap has received a lot of attention in the field of personnel selection (Anderson, Herriot, & Hodgkinson, 2001; Drenth, 2008) and we think that it is also time to address this issue in higher education admissions. It is, therefore, extremely important that colleges should be transparent about how they select students and how their admissions criteria relate to later performance or other outcomes that one would like to predict. Zwick (2017a) warned against the risks of the use of character-based admissions criteria and suggested first to investigate the possible implications of these admissions criteria on the behavior of different stakeholders like students and parents and to conduct pilot studies to evaluate character-based admissions criteria.

Many studies have shown that the most common admissions criteria – high school GPA and standardized test scores – are good predictors of academic performance, although admittedly they have their drawbacks. Character-based admissions criteria may be able to deal with those drawbacks and, therefore, may be rightfully defined as the next frontier in college admissions (Hoover, 2013).

However, the main problem is that we need much more research to find effective methods to measure character in high-stakes testing. In addition, it is surprising that there is such little evidence available about admissions procedures outside the United States and that there are few studies that provide compelling evidence that other commonly used or recommended admissions criteria predict academic performance or other relevant outcomes. In our view, we should address this scientist–practitioner gap by conducting more studies in operational–admissions settings in collaboration with admissions officers and emphasizing that procedures with such high-societal impact be evidence-based. We advocate for more transparency by sharing information and results in international journals or other platforms to build a broader evidence-based educational-selection literature that can inform further research, policies, and practice, especially for character-based measures.

REFERENCES

Allman, M. (2009). Quintessential questions: Wake Forest's director gives insight into the interview process. Retrieved from http://blog.rethinkingadmissions .wfu.edu/2009/08/quintessential-questions-wake-forest%E2%80%99s-admission-director-gives-insight-into-the-interview-process/.

Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner–researcher divide in industrial, work and organizational (IWO) psychology: Where we are now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*, 391–411. https://doi.org/10.1348/096317901167451.

Anglim, J., Bozic, S., Little, J., & Lievens, F. (2018). Response distortion on personality tests in applicants: Comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education: Theory and Practice*, *23*, 311–321. https://doi.org/10.1007/s10459-017-9796-8.

Bartels, M., Rietveld, M. H., Van Baal, G. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research*, *5*, 544–553. https://doi.org/10.1375/136905202762342017.

Berry, C. M. (2015). Differential validity and differential prediction of cognitive ability tests: Understanding test bias in the employment context. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 435–463. https://doi.org/10.1146/annurev-orgpsych-032414-1112566.

Borghans, L., Golsteyn, B. H., Heckman, J., & Humphries, J. E. (2011). Identification problems in personality psychology. *Personality and Individual Differences*, *51*, 315–320. https://doi.org/10.1016/j.paid.2011.03.029.

(2016). What grades and achievement tests measure. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *113*, 13354–13359. https://doi.org/10.1073/pnas.1601135113.

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, *18*, 36–52. https://doi.org/10.1037/a0030641.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, *29*, 1057–1068. https://doi.org/10.1016/S0191-8869(99)00253-6.

Camara, W. J. (2009). College admission testing: Myths and realities in an age of admissions hype. In R. P. Phelps (Ed.). *Correcting fallacies about educational and psychological testing* (pp. 147–180). Washington, DC: American Psychological Association. https://doi.org/10.1037/11861-004.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance*, *18*, 267–307. https://doi.org/10.1207/s15327043hup1803_4.

Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment*, *13*, 56–75. https://doi.org/10.1080/10627190801968240.

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology*, *53*, 325–351. https://doi.org/10.1111/j.1744-6570.2000.tb00204.x.

Credé, M. & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate performance. *Perspectives on Psychological Science*, *3*, 425–453. https://doi.org/10.1111/j.1745-6924.2008.00089.x.

Cremonini, L., Leisyte, L., Weyer, E., & Vossensteyn, J. J. (2011). *Selection and matching in higher education: An international comparative study*. Enschede: Center for Higher Education Policy Studies (CHEPS).

Crombag, H. F., Gaff, J. G., & Chang, T. M. (1975). Study behavior and academic performance. *Tijdschrift voor Onderwijsresearch*, *1*, 3–14.

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology*, *102*, 1403–1420. https://doi.org/10.1037/apl0000234.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgment and Decision Making*, *8*, 512–520.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision-making. *American Psychologist*, *34*, 571–582. https://doi.org/10.1037/0003-066X.34.7.571.

de Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. N. (2018). Integrity situational judgment test for medical school selection: Judging "what to do" versus "what not to do." *Medical Education*, *52*, 427–437. https://doi.org/10.1111/medu.13498.

de Visser, M., Fluit, C., Fransen, J., Latijnhouwers, M., Cohen-Schotanus, J., & Laan, R. (2017). The effect of curriculum sample selection for medical school. *Advances in Health Science Education*, *22*, 43–56. https://doi.org/10.1007/s10459–016-9681-x.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21. https://doi.org/10.1016/j.intell.2006.02.001.

Drenth, P. J. D. (2008). Psychology: Is it applied enough? *Applied Psychology: An International Review*, *57*, 524–540. https://doi.org/10.1111/j.1464-0597.2008.00337.x.

Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful noncognitive predictor of school achievement in adolescents. *Journal of Individual Differences*, *37*, 8–15. https://doi.org/10.1027/1614-0001/a000182.

Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004). The ability of the multiple mini-interview to predict pre-clerkship performance in medical

school. *Academic Medicine*, *79*, 40–42. https://doi.org/10.1097/00001888-200410001-00012.

Goho, J., & Blackman, A. (2006). The effectiveness of academic admission interviews: An exploratory meta-analysis. *Medical Teacher*, *28*, 335–340. https://doi.org/10.1080/01421590600603418.

Griffin, B. & Wilson, I. G. (2012). Faking good: Self-enhancement in medical school applicants. *Medical Education*, *46*, 485–490. https://doi.org/10.1111/j.1365-2923.2011.04208.x.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x.

Highhouse, S., & Kostek, J. A. (2013). Holistic assessment for selection and placement. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.). *APA handbook of testing and assessment in psychology* (Vol. 1) *Test theory and testing and assessment in industrial and organizational psychology* (pp. 565–577). Washington, DC: American Psychological Association. https://doi.org/10.1037/14047-031.

Hoover, E. (2013). Noncognitive measures: The next frontier on college admissions. *Chronicle of Higher Education.* Retrieved from www.chronicle.com/article/Noncognitive-Measures-The/136621.

Horn, C. (2005). Standardized assessments and the flow of students into the college admission pool. *Educational Policy*, *19*, 331–348. https://doi.org/10.1177/0895904804274057.

Husbands, A., & Dowell, J. (2013). Predictive validity of the Dundee Multiple Mini-Interview. *Medical Education*, *47*, 717–725. https://doi.org/10.1111/medu.12193.

Jones, B. M. (2011). Assessing admission interviews at residential STEM schools. *NCSSSMST Journal*, *16*, 20–29.

Keiser, H. N., Sackett, P. R., Kuncel, N. R., & Brothen, T. (2016). Why women perform better in college than admission scores would predict: Exploring the roles of conscientiousness and course-taking patterns. *Journal of Applied Psychology*, *101*, 569–581. https://doi.org/10.1037/apl0000069.

Kelly, M. E., Patterson, F., O'Flynn, S., Mulligan, J., & Murphy, A. W. (2018). A systematic review of stakeholder views of selection methods for medical schools admission. *BMC Medical Education*, *18*, 139. https://doi.org/10.1186/s12909-018-1235-x.

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science*, *19*, 339–345. https://doi.org/10.1177/0963721410389459.

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*, 1060–1072. https://doi.org/10.1037/a0034156.

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for

hope. *International Journal of Selection and Assessment*, 22, 101–107. https://doi.org/10.1111/ijsa.12060.

Kunina, O., Wilhelm, O., Formazin, M., Jonkmann, K., & Schroeders, U. (2007). Extended criteria and predictors in college admission: Exploring the structure of study success and investigating the validity of domain knowledge. *Psychology Science*, 49, 88–114.

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). Personality, motivation, and college readiness: A prospectus for assessment and development (ETS Research Report No. RR-14-06). https://doi.org/10.1002/ets2.12004.

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2005). Noncognitive constructs and their assessment in graduate education: A review. *Educational Assessment*, *10*, 153–184. https://doi.org/10.1207/s15326977ea1003_2.

Lamb, S., Markussen, E., Teese, R., Polesel, J., & Sandberg, N. (2011) *School dropout and completion*. Dordrecht: Springer. https://doi.org/10.1007/978-90-481-9763-7.

Le, H., Casillas, A., Robbins, S. B., & Langley, R. (2005). Motivational and skills, social, and self-management predictors of college outcomes: Constructing the student readiness inventory. *Educational and Psychological Measurement*, *65*, 482–508. https://doi.org/10.1177/0013164404272493.

Lee, J., Connelly, B. S., Goff, M., & Hazucha, J. F. (2017). Are assessment center behaviors' meanings consistent across exercises? A measurement invariance approach. *International Journal of Selection and Assessment*, 25, 317–332. https://doi.org/10.1111/ijsa.12187.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York, NY: Farrar, Straus, and Giroux.

Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgment tests. *Medical Education*, 47, 182–189. https://doi.org/10.1111/medu.12089.

Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, *10*, 245–257. https://doi.org/10.1111/1468-2389.00215.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*, 460–468. https://doi.org/10.1037/a0025741.

Lyren, P. (2008). Prediction of academic performance by means of the Swedish scholastic assessment test. *Scandinavian Journal of Educational Research*, 52, 565–581. https://doi.org/10.1080/00313830802497158.

Markle, R., Olivera-Aguilar, M., Jackson, T., Noeth, R., & Robbins, S. (2013). Examining evidence of reliability, validity, and fairness for the SuccessNavigator™ assessment (ETS Research Report No. RR-13-12). Retrieved from www.ets.org/Media/Research/pdf/RR-13-12.pdf.

Mattern, K., Sanchez, E., & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice*, *36*(1), 47–57. https://doi.org/10.1111/emip.12138.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota

Ministry of Education, Culture, and Science (2014, August 29). *Informatie over de afschaffing van loting bij numerusfixusopleidingen* [Information concerning the abolishing of lottery admission]. Kamerbrief [Government information]. Retrieved from www.rijksoverheid.nl/documenten-en-publicaties/kamerstuk ken/2014/08/30/kamerbrief-met-informatie-over-de-afschaffing-van-loting-bij-numerusfixusopleidingen.html.

Murphy, S. C., Klieger, D. M., Borneman, M. J., & Kuncel, N. R. (2009). The predictive power of personal statements in admissions: A meta-analysis and cautionary tale. *College and University*, *84*(4), 83–86.

National Association for College Admission Counseling (2017). *State of College Admission*. Retrieved from www.nacacnet.org/globalassets/documents/publi cations/research/soca17final.pdf.

Niessen, A. S. M., & Meijer, R. R. (2016). Selection of medical students on the basis of non-academic skills: Is it worth the trouble? *Clinical Medicine*, *16*, 339–342. https://doi.org/10.7861/clinmedicine.16-4-339.

(2017). On the use of broadened selection criteria in higher education. *Perspectives on Psychological Science*, *12*, 436–448. https://doi:10.1177/1745691616683305.

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017a). Applying organizational justice theory to admission into higher education: Admission from a student perspective. *International Journal of Selection and Assessment*, *25*, 72–84. https://doi.org/10.1111/ijsa.12161.

(2017b). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality and Individual Differences*, *106*, 183–189. https://doi.org/10.1016/j.paid.2016.11.014.

(2018) Admission testing for higher education: A multi-cohort study on the validity of high-fidelity curriculum-sampling tests. *PLoS ONE* *13*(6), 1–21. https://doi.org/10.1371/journal.pone.0198746.

Organisation for Economic Co-operation and Development (2010). *OECD reviews of migrant education: Closing the gap for immigrant students; policies, practice and performance*. Paris: Organisation for Economic Co-operation and Development.

Orr, D., Usher, A., Haj, C, Atherton, G, & Geanta, I. (2017). *Study on the impact of admission systems on higher education outcomes* (Vol. 1): *Comparative report*. Brussels: European Commission. Retrieved from https://publications .europa.eu/en/publication-detail/-/publication/9cfdd9c1-98f9-11e7-b92d-01aa75ed71a1.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as

predictors of college student performance. *Journal of Applied Psychology*, *89*, 187–207. https://doi.org/10.1037/0021-9010.89.2.187.

Patterson, F., Cousans, F., Edwards, H., Rosselli, A., Nicholson, S., & Wright, B. (2017). The predictive validity of a text-based situational judgment test in undergraduate medical and dental school admissions. *Academic Medicine*, *92*, 1250–1253. https://doi.org/10.1097/ACM.0000000000001630.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education*, *50*, 36–60. https://doi.org/10.1111/medu.12817.

Ramsay, L. J., Schmitt, N., Oswald, F. L., Kim, B. H., & Gillespie, M. A. (2006). The impact of situational context variables on responses to biodata and situational judgment inventory items. *Psychology Science*, *48*, 268–287.

Resing, W. C. M., & Drenth, P. J. D. (2007). *Intelligentie: Weten en meten* [Intelligence: Measuring and knowing]. Amsterdam: Uitgeverij Nieuwezijds.

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

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*, 261–288. https://doi.org/10.1037/0033-2909.130.2.261.

Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, *50*, 707–721. https://doi.org/10.1111/j.1744-6570.1997.tb00711.x.

Salgado, J. F., & Táuriz, G. (2014). The five-factor model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, *23*, 3–30. https://doi.org/10.1080/1359432X.2012.716198.

Schmidt, F. L., & Hunter, J. E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. https://doi.org/10.1037/0033-2909.124.2.262.

Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist*, *47*, 18–29. https://doi.org/10.1080/00461520.2011.610680.

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*, 1479–1497. https://doi.org/10.1037/a0016810.

Schwager, I. T. L., Hülsheger, U. R., Lang, J. W. B., Klieger, D. M., Bridgeman, B., & Wendler, C. (2014). Supervisor ratings of students' academic potential as predictors of citizenship and counterproductive behavior. *Learning*

and Individual Differences, *35*, 62–69. https://doi.org/10.1016/j.lindif
.2014.07.005.

Sedlacek, W. E. (2004) *Beyond the big test: Noncognitive assessment in higher
education*. San Francisco, CA: Jossey-Bass.

 (2005). The case for noncognitive measures. In W. J. Camara, E. W. Kimmel,
W. J. Camara, & E. W. Kimmel (Eds.). *Choosing students: Higher education
admissions tools for the 21st century* (pp. 177–193). Mahwah, NJ: Lawrence
Erlbaum Associates Publishers.

Shultz, M. M., & Zedeck, S. (2012). Admission to Law school: New
measures. *Educational Psychologist*, *47*, 51–65. https://doi.org/10.1080/
00461520.2011.610679.

Sinha, R., Oswald, F., Imus, A., & Schmitt, N. (2011). Criterion-focused approach
to reducing adverse impact in college admissions. *Applied Measurement in
Education*, *24*, 137–161. https://doi.org/10.1080/08957347.2011.554605.

Steenman, S. C. (2018). Alignment of admission: An exploration and analysis of
the links between learning objectives and selective admission to programmes
in higher education (Doctoral dissertation). University of Utrecht.

Stegers-Jager, K. M. (2018). Lessons learned from 15 years of non-grades-based
selection for medical school. *Medical Education*, *52*, 86–95. https://doi.org/
10.1111/medu.13462.

Stemler, S. E. (2012). What should university admissions tests predict? *Educational
Psychologist*, *47*, 5–17. https://doi.org/10.1080/00461520.2011.611444.

Sternberg, R. J. (2016). *What universities can be: A new model for preparing
students for active concerned citizenship and ethical leadership*. Ithaca, NY:
Cornell University Press.

Tagesschau (2017). Medizin-Zulassung muss überarbeitet werden [Admission to
medical school must be revised]. Retrieved from www.tagesschau.de/inland/
medizinstudium-verfassungsgericht-101.html.

Thomas, L. L., Kuncel, N. R., & Credé, M. (2007). Noncognitive variables in
college admissions: The case of the non-cognitive questionnaire. *Educational
and Psychological Measurement*, *67*, 635–657. https://doi.org/10.1177/
0013164406292074.

Till, H., Myford, C., & Dowell, J. (2013). Improving student selection using
multiple mini-interviews with multifaceted Rasch modeling. *Academic Medi-
cine*, *88*, 216–223. https://doi.org/10.1097/ACM.0b013e31827c0c5d.

Valli, R., & Johnson, P. (2007). Entrance examinations as gatekeepers. *Scandi-
navian Journal of Educational Research*, *51*, 493–510. https://doi.org/
10.1080/00313830701576631.

Vasilopoulos, N. L., Cucina, J. M., Dyomina, N. V., Morewitz, C. L., & Reilly,
R. R. (2006). Forced-choice personality tests: A measure of personality and
cognitive ability? *Human Performance*, *19*, 175–199. https://doi.org/
10.1207/s15327043hup1903_1.

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review
and meta-analysis. *Psychological Bulletin*, *139*, 841–869. https://doi.org/
10.1037/a0030746.

Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, *41*, 221–229. https://doi.org/10.1016/j.jrp.2006.03.001.

Wammes, J. G., van den Akker-van Marle, M. E., Verkerk, E. W., van Dulmen, S. A., Westert, G. P., van Asselt, A. I., & Kool, R. B. (2016). Identifying and prioritizing lower value services from Dutch specialist guidelines and a comparison with the UK do-not-do list. *BMC Medicine*, *14*, 1–9. https://doi.org/10.1186/s12916–016-0747-7.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, *20*, 23–45. https://doi.org/10.1080/10627197.2015.997614.

Witzburg, R. A., & Sondheimer, H. M. (2013). Holistic review: Shaping the medical profession one applicant at a time. *New England Journal of Medicine*, *368*, 1565–1567. https://doi.org/10.1056/NEJMp1300411.

Wouters, A. (2017). Effects of medical school selection on the motivation of the student population and the applicant pool (Doctoral dissertation). VU University, Amsterdam.

Wouters, A., Croiset, G., Schripsema, N. R., Cohen-Schotanus, J., Spaai, G. G., Hulsman, R. L., & Kusurkar, R. A. (2017). A multi-site study on medical school selection, performance, motivation and engagement. *Advances in Health Sciences Education: Theory and Practice*, *22*, 447–462. https://doi.org/10.1007/s10459-016-9745-y.

Zwick, R. (2013). *Disentangling the role of high school grades, SAT scores, and SES in predicting college achievement*. (Research Report No. RR-13-09). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/j.2333-8504.2013.tb02316.x.

  (2017a). The risk of focusing on character in admissions. *Chronicle of Higher Education*, *63*(42). Retrieved from www.chronicle.com/article/The-Risks-of-Focusing-on/240787.

  (2017b). *Who gets in? Strategies for fair and effective college admissions*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674977648.