

# Chapter 5

## Selection Practices and Methods in Other Fields

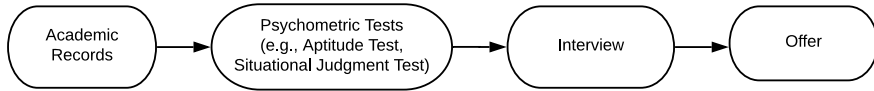


**Abstract** The research base for selection into teacher education programs and teaching practice is only recently emerging (Klassen & Kim, 2019; Klassen et al., 2017). In this light, reviewing selection practices and methods used in other fields—especially those where the methods are well-developed and well-researched—provides a lens through which to view and consider teacher selection. Various selection methods have been used to select individuals into educational (training) programs and into employment. Though the methods used in other fields have some degree of overlap with each other, each area also has its own distinct methods and research base that characterize the field. As such, in this chapter, we will review the practices and the evidence base for the methods that are used to select individuals into medical schools, law schools, and into large organizations.

### 5.1 Selection Practices in Other Fields

**Selection into medical schools.** Much research has been dedicated to investigating the validity of selection methods in order to identify candidates who are most likely to be successful in medical training and progress to become competent clinicians. One of the distinctive characteristics of medical school selection procedures is the quantity and quality of the evidence used to make the selection decisions. Using evidence-informed selection methods is important for medical school selection researchers and practitioners, at least partly due to the high levels of competition for admission and the evidence-based practice ethos prevalent in modern medicine.

Although selection approaches vary across medical schools and countries, a typical selection procedure consists of the following stages (shown in Fig. 5.1). First, prior academic records are used as a screening tool. Next, applicants' documents and test scores are reviewed, which often includes a review of their personal statements, references, and psychometric test scores. Psychometric tests can include aptitude tests to test academic attributes (cognitive attributes associated with academic skills and abilities), sometimes including tests of numerical ability and verbal ability. Non-cognitive attributes are often also assessed, which are attributes not associated with academic skills and abilities; otherwise known as non-academic competencies, 'soft'



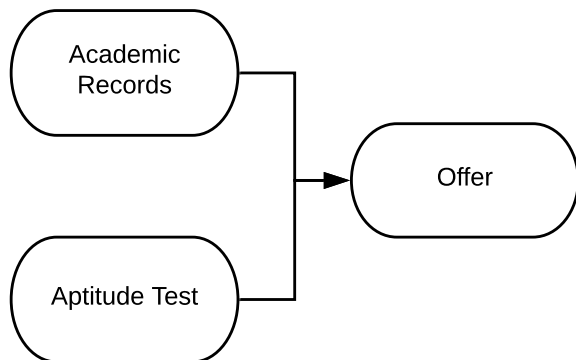
**Fig. 5.1** A simplified model of a medical school selection process

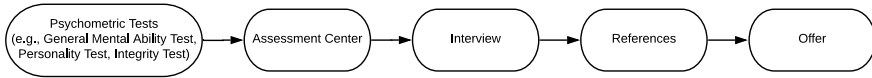
skills, or people skills. Situational judgment tests (SJTs; discussed in a later section of the chapter) are increasingly being used at this stage of selection to assess non-cognitive attributes. After first-stage or screening tests, eligible applicants are then invited to the final stage of selection, which is usually an interview of some format such as a multiple mini-interview (MMI), although assessment centers (with multiple activities) are also sometimes used. Each medical school weights the scores from the various stages differently. From this total score, a candidate pool is formed and training places are offered.

**Selection into law schools.** In some countries, passing the law school selection process is claimed to be the most difficult stage in entering the legal profession (Shultz & Zedeck, 2012). Unlike research in medical school selection where the validity and reliability of a variety of selection methods have been widely explored by multiple researchers, such breadth of research seems to be missing in law school selection. Existing research is largely focused on the academic predictors of first year GPA during law school and not so much on the non-cognitive predictors. Moreover, there is limited research that considers a range of outcomes beyond academic success in law school, such as job performance and job satisfaction.

Unlike the multi-stage process in medical school selection, law school selection often consists of one stage—the assessment of prior academic records and an aptitude test (see Fig. 5.2). Since the 1940s, two scores have largely determined law school selection decisions in the US: undergraduate grade point average (UGPA, as a measure of academic records) and the Law School Admission Test (LSAT, as a measure of aptitudes; Shultz & Zedeck, 2012). Both of these measures are assessments of applicants’ academic attributes. Like medical school selection, each law school determines the weighting of each of the scores (i.e., UGPA and LSAT).

**Fig. 5.2** A simplified model of law school selection process





**Fig. 5.3** A simplified model of personnel selection process

**Personnel selection in large organizations.** Selection programs in large organizations often include a wide variety of predictors and a great range of outcome measures. Organizations often tailor their selection methods to their unique organizational factors, such as size, context, and the nature of the tasks required by the various jobs. It is sometimes the case that employers choose their selection methods based on idiosyncratic preferences or tradition rather than the reliability and validity of the methods (Anderson & Witvliet, 2008).

Despite the absence of a commonly agreed set of selection methods, many organizations follow similar stages of practice for selection (Fig. 5.3). After an initial screen of resumés and job applications to assess applicants' suitability for the job, applicants may be invited to complete some psychometric tests, such as those assessing general mental ability, personality, and/or integrity. Applicants may then be invited to assessment centers, in which they demonstrate how they would handle various work situations through a range of activities such as group tasks, group interviews, and individual tasks. Individual interviews often form the last stage of selection, after which applicants' references are checked for cross-verification purposes.

## 5.2 Selection Methods and Their Evidence

Multiple review articles and meta-analyses have been published on the validity of selection methods used in each field (e.g., Lievens et al., 2021; Patterson et al., 2016; Shultz & Zedeck, 2012). A selection method can be used to select *out* individuals (e.g., a screening of academic records to reduce the pool), select *in* individuals (e.g., conducting structured interviews to make selection decisions), or to verify applicant details (e.g., reference checks). In the next section, we will summarize the evidence behind a range of selection methods that are used frequently in each of the three fields of interest: selection into training programs in medicine and law, and selection into large organizations (personnel selection). A summary of the evidence can be found in Table 5.1.

**Academic records.** Records of prior academic achievement are frequently considered as part of selection decision-making for educational programs and employment. Forms of academic records used for selection include grades from secondary school, if applying for undergraduate programs; and grades from both secondary school and from undergraduate study, if applying for graduate school programs or jobs.

Academic records are generally used for selection in two ways. First, academic records are used as an indicator of cognitive ability or used in conjunction with other information (e.g., resumés) that highlight relevant experience. Here, a record

**Table 5.1** An interpretation of the wider literature on various selection methods

Selection method	Summary of evidence
Academic records	Most evidence is from educational program selection, with moderate predictive validity. Issues of fairness for minority ethnic groups
Aptitude tests	Most evidence is from education program selection. Good evidence of predictive validity and reliability. Issues of fairness for minority ethnic groups
General Mental Ability (GMA) tests	Most evidence is from personnel selection. Good evidence of predictive validity and reliability. Issues of fairness for minority ethnic groups
Personality tests	Most evidence is from personnel selection. Good evidence of predictive validity, reliability, and fairness
Integrity tests	Most evidence is from personnel selection. Good evidence of predictive validity and reliability
Situational judgment tests	Most evidence is from medical education and personnel selection. Good evidence of predictive validity and fairness. Low-to-moderate reliability given their multidimensionality
Interviews (including MMIs)	Structured interviews (particularly MMIs) have strong evidence of predictive validity, reliability, and fairness
Assessment centers	Most evidence is from personnel selection. High face validity but little incremental validity. Some evidence of reliability. Issues of fairness for minority ethnic groups
Reference checks	Used mostly for face validity. Limited evidence of predictive validity and fairness. Low reliability

of minimum level of academic achievement is used for screening. Second, academic records are used to give priority to applicants with higher levels of academic achievement and used as an indicator of the quality of the accepted applicants.

For medical school selection, most studies indicate that applicants' academic records are a good predictor of academic and clinical performance. For example, studies found that applicants' prior academic records reliably predicted various measures of success during and after medical training, including medical school academic grades (McManus et al., 2013), licensing examination scores (Julian, 2005), internship performance (Ferguson et al., 2002), and career progression (McManus et al., 2003). Moreover, a meta-regression study found that secondary school academic records in the UK (A-levels) were a stronger predictor of medical students' first year academic performance than aptitude test scores (McManus et al., 2013). However, the strength of prediction declined throughout the undergraduate and postgraduate years, though it was still statistically significant.

First year academic performance in law school is the criterion outcome that a majority law school selection studies use to assess the validity of selection tools. Studies have found that undergraduate academic performance was a relatively good predictor of performance in first year of law school, with correlations ranging from

0.26 to 0.29 (Law School Admission Council, 2019). Such empirical relationships are unsurprising since it is established that an academic measure is moderately associated with future measures of academic success (e.g., Geiser & Santelices, 2007). Unlike research in medical school selection research, limited research has been conducted on how law students' UGPA predicts outcomes beyond law school. An exception is a study by Lempert et al. (2000), which found that law students' UGPA did not predict post-law school measures of success (i.e., income, career satisfaction, and service contributions).

Academic records are sometimes used for personnel selection. Although sparsely researched, there is evidence indicating that academic records can add useful information for selection. A meta-analytic study by Roth et al. (1996) found that academic records were associated with job performance with a corrected correlation in the mid 0.30. Moreover, the relationship was stronger the closer to the time job performance was measured to the time of the GPA, which indicate that academic records are reasonable proximal predictors (i.e., they are better at predicting outcomes the closer they are to each other temporally). However, it has been noted that using college academic performance for selection may create an adverse impact given issues of ethnic group differences in GPA scores (Roth & Bobko, 2000; see Chap. 4 for discussion on adverse impact).

In sum, academic records seem to be most helpful for predicting outcomes that are academic in nature, though care must be taken for selection use as it can disadvantage certain minority groups. Furthermore, since the sizes of the associations tend to be stronger the closer the timing is between the measurement of the academic record and the criterion outcome, considering the time lag between two measurements should be considered.

**Psychometric tests.** Using psychometric tests can be an efficient way to screen applicants before more resource-intensive selection methods (e.g., interviews or assessment centers) are used to further select individuals to the next stages of the selection process. Psychometric tests can be relatively easy to administer and score, especially when they are computer-administered. There are generally two types of psychometric tests: tests that assess one's academic attributes (e.g., aptitude tests, general mental ability tests) and tests that assess one's non-academic attributes (e.g., personality tests, integrity test, and situational judgment tests). The evidence behind each of these methods will be outlined below.

**Aptitude tests.** Aptitude tests are assessments that examine qualities important for a skill, job, and/or field. In personnel selection, aptitude tests that examine specific aptitudes (e.g., verbal ability and numerical ability) as well as general aptitudes (i.e., general mental ability; discussed in a later section of the chapter) are used. In medical and law school selection, aptitude tests tend to measure multiple aptitudes, including subject-specific areas. In this section, we will focus on the aptitude tests used in medical school and law school selection.

In medical school selection, the aptitude tests aim to assess the academic and non-cognitive attributes associated with medical school academic and clinical performance. The type of aptitude test that applicants must sit depends on the country in which they wish to study medicine, and sometimes the level of entry. For example,

applicants for British medical schools must sit either the University Clinical Aptitude Test (UCAT) or the Biomedical Admissions Test (BMAT), whereas applicants for Australian medical schools must sit either the Undergraduate Medicine and Health Sciences Admissions Test (UMAT) for undergraduate medical programs or the Graduate Medical Schools Admissions Test (GAMSAT) for graduate medical programs. In North America, the Medical College Admissions Test (MCAT) is most commonly used for admissions. These tests share common ground in assessing a range of similar aptitudes but also have varying emphases on different aptitudes. For example, the UCAT includes subtests of *verbal reasoning*, *decision making*, *quantitative reasoning*, *abstract reasoning*, and *situational judgment*, but no specific scientific knowledge (UCAT, 2021). The MCAT includes questions that assess *critical analysis and reasoning skills*, but also those that assess knowledge of concepts and principles associated with medicine (e.g., biology, biochemistry, and psychobiology; AAMC, 2021).

Despite the wide usage of aptitude tests to select individuals into medical schools, the evidence of their validity is mixed. Some studies report that aptitude tests do predict academic and clinical performance (e.g., Puddey & Mercer, 2014) and do so above and beyond prior academic records (e.g., McManus et al., 2013; Sartania et al., 2014). On the other hand, some studies report that aptitude tests do not predict academic and clinical performance (e.g., Yates & James, 2010). Discrepancies in the predictive validity of medical school entrance exams indicate that closer examinations of the content of aptitude test sections may be needed.

Research on students' perceptions of the fairness and usefulness of medical school selection aptitude tests seem to report mixed findings. Studies have reported that students viewed aptitude tests as neither fair nor useful (Dhar et al., 2012), some viewed them as generally useful and suitable for selection (Cleland et al., 2011), while others viewed only particular sections of the aptitude tests (e.g., logical reasoning and problem solving, or interpersonal understanding) as useful and well-designed (Stevens et al., 2014). In a summary of evidence of medical school selection tools, Patterson et al. (2016) reported that evidence on the fairness and usefulness of aptitude tests seemed mixed. They advised that closer examination may be necessary for each section of the aptitude tests as well as for specific types of aptitude tests (e.g., UCAT, BMAT) rather than generalizing to all aptitude tests. Nevertheless, they concluded that aptitude tests are good screeners for medical school selection.

For law school selection, the type of aptitude test used varies by country and institution. In the US, Canada, and some universities in Australia, the Law School Admission Test (LSAT) is a required component of the admissions process. In the UK, the Law National Aptitude Test is used for admissions by some law schools. As LSAT is the most researched type of law school aptitude test, the LSAT will be the focus of this section.

The LSAT has been in use since 1948 (LaPiana, 2004) and consists of sections testing *analytical and logical reasoning skills*, *reading comprehension skills*, and *writing skills* (Law School Admissions Council, 2021). The LSAT was designed to predict first year GPA (LaPiana, 2004), which is supported by findings that the correlation between average LSAT and first year GPA ranges between 0.34 and 0.41,

compared to the correlation between UGPA and first year GPA that ranges between 0.26 and 0.29 (Law School Admissions Council, 2019). Although academic attributes are known to predict job performance and job knowledge in general, there is limited evidence that LSAT scores predict performance in law school beyond the first year in the program.

The issue of fairness in using LSAT scores to make law school selection decisions has been a source of debate (Holmquist et al., 2014), largely due the findings that LSAT scores seem to differ between the ethnicities of the applicants. For example, findings from the LSAT administered between 2007–2008 and 2013–2014 indicate that Caucasian test-takers consistently had the highest mean scores, followed by Asian/Pacific Islander test-takers, with Puerto Rican test-takers consistently had the lowest (Dalessandro et al., 2014). Associated with this issue is the finding that LSAT scores tend to overpredict first year GPA performance for minority test-takers, indicating that LSAT scores may have differential powers in predicting academic performance depending on the test-takers' race/ethnicity (see Kidder, 2001 for review of studies). As such, heavy emphasis on the LSAT for admissions has been identified as a factor that is restricting the diversity of individuals in legal education and in the legal profession. Including assessments of non-academic attributes as part of the aptitude test suite has been identified as a way to reduce these adverse effects (Holmquist et al., 2014).

**General mental ability tests.** Psychometric tests of general mental ability (GMA; otherwise known as general cognitive ability or intelligence) aim to assess a composite of multiple cognitive abilities. GMA has commonly been used in personnel selection, although aspects of GMA are included in admissions tests for medical schools and law schools. Multiple meta-analyses (e.g., Lang et al., 2010) indicate that GMA predicts both overall job performance and specific job performance dimensions (e.g., task performance, contextual performance, counterproductive work behavior). However, GMA is not as predictive for less complex jobs (Gottfredson, 1997). Nevertheless, tests of GMA seem to be generalizable and valid predictors of a range of outcomes, including academic performance, career potential, creativity, and job performance (Kuncel et al., 2004). The reliability of GMA tests is one of the highest of the selection methods, often in the 0.80 to 0.90 (Ones et al., 2012). Using multiple cognitive ability tests (e.g., numerical ability, literacy skills) to capture GMA is encouraged rather than a reliance on a single test (Lubinski, 2000). Overall, GMA tests seem to be useful for inclusion in selection programs given that cognitive ability is a good predictor of multiple outcomes.

**Personality tests.** Personality tests aim to examine the underlying non-cognitive attributes influencing thoughts, feelings, and behaviors (John et al., 2008). In the context of medical school selection, whether personality tests should be used for selection has been a source of debate (Patterson et al., 2016). On the one side, the 'explicit' nature of many personality tests is concerning. As applicants can often know what each of the items are measuring, issues such socially desirable response patterns arise (e.g., Lievens & Sackett, 2017). However, there is evidence that medical school students' personality is associated with their outcomes during medical training and beyond. For example, McLarnon et al. (2017) examined the utility of including

personality in addition to the traditional predictors for medical school selection (i.e., MCAT, GPA, and scores from a semi-structured panel interviews). They found that personality predicted medical school academic performance and clinical performance above and beyond these traditional predictors and the regression coefficients were stronger. Furthermore, personality was the only significant predictor of clinical performance, providing validity evidence for the use of personality for selection. Given the empirical relationship between personality and important outcomes in medicine, more research into their potential use in medical school selection is warranted.

Personality tests have rarely been included in research on law school admissions. An exception is a study by Shultz and Zedeck (2011), who found that the personality test subscale scores correlated with more lawyer effectiveness criterion outcomes than the LSAT and UGPA scores, and the size of the correlations was bigger. This finding suggests that personality tests could potentially play a role in predicting important lawyer outcomes that are not adequately accounted for by academic attributes alone, and hence worth considering as part of a suite of law school selection tests (Holmquist et al., 2014).

In contrast to research in medical school and law school selection, personnel selection has a long history of using personality tests as part of the selection procedure. Its use is supported by numerous meta-analytic findings that conscientiousness, a domain of the Big Five personality framework, consistently predicts job performance (e.g., Judge et al., 2013). Contrary to common belief though, the relationship between personality and job performance may not be linear but curvilinear (Le et al., 2011). This finding suggests that very high scores on personality tests are not necessarily optimal for higher performance, which has implications for how personality scores are used for selection. Another consideration is the potential for adverse impact. Though there are only modest variations of personality across racial groups (Hough et al., 2001), there can still be adverse impact depending on how personality scores are used (Risavy & Hausdorf, 2011), which warrants further research.

**Integrity tests.** Integrity tests are used more often for personnel selection than for selection into medical and law schools. Integrity tests measure applicants' levels of honesty, which can be considered a hybrid of personality factors including conscientiousness, agreeableness, and adjustment (Ones & Viswesvaran, 1998). There are two types of integrity tests: overt tests and personality-based tests (Wanek et al., 2003). Overt tests explicitly examine test-takers' levels of honesty, their attitudes about theft, and their past cases of theft. Personality-based tests are a less explicit assessment of antecedent qualities associated with dishonest behaviors, assessing qualities such as dependability, trouble with authority, and hostility. A meta-analysis indicated that overt integrity tests are more closely associated with job performance than personality-based integrity tests (Van Iddekinge et al., 2012). The focus of research on integrity tests has primarily been on their relationship with counterproductive work behaviors (negative behaviors at work, such as theft and lying), with results showing moderate correlations with job performance, training performance,



and turnover. A review on integrity tests found that applicants seem to report reasonably positive reactions to the test (Berry et al., 2007). Thus, integrity tests seem to be particularly useful for personnel selection.

**Situational judgment tests.** Situational judgment tests (SJTs) are a scenario-based method used to measure a range of attributes, especially non-cognitive (non-academic) attributes such as interpersonal skills, contextualized judgment, and other work-related ‘soft skills’. SJTs present a series of context-related scenarios and ask applicants for their responses to the scenarios (Motowidlo et al., 1990). There are a range of design possibilities, including in the response instructions (e.g., asking what the applicants *would do* or *should do*), and response format (e.g., multiple-choice options or Likert-scale options; e.g., de Leng et al., 2018). Increasingly, digitalized and gamified versions are becoming more available (Gkorezis et al., 2021) with the increasing uptake of digital selection procedures (see Woods et al., 2020 for a review).

In medicine, SJTs are widely used for selection purposes with substantial evidence to support their use (Webster et al., 2020). SJTs assessing the ability to regulate emotions have been shown to predict academic performance in medical training, above and beyond cognitive ability and conscientiousness (Libbrecht et al., 2014). Similarly, an SJT assessing interpersonal skills was shown to predict academic and clinical performance in medicine, above and beyond cognitive ability (Lievens, 2013). Both text-based and video-based SJT formats have been trialed in medicine and video-based formats have demonstrated higher predictive validity (Lievens & Sackett, 2006) and generally received more favorable evaluations (Kanning et al., 2006).

Limited research has been dedicated to exploring how SJTs can be used for law school selection. One study, however, found that an SJT measuring non-cognitive qualities associated with lawyer effectiveness correlated with 23 of 26 lawyer effectiveness criterion outcomes, with correlations ranging from 0.11 to 0.21 (Shultz & Zedeck, 2011). Here, SJT scores were associated with more outcomes than LSAT scores and the size of the correlations appeared to be bigger, suggesting that research in this area may be worth conducting.

Large organizations have used SJTs for personnel selection for decades, using the method to assess a wide range of constructs, including interpersonal skills, leadership, personality, and heterogeneous composites of multiple constructs. A meta-analysis reported that the associations between SJTs and job performance ranged from 0.19 (for job knowledge and skills) to as high as 0.43 (for personality composites; Christian et al., 2010). In line with medical school selection research findings, video-based SJTs demonstrated stronger criterion-related validity than text versions (Christian et al., 2010). Given the multidimensional nature of SJTs, their internal consistency is often low to moderate (Lievens et al., 2008). However, there are minimal group differences (including gender and race) on SJT performance (Whetzel et al., 2008). More information on SJTs, especially pertaining to teacher selection, can be found in Chaps. 7 and 9.

**Interviews.** Interviews are face-to-face interactions between one or more interviewers with one or more interviewees. There are two broad types of interview formats: structured and unstructured. Structured interviews consist of pre-determined

questions with set scoring keys that are consistent across applicants. Unstructured interviews have no set format: the questions can differ between applicants and the scoring of the applicants are often overall impressions with no set scoring key. Panel interviews are an interview format with multiple interviewers and, although they are often perceived to have higher reliability and validity than individual interviews, evidence for this seems to be inconclusive (Dixon et al., 2002).

Interviews have been long used for medical school selection. The interviewers for medical school selection may be faculty members of the medical school and/or community members who are given interview training by medical school selection staff. Despite the popular use of interviews, especially in the final stage of medical school selection, past research shows that the reliability of structured interviews used for medical school selection can be low to moderate (Kreiter et al., 2004). This finding is surprising because interviews have high face validity and are widely used across numerous medical schools globally.

Interviews are infrequently used in law school selection, but they are one of the most frequently used methods for personnel selection (Anderson & Witvliet, 2008). Structured interviews show higher validity and reliability than unstructured interviews (Posthuma et al., 2002). There are only small inter-group differences when using structured interviews, with particularly low differences for interviews focusing on behavior (Moscoso, 2000). Applicants perceive interviews as a fairer method of selection than other methods (Hausknecht et al., 2004), with applicants typically expecting that interviews will take place as part of the selection process (Lievens et al., 2003). From the employers' perspective, interviews are seen as a chance to assess the applicants' social and communication skills.

**Multiple mini-interviews (MMIs).** A specific type of interview, multiple mini-interviews (MMIs; see Chap. 8 for discussion on applications in teacher selection), are increasingly being adopted to replace traditional interviews for medical selection (Patterson et al., 2016). The underlying assumption of MMIs is that a greater sampling of behaviors provides more information about the suitability of applicants and increases the reliability of the interview process. In MMIs, applicants rotate through 5 to 12 stations for brief interviews (typically 5–10 min) and to complete various tasks (Gafni et al., 2012). These stations differ in the type of interview administered. For example, one station may require applicants to describe what they would do in certain situations (situational judgment stations), another station might require applicants to describe what they did in the past in a particular situation (behavioral interview station), and another might ask the applicant to engage in a conversation with an actor playing the role of a patient. Of these types, behavioral interview stations seem to best differentiate among applicants (Eva & Macala, 2014).

Investigating MMI's overall validity in medical school selection, Eva et al. (2012) found that those who were rejected at an institution because of low MMI scores also received lower scores on the Canadian national licensing examinations than those who were not rejected given their MMI scores. Furthermore, the magnitude and the direction of MMI's concurrent validity with a national medical school aptitude test seemed to differ depending on the content of the aptitude test. Specifically, there was a small positive correlation with a section on reasoning in humanities and

social sciences (0.26) and a small negative correlation with a section on reasoning in biological and physical sciences ( $-0.15$ ), suggesting a multi-faceted relationship with different aptitudes (Roberts et al., 2008).

MMIs seem to be moderately to highly reliable (Rahim & Yusoff, 2016) and are favorably perceived by both applicants and examiners (Eva et al., 2004). Nevertheless, sufficient interviewer training and consistent scoring scheme is necessary given the inevitable subjective nature of interviews in general. All in all, MMIs appear to be a promising interview format for medical school selection with the potential for application in other fields, including teacher education.

**Assessment centers.** Assessment centers (ACs) are a suite of individual and group exercises that aim to assess a variety of applicant attributes that relate both directly and indirectly to a particular training or job opportunity. ACs can consist of group exercises, written/in-tray tasks, oral presentations, and interactive exercises. The AC format has not been widely explored for entrance into law school or medical education; however, some similar formats have been used for selection into postgraduate medical training (e.g., Randall, Davies et al., 2006; Randall, Stewart et al., 2006). Further evidence is needed before the use of ACs could be widely endorsed for medical school selection (Patterson et al., 2016).

Since their introduction over 50 years ago, ACs have been extensively used for personnel selection given their high face validity (Sackett & Lievens, 2008). A meta-analysis reported that ACs most commonly assess six dimensions of cognitive and non-cognitive attributes: *consideration of others*, *communication skills*, *motivation*, *persuasive power*, *organization and planning skills*, and *problem-solving skills* (Arthur et al., 2003). One of the barriers to using ACs is the high cost associated with administering this type of selection method. Job analyses are typically conducted to identify important dimensions of jobs, followed by training of multiple assessors who would rate applicants' performance in each of the identified dimensions (International Taskforce on Assessment Center Guidelines, 2015). As with interviews, highly trained assessors are needed to ensure that the selection method is useful in making predictions about future behaviors and outcomes. ACs tend to have weak measurement properties, especially in relation to the reliability of the scoring (Jackson et al., 2016). Moreover, in Schmidt and Hunter's (1998) meta-analysis, AC methods did not demonstrate significant incremental validity when combined with GMA in predicting overall job performance (a gain of 0.02 points). Furthermore, a meta-analysis reported that the standardized differences between ethnic subgroups ( $d$ ) can be large, with White applicants scoring higher than Black applicants ( $d = 0.52$ ) and Hispanic applicants ( $d = 0.28$ ; Dean et al., 2008). Thus, more research to ensure the rigorously and fairness of the method and its scoring may be necessary.

**Reference checks.** References provided by applicants typically include letters (or contact details) from individuals who know the applicant from a personal or professional context. These are used to verify the character (e.g., teamwork skills, organization skill) and work history of the applicants.

Reference checks are widely used in medical school selection (Kuncel et al., 2014). However, evidence of their predictive validity is mixed. Some researchers have found evidence supporting their predictive validity (e.g., DeZee et al., 2014)

but other researchers found inconsistent or no evidence in predicting medical school performance (e.g., Poole et al., 2009). However, it difficult to differentiate applicants using references. Most references provided are exclusively or mostly positive, with any negative indicators given ‘in code’—a referee may have sent a ‘hidden message’ but is missed by the assessor, which raises questions as to the appropriateness of including reference checks as part of medical school selection procedure (Stedman et al., 2009).

There is scant research examining how law schools use references as part of their selection system. We can, however, draw from the literature on how references are used for college and graduate admission generally. A meta-analysis on the relationship between references and undergraduate, graduate, and medical school performance found that there are modest correlations and little incremental validity over traditional academic predictors (Kuncel et al., 2014).

Reference checks add very little to predicting overall job performance when combined with GMA (Schmidt & Hunter, 1998). Their low predictive validity may be the result of uncertainty about what content should be assessed and how the content, typically positive, can be used to differentiate between applicants. Furthermore, low inter-rater reliability (around 0.40) means that there is a problem in the trustworthiness of the data from reference letters (Kuncel et al., 2014). This issue is in addition to limited research on adverse impact for particular groups and limited evidence supporting the use of reference letters except for their face validity. Thus, if references are used for selection, they may be useful for cross-checking or confirming applicant details at the final stages of selection rather than as a criterion for progression to a subsequent stage of selection.

### 5.3 What Can We Learn from Selection in Other Fields?

We have reviewed the common methods used in medical school selection, law school selection, and personnel selection. What can we learn from these fields that we can apply to teacher selection?

**1. Use multiple selection methods.** There is no magical selection method that can do everything. That is, one selection method should not be used as the only method to choose individuals into teacher training or employment. Rather, a carefully designed selection procedure consisting of multiple methods is best. Using multiple selection methods can: (a) allow assessments of multiple important constructs, (b) allow certain selection methods to ‘select in’ or ‘select out’ applicants at various stage of the selection procedure, (c) reduce costs as certain selection methods can be used to mass-screen a large pool of applicants so that a smaller pool of applicants can be invited to undergo more intensive (and expensive) selection methods, and (d) increase predictive validity by using multiple predictors rather than a single predictor.

**2. Use evidence-informed selection methods.** It is easy to include selection methods that have been used in the field for a long time and that are easy to administer. However, as we have reviewed in this chapter, there is varying strength of

evidence for each selection method. It is best to use selection methods that have the strongest evidence behind them (see Chap. 4 for issues to consider in decision-making). Examining the evidence that is available within and outside of the target field helps develop a strong theoretical and empirical rationale for the use of particular selection methods.

**3. Distinguish between selection method and selection constructs.** Before a selection method is chosen, consideration should be given to decisions about which constructs the selection process is targeting for assessment. After the constructs of interest are identified, one should assess which methods can best assess these constructs. For example, if communication skills are the target construct, one should consider a variety of methods that can assess this construct, such as MMIs and SJTs, and compare the evidence behind each of the methods. After choosing the selection method(s), one should make sure that the construct is indeed featured as a criterion in the procedure. For example, if communication skills are the construct of interest and interviews were chosen as a method to assess this, communication skills need to be explicitly featured in the structure of the interviews and the scoring criteria.

**4. Include structure in interviews.** Interviews are one of the most popular selection methods, with considerable evidence supporting the use of structured interviews over unstructured interviews. Unfortunately, structured interviews are less often used in practice for personnel selection (Lievens & De Paepe, 2004). Some of the explanations given by human resources personnel chime with why unstructured interviews are more frequently used in teacher selection; for example, interviewers' desire to establish an informal contact with the interviewees and to have greater discretion over the interview questions. Furthermore, conducting highly structured interviews can be quite costly both in time and money as they require additional processes, such as a job-analysis to form the interview questions, formulation of scoring procedures, and interviewer training to ensure standard scoring procedures. In contrast to common concerns, there is considerable flexibility when creating structured interviews (Levashina et al., 2014). Although it may be easy for interviewers to feel that unstructured interviews give a greater 'feel' for the applicants, the evidence is clear: structured interviews have greater predictive validity evidence than unstructured interviews. In the future, MMIs may become a more common method of implementing structured interviews in teacher selection (see Chap. 8).

## 5.4 Chapter Summary

To improve teacher selection methods, it is important to examine research from multiple contexts, including from fields outside of education, where there are existing, and often stronger, research foundations. This chapter reviewed the selection methods used for entrance into medical schools, law schools, and for employment in organizations (personnel selection). We outlined that the research base for different selection methods varies widely—from those with little evidence (e.g., reference checks) to those with stronger evidence (e.g., SJTs and MMIs). The combination and weighting

of the methods included in a selection program should be carefully considered and scrutinized against the needs of the program or organization, and resources that are available. In the next chapter, we will turn specifically to the selection of prospective teachers and explore historical and current research and practices in this field.

## References

- AAMC (2021). *What's on the MCAT exam?* Retrieved from <https://students-residents.aamc.org/applying-medical-school/article/whats-mcat-exam/>
- Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment, 16*, 1–13.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–153.
- Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology, 60*, 271–301.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83–117.
- Cleland, J. A., French, F. H., Johnston, P. W., & Scottish Medical Careers Cohort Study Group. (2011). A mixed-methods study identifying and exploring medical students' views of the UKCAT. *Medical Teacher, 33*, 244–249.
- Dalessandro, S. P., Stilwell, L. A., & Reese, L. M. (2014). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 2007–2008 through 2013–2014 testing years (TR 14–02)*. Law School Admission Council Research Rep. No. TR-00–01. Retrieved from [http://www.lsac.org/docs/default-source/research-\(lsac-resources\)/tr-14-02.pdf](http://www.lsac.org/docs/default-source/research-(lsac-resources)/tr-14-02.pdf)
- Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93*(3), 685–691.
- de Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. P. N. (2018). Influence of response instructions and response format on applicant perceptions of a situational judgement test for medical school selection. *BMC Medical Education, 18*(1), 282.
- DeZee, K. J., Magee, C. D., Rickards, G., Artino, A. R., Jr., Gilliland, W. R., Dong, T., & Durning, S. J. (2014). What aspects of letters of recommendation predict performance in medical school? Findings from one institution. *Academic Medicine: Journal of the Association of American Medical Colleges, 89*, 1408–1415.
- Dhar, D., Perry, W. R. G., & Poole, P. (2012). Students' perceptions of the Undergraduate Medicine and Health Sciences Admissions Test (UMAT). *The New Zealand Medical Journal, 125*(1361), 29–36.
- Dixon, M., Wang, S., Calvin, J., Dineen, B., & Tomlinson, E. (2002). The Panel Interview: A Review of Empirical Research and Guidelines for Practice. *Public Personnel Management, 31*, 397–428.
- Eva, K. W., & Macala, C. (2014). Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education, 48*(6), 604–613.
- Eva, K. W., Reiter, H. I., Rosenfeld, J., Trinh, K., Wood, T. J., & Norman, G. R. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *Journal of the American Medical Association, 308*, 2233–2240.
- Eva, K. W., Rosenfeld, J., & Reiter, H. I. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education, 38*, 314–326.
- Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: Systematic review of the literature. *BMJ, 324*(7343), 952–957.

- Gafni, N., Moshinsky, A., Eisenberg, O., Zeigler, D., & Ziv, A. (2012). Reliability estimates: Behavioural stations and questionnaires in medical school admissions. *Medical Education, 46*, 277–288.
- Geiser, S., & Santelices, M. V. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. *UC Berkley: Center for Studies in Higher Education*. Retrieved from <https://escholarship.org/uc/item/7306z0zf>
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.
- Gkorezisi, P., Georgiou, K., Nikolaou, I., & Kyriazati, A. (2021). Gamified or traditional situational judgement test? A moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and Organizational Psychology, 30*, 240–250.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639–683.
- Holmquist, K., Shultz, M., Zedeck, S., & Oppenheimer, D. (2014). Measuring merit: The Shultz-Zedeck research on law school admissions. *Journal of Legal Education, 63*, 565–584.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1–2), 152–194.
- International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*(4), 1244–1273.
- Jackson, D. J. R., Michaelides, G., Dewberry, C., & Kim, Y.-J. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of Applied Psychology, 101*(7), 976–994.
- 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). Guilford Press.
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology, 98*(6), 875–925.
- Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine: Journal of the Association of American Medical Colleges, 80*(10), 910–917.
- Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view. *European Journal of Psychological Assessment, 22*(3), 168–176.
- Kidder, W. (2001). Does the LSAT mirror or magnify racial and ethnic differences in educational attainment?: A study of equally achieving “Elite” college students. *California Law Review, 89*(4), 1055–1124
- Klassen, R. M., Durksen, T. L., Kim, L. E., Patterson, F., Rowett, E., Warwick, J., Warwick, P., & Wolpert, M.-A. (2017). Developing a proof-of-concept selection test for entry into primary teacher education programs. *International Journal of Assessment Tools in Education, 4*, 96–114.
- Klassen, R. M., & Kim, L. E. (2019). Selecting teachers and prospective teachers: A meta-analysis. *Educational Research Review, 26*, 32–51.
- Kreiter, C. D., Yin, P., Solow, C., & Brennan, R. L. (2004). Investigating the reliability of the medical school admissions interview. *Advances in Health Sciences Education: Theory and Practice, 9*(2), 147–159.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*(1), 148–161.
- Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1), 101–107.

- Lang, J. W. B., Kersting, M., Hülshager, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested-factors model of cognitive abilities. *Personnel Psychology, 63*, 595–640.
- LaPiana, W. P. (2004). Merit and diversity: The origins of the law school admissions test. *Saint Louis University Law Journal, 48*(3), 955–990.
- Law School Admissions Council. (2021). *About the LSAT*. Retrieved from <http://www.lsac.org/jd/lsat/about-the-lsat>
- Law School Admission Council. (2019). *2017–2019 LSAT correlation study results*. [https://www.lsac.org/system/files/inline-files/lsat-correlation-results\\_combined\\_2017\\_2018\\_2019\\_0.pdf](https://www.lsac.org/system/files/inline-files/lsat-correlation-results_combined_2017_2018_2019_0.pdf)
- Le, H., Oh, I.-S., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology, 96*, 113–133.
- Lempert, R. O., Chambers, D. L., & Adams, T. K. (2000). Michigan’s minority graduates in practice: The river runs through law school. *Law & Social Inquiry, 25*(2), 395–505.
- Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*(1), 241–293.
- Libbrecht, N., Lievens, F., Carette, B., & Côté, S. (2014). Emotional intelligence predicts success in medical school. *Emotion, 14*(1), 64–73.
- Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education, 47*(2), 182–189.
- Lievens, F., De Corte, W., & Brysse, K. (2003). Applicant perceptions of selection procedures: The role of selection information, belief in tests, and comparative anxiety. *International Journal of Selection and Assessment, 11*(1), 67–77.
- Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior, 25*(1), 29–46.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*(4), 426–441.
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*(5), 1181–1188.
- Lievens, F., & Sackett, P. R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology, 102*(1), 43–66.
- Lievens, F., Sackett, P. R., & Zhang, C. (2021). Personnel selection: A longstanding story of impact at the individual, firm, and societal level. *European Journal of Work and Organizational Psychology, 30*, 444–455.
- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: “sinking shafts at a few critical points.” *Annual Review of Psychology, 51*(1), 405–444.
- McLarnon, M. J. W., Rothstein, M. G., Goffin, R. D., Rieder, M. J., Poole, A., Krajewski, H. T., & Mestdagh, T. (2017). How important is personality in the selection of medical school students? *Personality and Individual Differences, 104*, 442–447.
- McManus, I. C., Dewberry, C., Nicholson, S., Dowell, J. S., Woolf, K., & Potts, H. W. W. (2013). Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies. *BMC Medicine, 11*(1), 1–21.
- McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. R. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ, 327*, 139–142.
- Moscato, S. (2000). Selection interview: A review of validity evidence, adverse impact and applicant reactions. *International Journal of Selection and Assessment, 8*(4), 237–247.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*, 640–647.



- Ones, D.S. and Viswesvaran, C. (1998) Integrity testing in organizations. In R.W. Griffin, A. O'Leary-Kelly and J.M. Collins (eds) *Dysfunctional Behavior in Organizations: Violent and Deviant Behavior*. JAI Press.
- Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive Abilities. In N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 1–92). Oxford University Press.
- 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*(1), 36–60.
- Poole, P. J., Moriarty, H. J., Wearn, A. M., Wilkinson, T. J., & Weller, J. M. (2009). Medical student selection in New Zealand: looking to the future. *The New Zealand Medical Journal, 122*(1306), 88–100.
- Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*, 1–81.
- Puddey, I. B., & Mercer, A. (2014). Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education, 14*(31), 1–12.
- Rahim, A. F. A., & Yusoff, M. S. B. (2016). Validity evidence of a Multiple Mini Interview. *Education in Medicine Journal, 8*(2), 49–63.
- Randall, R., Davies, H., Patterson, F., & Farrell, K. (2006). Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Archives of Disease in Childhood, 91*(5), 444–448.
- Randall, R., Stewart, P., Farrell, K., & Patterson, F. (2006). Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *The Obstetrician & Gynaecologist, 8*(4), 257–262.
- Risavy, S. D., & Hausdorf, P. A. (2011). Personality testing in personnel selection: Adverse impact and differential hiring rates. *International Journal of Selection and Assessment, 19*(1), 18–30.
- Roberts, C., Walton, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K., & Tiller, D. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education, 42*(4), 396–404.
- Roth, P. L., BeVier, C. A., Switzer, F. S., III., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*(5), 548–556.
- Roth, P. L., & Bobko, P. (2000). College grade point average as a personnel selection device: Ethnic group differences and potential adverse impact. *Journal of Applied Psychology, 85*(3), 399–406.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419–450.
- Sartania, N., McClure, J. D., Sweeting, H., & Browitt, A. (2014). Predictive power of UKCAT and other pre-admission measures for performance in a medical school in Glasgow: A cohort study. *BMC Medical Education, 14*(116), 1–10.
- 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*(2), 262–274.
- Shultz, M. M., & Zedeck, S. (2011). Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry: Journal of the American Bar Foundation, 36*(3), 620–661.
- Shultz, M. M., & Zedeck, S. (2012). Admission to law school: New measures. *Educational Psychologist, 47*(1), 51–65.
- Stedman, J. M., Hatch, J. P., & Schoenfeld, L. S. (2009). Letters of recommendation for the predoctoral internship in medical schools and other settings: Do they enhance decision making in the selection process? *Journal of Clinical Psychology in Medical Settings, 16*(4), 339–345.
- Stevens, L., Kelly, M. E., Hennessy, M., Last, J., Dunne, F., & O'Flynn, S. (2014). Medical students views on selection tools for Medical School—a mixed methods study. *Irish Medical Journal, 107*(8), 229–231.
- UCAT (2021). *UCAT test format*. Retrieved from: <https://www.ucat.ac.uk/ucat/test-format/>

- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology, 97*(3), 499–530.
- Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology, 56*, 873–894.
- Webster, E. S., Paton, L. W., Crampton, P. E. S., & Tiffin, P. A. (2020). Situational judgement test validity for selection: A systematic review and meta-analysis. *Medical Education, 54*(10), 888–902.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: a meta-analysis. *Human Performance, 21*(3), 291–309.
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology, 29*, 64–77.
- Yates, J., & James, D. (2010). The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: A prospective cohort study at Nottingham Medical School. *BMC Medical Education, 10*(55), 1–9.