2018 Ottawa Consensus Statement: Selection and Recruitment in the Healthcare Professions

Journal:	*Medical Teacher*
Manuscript ID	Draft
Manuscript Categories:	Articles
Date Submitted by the Author:	n/a
Complete List of Authors:	Patterson, Fiona; Work Psychology Group, ; University of Cambridge, Psychology Department Roberts, Chris; The University of Sydney, Northern Clinical School Hanson, Mark Hampe, Wolfgang; University of Hamburg Ponnamperuma, Gominda; University of Colombo, Medical Education Development and Research Centre; University of Dundee, Centre for Medical Education Eva, Kevin; University of British Columbia, Department of Medicine Magzoub, Mohi-Eldin; King Saud bin Abdulaziz University of Health Sciences, Medical Education Tekian, Ara; University of Illinois at Chicago, Medical Education Cleland, Jennifer; University of Aberdeen, Division of Medical and Dental Education
Keywords:	Selection < Management, Medical education research < Management, Assessment

2018 Ottawa Consensus Statement: Selection and Recruitment to the Healthcare Professions

Authorship. Patterson F, Roberts C, Hanson MD, Hampe W, Eva K, Ponnamperuma G, Magzoub M, Tekian A & Cleland J.

Corresponding author. Fiona Patterson. F.patterson@workpsychologygroup.com

Acknowledgements. Grateful thanks go to Tom Kinirons and Sarah Stott of Work Psychology Group for supporting the consensus group discussions and workshops, and in preparing the final manuscript.

Declaration of interests:

FP is a Director in Work Psychology Group, a research consulting practice which receives funding from Health Education England, UKCAT and other educational institutions to design and evaluate selection methods and systems.

JC is a member of the UKCAT Research Committee.

No other competing interests declared.

2018 Ottawa Consensus Statement: Selection and Recruitment to the Healthcare Professions

Introduction

Selection and recruitment into healthcare education and practice is a major area of endeavor for health professional educators. Key insights have emerged to guide further inquiry and influence future policy and practice since the last Ottawa consensus statement on selection (Prideaux et al., 2011). However, much remains to be done, not only in refining the science of selection, but also in considering the opportunities and challenges of translating current research understandings into established practice.

This updated consensus statement is based on strong theoretical research evidence and was developed using a multi-stage process. First, a group of international researchers with expertise in selection and recruitment, assessment, curriculum, and educational theory was specially convened to critically appraise the literature and develop a draft consensus statement. Second, the group shared the draft statement with the wider community via a workshop and symposium at the 2018 Ottawa-ICME Joint Conference on the Assessment of Competence in Medicine and the Healthcare Professions, and at the Second International Conference of Selection into the Health Professions. Colleagues from diverse countries and regions actively engaged with these activities face-to-face and through social media. Finally, the statement was refined on the basis of feedback, questions and comments, sent for independent peer review, and further revised.

Reflecting the state of the science at the time, the previous consensus (Prideaux et al., 2011) characterised selection as an assessment system and focussed on the quality of different selection methods. In this updated statement, we present the latest research findings on selection methods and, reflecting progression within the field, we also consider potentially more complex issues including: selection policies; methodological concerns (beyond psychometric issues); social accountability, diversity and fairness, workforce shortages in some specialities (e.g. General Practice and Psychiatry) and in certain contexts (e.g. remote and rural working, and emerging countries), globalisation issues, developments in theory, and evaluation frameworks.

We conclude with a synthesis of the key issues and a series of recommendations to guide future research and practice and encourage debate between colleagues across the globe. Our findings are

summarised into four areas as follows: (1) Selection philosophy and policies, (2) The effectiveness of different selection methods, (3) Diversity and globalisation issues, and (4) Theory and evaluation.

(1) Selection Philosophy and Policies

Individual education and training institutions define their selection philosophy and uniquely enact a selection policy (or policies) within the context of their national or state-based regulations, their own history, mission and goals, and stakeholder organisations. They choose and implement a combination of specific methods from the range available to them. Globally, guidance has been provided by health professional education policies and initiatives that have emphasized incorporation of a broader range of selection approaches ("The Edinburgh Declaration," 1988; Eva, 2018). Locally, institutions (e.g. The Association of American Medical Colleges (AAMC) Holistic Admissions Review (Colleges, 2018), the future of medical education in Canada (FMEC MD) (Canada, 2010) have developed versions of core competencies for their entry level applicants. However, despite an acknowledged need, there is little empirical research exploring the opportunities and challenges of developing, enacting, and evaluating selection at the policy level. As examples of this gap in the research, there has been a longstanding policy directive of the World Federation of Medical Education to ensure admission policies match the numbers of students trained with national needs for doctors ("The Edinburgh Declaration", 1988). Yet, a recent synopsis suggested that whilst much needed empirical evidence is awaited, selection policies are not taking account of the best available health outcomes and desirable career uptake data (Gorman, 2018). In the UK, there has been a lack of local enactment of national policy to widening participation in medical school selection, leading to a call for interdisciplinary working with those involved in health policy planning (for example, economists), as a way forward to develop more coherent and measurable policy in selection (Cleland, Kelly, Moffatt, & Nicholson, 2015). Some suggest that a major barrier to policy enactment is various stakeholders having a political interest in selection (e.g., government, the regulators and trade unions). Rather than remaining 'secretive' and immune to policy (Tekian, 1998), interdisciplinary working across stakeholder groups could establish the political validity of a more transparent and fair selection policy (Patterson, Lievens, Kerrin, Zibarras, & Carette, 2012).

There are at least three approaches to enacting selection policies that involve; individually-focused processes, competency-based frameworks, and a third approach which seeks to encompass factors such as social accountability, widening access/diversity, and workforce planning (Roberts, Khanna, & Rigby, 2017). In *individually-focused* processes, the capacity for academic success is typically the basis for

selection in the majority of medical school and specialty selection systems globally. The focus on academic success is a sequelae of the Flexner Report (Finnerty et al., 2010; Flexner, Pritchet, & Henry, 1910) and a precursor for the United States Medical College Admission Test (MCAT). This aptitude test was implemented to assess applicants' "*academic readiness*" for medical education (McGaghie, 2002). MCAT prioritizes the merit (defined most commonly as academic achievement) of individual applicants over broader social factors (Razack, Hodges, Steinert, & Maguire, 2015). Recently a range of standardized testing methods have been developed and tailored to local geographic definitions of "academic readiness" (e.g. UK Clinical Aptitude Test; UKCAT, and the Graduate Australian Medical School Admissions Test; GAMSAT). In specialty training settings, for example residency training in the US, rather than developing a specific test, the academic focus in selection uses national licensing scores such as the USMLE (Prober, Kolars, First, & Melnick, 2016), even though this assessment is not designed for this purpose. In medical school admissions, the mainstay of selection internationally continues to rely on academic qualifications which again, are not designed specifically for selection purposes.

In order to broaden selection to encompass personal attributes, *competency-based frameworks* (Mahon, Henderson, & Kirch, 2013; Ferguson, James, & Madeley, 2002) are built on a taxonomy of desirable behaviours and attitudes that are thought to indicate success as a healthcare practitioner or student, (Patterson et al., 2016b) and bear some resemblance to the tenets of competency-based medical education (Frank & Danoff, 2007). Entry-level selection criteria are often derived through multi-source, multi-method job analysis studies. For example, in the UK, the principles of job analysis have been used to develop selection criteria and methods for a range of specialties by identifying core and specialty specific competencies including those thought important for training in general practice (family medicine) (Patterson, Ferguson, & Thomas, 2008; Patterson et al., 2013). The subsequent multi-method selection processes may include structured interviews/multiple mini-interviews (MMI) and situational judgment testing (SJT) in addition to metrics of academic success.

This confluence of initiatives which adapt varying but congruent frameworks of personal attributes has seen an international shift towards the assessment of both applicants' "*academic readiness*" and their "*personal readiness*" for medical education (Kirch, 2012; Niessen & Meijer, 2017). As we discuss later, both types of measure are socio-culturally bound.

The third selection philosophy assumes some level of academic capability but attempts to address a dynamic interplay between individual personal competencies and collective societal expectations, appearing globally under a range of designations including *student diversity, social accountability,*

workforce planning and widening participation. Efforts to increase participation of previously under-represented student populations in medical and health professional education are thought to mitigate the impact of institutionally or societally determined barriers to their achieving the same academic success. Policies and initiatives with the potential to impact this third selection philosophy include; US federal disability legislation to increase societal participation and, by extension, participation in medical education of individuals with disabilities (DeLisa & Lindenthal, 2016); in Australia, much research and many government programs have focused on ways to increase the medical workforce supply, recruitment, and retention (Humphreys et al., 2008) to address physician maldistribution between urban, regional and rural areas. Other approaches include locating medical schools in remote and under-serviced areas where students are selected from these localities such as in Norway, Saudi Arabia and Sudan. Further policy, practice and research regarding patient engagement and co-development of selection processes will contribute to and advance this new and burgeoning selection philosophy (Hanson et al., 2018).

Our synthesis of the literature leads us to conclude that there is both a gap in the literature around, and a compelling case for, further theoretical and empirical literature to underpin the development of overall selection philosophies and policies and their enactment (Patterson et al., 2016a; Roberts et al., 2017; Patterson, Cleland, & Cousans, 2017).

(2) How Effective are Selection Methods?

The initial consensus statement (Prideaux et al., 2011) largely focused on what was known about selection methods at the time, mostly based on reviewing evidence for undergraduate selection. Drawing upon recent systematic reviews of the research literature (Patterson et al. 2016a; Roberts et al. 2017), there are now clearer messages about the comparative reliability, validity and cost-effectiveness of various selection methods, with more evidence emerging for postgraduate selection practices than previously. The research evidence for eight different selection methods is summarised in Table 1.

*****INSERT TABLE 1 HERE*****

As in the previous statement, prior academic attainment of candidates remains a universal feature of selection policies internationally and the strength of evidence for this remains strong. Despite ongoing concerns about comparability across institutions and programs, and concerns about grade inflation,

prior grades remain a robust predictor of future educational achievements (Ferguson et al., 2002; Siu & Reiter, 2009).

The picture remains less clear for aptitude tests mainly due to the large number of different aptitude tests available, ranging from tests of general cognitive ability (e.g. United Kingdom Clinical Aptitude; UKCAT) versus those that also comprise tests of scientific knowledge (e.g. Undergraduate Medical and Health Sciences Aptitude Test; UMAT) versus hybrids of the two types (e.g., the Medical College Admissions Test MCAT 2015). Performance on aptitude tests has been compared to mainly midpoint or endpoint training outcome measures, such as exam performance early in the curriculum (McManus, Dewberry, Nicholson, & Dowell, 2013) and at the end of medical school (Tiffin et al., 2016). There is equivocal evidence that aptitude tests may or may not favour certain types of candidates, which may have implications for fairness and widening access to medicine (Tiffin, Dowell, & McLachlan, 2012; Lievens, Patterson, Corstjens, Martin, & Nicholson, 2016; Kumar, Roberts, Bartle, & Eley, 2018).

Similarly, although some personality traits have been linked to in-training performance (e.g. Lievens et al., 2009) the evidence regarding the extent to which personality assessment is useful in selection is mixed, at best (see Patterson et al., 2016a; Roberts et al., 2017). Recent evidence also suggests that personality traits are no longer regarded as stable deterministic predictors of behaviour and are instead found to change across life span and in response to environmental contingencies (Ferguson & Lievens, 2017). This implies that approaches focusing on selecting out for specific traits might be of limited validity given the dynamic nature of traits and the context specificity of trait expression, and so further research is required to explore how personality assessment may be best operationalised in selection practices.

Since the previous consensus statement (Prideaux et al., 2011), more research has emerged demonstrating MMIs and SJTs to be more valid predictors of inter- and intra-personal attributes such as communication skills, ethical reasoning, empathy and integrity than personal statements, letters of recommendation or references (Patterson et al., 2016a). MMIs and SJTs are complementary and most commonly used at different stages of selection processes. SJTs' capacity to include many items can offer feasible measurement of a broader range of constructs and can be used for larger numbers of applicants as they can be computer-delivered and machine-marked. As such, SJTs are often used for the purpose of screening who is invited to an interview. On the other hand, MMIs offer a face-to-face encounter

allowing the perspectives of a preselected group of applicants to be probed in more depth. Face-to-face encounters may offer additional gains unrelated to the psychometric properties of the selection method: for example, they indicate institutional valuing of personal connections and visiting the medical school or specialty training setting which help applicants to assess the fit between themselves and the learning environment (Burgess, Roberts, Clark, & Mossman, 2014).

Although research is developing rapidly in this area, with many researchers seeking to identify precisely what is measured by both MMIs and SJTs, seeking universal evidence of construct validity may be misplaced as both SJTs and MMIs are measurement methodologies and the design parameters for each can vary significantly depending on the purpose of the recruiting institution (Knorr & Hissbach, 2014). Whilst there might be an available format whose components perfectly align with an institution's needs, the purpose and context might require a specific design (Reiter & Roberts 2018).

All these selection methods comprise a multitude of different design parameters, which are informed by the purpose of selection and the particular context. The labelling of differing methods indicates the categories of ways in which data can be collected to support selection rather than allowing definitive inferences to be drawn regarding how the methods were implemented. As a result, universal claims of validity and quality of any specific version of a method should be judged cautiously. For example, when evaluating MMIs and SJTs in differing contexts with differing formats of an interview or written test, each test must be evaluated individually before reaching conclusions about both its effectiveness (Reiter & Eva, 2017), and generalisability to other selection contexts. Although results from large-scale and meta-analytic studies can indicate the quality of different selection methods in general, local validation studies are required to determine the effectiveness of any given selection process. Validity also depends on the type of outcome measures against which performance on selection tests is compared, and the study phase in which the outcome is measured (Siu & Reiter, 2009).

A consistent problem in judging the quality of different selection methods relates the so-called '*criterion-problem' i*.e. what outcomes are we trying to predict? Predictive validity studies require a generally agreed-upon 'gold standard', but there is no single undisputed gold standard that measures the performance of a practicing health professional (Patterson et al., 2017; Roberts et al., 2017). Often the criteria used to measure performance in the student or job role do not directly relate to the criteria used for selection, which presents difficulties for data interpretation. Conversely, sometimes the criterion and predictor are very similar (e.g. using knowledge-based multiple-choice tests at selection to predict knowledge-based multiple-choice exam performance in medical school), which may lead to

common method variance complications. It is artificial to examine the influence of individual methods when they are used and weighted or sequenced in various ways across selection processes. There is little research into this aspect of selection although what evidence there is clearly identifies that diverse ways of weighting different methods will lead to different outcomes in terms of the population selected (e.g. Fernando, Cleland, & McKenzie, 2008; Tiffin et al., 2012; Griffin and Hu, 2015). Measurement can often report outcomes of a single aspect of the process (e.g., a new assessment format) and report evaluation outcomes at the level of the average reaction (c.f. Kirkpatrick Model) of the interviewers or applicants who are immediately impacted by the intervention (Roberts et al., 2017).

Outcome measures used to evaluate selection methods often focus on indicators of attainment and maximal performance (e.g. medical school achievements, performance in licensure exams) rather than indicators relating to clinical practice and typical (day-to-day) performance in the job role. Of the (few) longitudinal predictive validity studies available, additional clarity is needed to articulate the target outcome variables and their interpretation. As such, there is a need for more theory-driven, future-oriented research aimed at identifying what a 'competent' healthcare worker is at the various stages of education, training and practice to enable researchers to judge the quality of selection methods more accurately (see Patterson et al. 2016b for a detailed discussion). A further challenge for the selection community is the lack of agreement around taxonomies of desirable behaviours, for example based on the CanMEDS initiative for outcomes-based medical education (Frank and Danoff, 2007). There is thus a disconnect between selection, curriculum, in-program assessment and meaningful longitudinal evaluation.

From an international perspective, there is a breadth of selection methods currently available. Those that are of most value are likely to be context dependent, requiring the consideration of both an institution's philosophy regarding what they are trying to achieve, the communities it purports to serve, as well as the system within which the selection methods are to be used.

(3) Diversity and Globalisation Issues

People from certain groups, such as those from economically disadvantaged backgrounds, rural backgrounds or certain ethnic and cultural groups remain under-represented in medicine worldwide

(Bowes, Thomas, Peck & Nathwani, 2013). Razack et al.'s (2015) analysis of the multiple excellences that might need to be considered to enable inclusiveness in an exclusive process helps to demonstrate how power dynamics intervene between the professions and marginalized groups. This includes looking upstream at impediments put before some social groups that have not historically been noticed. For example, not so long ago, the UK admissions data indicated that there was a 600-fold difference between the most over-represented and most under-represented social classes in medical school (Hilton & Lewis 2004; Seyan, Greenhalgh, & Dorling, 2004). Regarding ethnicity, afro-Caribbean working-class males tend not to even apply to medicine in the UK (Kumwenda, Cleland, Greatrix, MacKenzie, & Prescott, 2017), and similarly, Turkish minority groups are significantly underrepresented in medical education in Germany. These trends have been stable over time (Mathers, Sitch, & Parry, 2016) and observed in almost all parts of the world (Young et al., 2012; Gale & Parker, 2013; Edwards, Maldonado, & Engelgau, 2000; Frazer, 2005; Crosby, Iyer, Clayton, & Downing, 2003).

In selection systems, traditionally dominated by measures of excellence, a lack of academic attainment can be driven by societal factors that preclude demonstrating sufficient success. The reproductive nature of medical education as elite is perpetually maintained. Barriers in demonstrating success for young people from less advantaged social groups include; routinely achieving lower levels of academic achievement throughout their schooling (Gorard et al., 2006), lack of self-confidence (Greenhalgh, Russell, Boynton, Lefford, Chopra, & Dunkley, 2006); not identifying with a medical career (Mathers & Parry, 2009; Southgate, Kelly, & Symonds, 2015); being unable to negotiate to information and resources in order to successfully apply to, and progress through, medical school (Nicholson & Cleland, 2017), being sensitive to the cost of medical education versus the need to earn money (Sianou-Kyrgiou & Tsiplakides, 2011); coming from a family without a history of university education (Robb, Dunkley, Boynton, & Greenhalgh, 2007; Southgate et al. 2015); and attending secondary (high) schools that do not motivate or support students to aspire to medicine (McHarg, Mattick, & Knight, 2007).

In many countries, major policy imperatives promote widening access into medicine, backed by governmental investment to increase the representation of certain groups (including those from lower socio-economic status, certain ethnic backgrounds, originating from remote and rural locations, and/or indigenous populations). The rationale for this is two-fold. First, to address societal issues of social justice and social mobility in terms of encouraging people from all backgrounds into higher education rather than birth dictating one's social and economic outcomes in life (Nicholson & Cleland, 2015). Second, a diverse healthcare workforce is considered essential to improving healthcare quality by

ensuring a supply of well-trained doctors who understand the context and culture of the communities they serve (Whitla et al., 2003; Saha, Guiton, Wimmers, & Wilkerson, 2008; Xu et al., 1997).

Paradoxically, efforts to move away from a reliance on academic achievement towards competency-based models can run the risk of amplifying these barriers as only the privileged can be expected to have opportunities to participate in many of the activities that admissions committee often treat as indicators of applicants' "personal readiness" (Wouters, 2018).

The approach to widening access varies widely by philosophy and context, raising a conundrum, between the individual rights of an applicant and the collective rights of the people they will serve. On one side, no applicant of equal ability should be discriminated against by virtue of which demographic group they come from (Woolf et al, 2011; Roberts et al., 2017). On the other side, the assumption that affirmative action policies such as quotas in classifying applicants on their demographic profile, will increase the probability of them practicing as a health professional in geographical areas with a similar demographic profile. There needs to be critical reflection and empirical evidence as to whether widening access on the basis of demographics can meet the purpose of providing a diverse workforce.

Some countries use quota-based systems to protect places for applicants from target groups which may be problematic in creating stigmas that under-represented applicants gained entry only because of the quota system. Quotas may work best in parallel with foundational or preparatory ("pipeline") programmes to help non-traditional applicants attain higher academic and personal readiness for medical school (Dalley, Podawiltz, & Castro, 2009; Cleland, Dowell, McLachlan, Nicholson, & Patterson, 2012; Southgate et al., 2015).

Other settings use contextual data in selection decision making, meaning that assessing an applicant's potential to succeed in higher education takes into consideration the context and circumstances in which their attainment to date has been achieved (Patterson & Price, 2017).

In Europe, France has little selection at entry into medical school but drastically reduces the number of students based on examinations early in the curriculum, which might reduce the impact of socio-economic status on selection. In Germany, the constitutional right for a free choice into medicine gave rise to a 20% quota of study places that are given to those applicants who have waited the longest from graduation from secondary school. However, the drop-out rates among these students are high and the constitutional court recently decided that this quota can be eliminated. By contrast in the US, state-wide

laws banning the consideration of ethnicity in selection pose serious obstacles for the medical profession to address inequities in care (Garces & Mickey-Pabello, 2015).

In the Netherlands, some perceive that candidates will use means available to them to 'strategically game' the selection system which may create a lack of diversity as it favours applicants of higher socio-economic status or social capital (Wouters, Croiset, Isik & Kusurkar, 2017; Stegers-Jager, 2018). Some argue that a national lottery may avoid these problems (Wouters et al., 2017; Hofstee, 1983). Others state however, that one cannot assume fairness if incoming academic success determines the number of lottery tickets one receives (Griffin & Hu, 2015). The secondary school system in the Netherlands is all publicly funded and tiered however, which means that only students from the top tier – academically – can apply for university, and during the secondary school period students who have entered in the lowest tier can move up. This is only partly influenced by SES. In any case, it must be recognized that there are costs associated with failing students in a selection system (or a student may 'luck out' in a lottery system, Foo et al., 2017). Further downstream in specialty training, there is a scarcity of data around diversity with the exception of evidencing independent predictors of rural medical practice, where there is a strong positive correlation between rural background and rural clinical school exposure (Farmer, Kenny, McKinstry, & Huysmans, 2015) suggesting there needs to be greater integration of selection with policy.

Overall, the evidence of effectiveness regarding various approaches to increasing diversity in medicine is equivocal. On a positive note, it is customary for newer selection tools to be tested for their neutrality toward different sections of the applicant population (Moreau, Reiter, & Eva, 2006; Lievens et al., 2016). The transition towards using more deliberate selection techniques appear to be yielding better student performance outcomes (Stegers-Jager, 2018). The movement towards broader competency-based models of selection appear to offset (to some degree) the deleterious effects of relying solely on academic achievement in some contexts e.g. in Australia (Griffin & Hu, 2015). Conditionalized admissions processes have been shown to successfully enable less prepared applicants from under-represented backgrounds to more readily integrate and succeed in medical school (Girotti, Park, & Tekian, 2015). Less positively, studying the outcomes of widening access initiatives and outreach has progressed little since publication of the original selection consensus statement. Technical solutions, which may address widening access by generating 'standardised admission ratios' (where the proportion admitted to study medicine from different social segments is matched with their population proportion: McManus, 2002; Seyan et al., 2004) or 'constrained optimisation' models (Kreiter, 2002) have not been

embraced by medical schools. On the other hand, there is promising results shown by conditional admissions and enrichment programs aimed toward promoting students from less advantaged groups who can receive the proper academic training, either prior to entry into medical school or early on in their medical school (Girotti et al, 2015). Through a 10-year study, Girroti and colleagues have shown comparable outcomes of students admitted through the conditional admissions and enrichment programs, with reasonable financial costs. Evidence such as these support calls for additional studies that can facilitate widening access for disadvantaged students who can practice medicine and contribute to healthcare of their communities.

There are two particular areas of selection where practice and policy are strongly influenced by market forces. In many locations, selection tools have been 'productized' and commercially marketed to applicants (e.g., fee-payable preparatory courses) and schools (e.g., "off-the-shelf" products/tests that can be included in the selection system). There is some research suggesting that coaching products do not necessarily improve application success (e.g. Griffin, Harding, Wilson, & Yeomans, 2008; see Griffin (2018) for a review). There is little research exploring the impact of inability to access commercially-driven selection practices on one's degree of aspiration towards medical school or specialty training, nor more generally, in the interaction between commercial product availability and efforts to increase the diversity of applicants. The ease with which institutions can require applicants (at their own cost) to complete commercially available selection products, may prove to be a greater determinant of their widespread uptake than indices of the product's psychometric or social responsibility qualities.

In a globalised economy, medical migration is an increasing issue. In many developing countries, such as in Africa and the Middle East, there are high rates of migration and intention to migrate (Bailey, Mandeville, Rhodes, Mipando, & Muula, 2012; Suciu, Popescu, Ciumageanu, & Buzoianu, 2017). These intentions may be transient (e.g. going abroad for specialty training and then returning to work in the country of origin; Tekian & Boulet, 2015) or permanent. Policy, regulations and legislation that enable and encourage active recruitment of overseas-trained doctors can propagate the 'brain drain' of health workers migrating from poor to wealthy countries (Smith, 2008). For example, Sub-saharan Africa has 24% of the burden (of global disease) but only 3% of health workers (World Health Organization Report, 2006). The question needs to be asked whether regulatory bodies have the agility to consider twenty-first century medical migration flows at a coordinated, global level (Hawthorne, 2015).

While such issues are not well represented in the health professional education literature on selection, we propose several research questions that need further investigation if we are to better understand

our practices and their influence on the international context. To what extent do applicants enter into medical school, or vocational training with the intention to migrate under current systems? In what ways might this be influenced – for example, do healthcare authorities view migration favourably (e.g., as a means of earning foreign exchange by 'exporting' healthcare workforce) or unfavourably (e.g., as a threat to meeting the healthcare workforce requirements of the home country)? Attempts have been made to establish a global code of ethics to guide the recruitment of healthcare professionals trained in one country by another country (Siyam, et al, 2013) although further research is required to establish how such initiatives can be best put into practice. In short, fundamental systems changes need to happen if we are to align input (medical students) with output (producing doctors to meet the public's health care needs). As a next step, we propose that much greater research attention be given to explaining why there are significant culturally acceptable differences in the way that different countries/contexts approach diversity and widening access issues.

(4) Theory and Evaluation

In appraising the rapidly increasing literature on selection into the health professions, the dominant agenda for those claiming to adopt "scientific methods" to inform the debates in the field has been psychometric (i.e. how best to achieve predictive validity and reliability in selection practices). This approach has focused on the reliability of particular single methods and assessing the linear relationship between students' performance on a specific selection method, and how they perform on later assessments using the same selection methods, as a medical student, resident/trainee, or as a fully-qualified doctor (see Patterson et al., 2016a). Newer psychometric developments such as machine learning, use a set of algorithms that are 'trained' on a data set to make predictions or take actions, such as classifying potential job applicants into good or bad prospects based on historical data (Mayer-Schonberger & Cukier, 2013). Big data approaches using learning analytics also bear promise in relating selection to in-training performance and potentially health outcomes data (Ellaway, Pusic, Galbraith, & Cameron, 2014). Given the interdisciplinary nature of such analysis we should expect more sophisticated predictive validity studies, applied to multi-site high quality data sets, that take differences in context and content of selection practices into account.

The issues with the dominance of the psychometric approach to selection research have been overlooked for too long and reflect the broader evolution of the field. We have made the case for the complex, multi-level, dynamic and social nature of selection. Selection does not occur in a

(psychometric) vacuum. In order to develop the appropriate theoretical frameworks and appropriate methodologies to investigate this, mixed methods approaches may be more appropriate. The philosophies, policies and outcomes of selection are not static, rather they shift on the basis of global and national, societal and political, drivers of medical education (Hodges & Lingard, 2012; Cooke, Irby, & O'Brien, 2010; Frenk et al., 2010). Broadly social forces underpinning selection may conflict as particular ways of defining excellence may lead to uncomfortable tensions for some medical schools and specialty selection boards that strive to increase the diversity of its student or trainee body. This reflects multiple and complex institutional demands and negotiations among actors and stakeholders including admissions chairs, institutional staff, applicants, patients, health providers, regulators, funders, universities and society, leading different health professional institutions to have different selection aims and objectives (Cleland et al., 2015; Alexander, Fahey Palma, Nicholson, & Cleland, 2017; Razack et al., 2015). There is a growing awareness in the health professional education literature that context matters, not only in the traditional sense of context specificity (poor relationships between performance on problems of different characteristics), but also in the macro sense of variability induced by differences in the environment in which educational interventions are implemented; Bates & Ellaway, 2016). These are all areas of much needed further research.

We urge those with interest in selection to shift their thinking, to embrace, rather than avoid, the exploration and consideration of forces that shape health professional selection. To do so requires more qualitative studies in the area to complement those focused on 'measurement'. Research that complements on-going longitudinal psychometric studies could help the field move towards addressing more complex questions such as the ways in which current selection practices may lead to unintended consequences in the makeup of the selected cohort. For example, the selection aims and objectives of a medical school in a country over-producing doctors with the understanding that many graduates will migrate (Bailey et al., 2012; Kizito et al., 2015; Suciu et al., 2017) are likely to be very different from those of a medical school in a country such as the UK where the number of medical schools are controlled by government on the basis of national workforce planning forecasts. We have emphasised the priority in extending research that informs policy to focusing on widening access/increasing diversity in medicine (e.g. Alexander et al., 2017; Razack et al., 2015; Cleland & Nicholson 2015; Greenhalgh, Seyan, & Boynton, 2004; Robb et al., 2007; Southgate et al., 2015).

A balance of qualitative and quantitative studies, ideally in the same programme of selection research, will help efforts to understand both measurement principles and social complexities, and to examine the tensions between efficiency and effectiveness in medical selection. There are many possible ways in which this could be undertaken, and we shall discuss three possibilities briefly – logic models, realist evaluation and evaluation that acknowledges complexity.

Dore and colleagues (Dore, Roberts, & Wright, 2017) suggest the use of program *logic models* to evaluate the process and effectiveness of selection processes. This approach highlights the critical components of a process that are often not explicitly stated and can enable linkages between studies with differing epistemological underpinnings, such as psychometric properties and social accountability. They suggest that logic models have the potential to illuminate the process of health care selection, to clarify what questions could and should be asked. Logic models have multiple components which are comprised of assumptions, biases, and expectations of how the inputs and activities (of, in this case, selection) will influence the outcome, output and impact (of a selection process; Cooksy, Gill, & Kelly, 2001).

Our second example is *realist evaluation*, defined as an iterative explanation-building process, that might draw from a suite of realist approaches based on the philosophy of scientific realism (Wong, Greenhalgh, Westhorp, & Pawson, 2012). This includes realist synthesis as a 'strategy for synthesising evidence and providing explanations (programme theories) on why interventions may, or may not, work (Kehoe et al., 2016). Originally proposed by Pawson and Tilley (1997), the principal concern of realist evaluation is to address, through a CMO (Context-Mechanism-Outcome) configuration, the question of 'what works, for whom, under what circumstances, and how' (Wong et al., 2016, p1; Dalkin, Greenhalgh, Jones, Cunningham, & Lhussier, 2015). Mechanisms can be viewed as the underlying entities, processes, or structures that operate in particular contexts to generate outcomes of interest (Dalkin et al., 2015). Realist evaluation could usefully be used to compare CMO configurations in different contexts, to provide insight into why one selection approach works in one context or country, and yet may not work in another.

Finally, Cleland, Patterson and Hanson (2018) propose that selection can be viewed as a "wicked problem" (Rittel & Webber, 1973) and offer a lens through which researchers can acknowledge the *complexity* of selection processes. It shifts thinking and action from seeking one elusive, objective truth, to recognizing that complexity is inevitable in health professional selection, managing uncertainty, and

questioning and considering "issues" associated with selection more productively. Labelling medical selection as "wicked" foregrounds the need for different responses than would be the case if it was simple, linear, or deterministic, and opens the door to shifting conceptualizations of selection. Similarly, Jorm and Roberts (2017) have proposed framing medical school evaluations within a complexity model and associated concepts. They suggest the tool of sensemaking to help medical educators apply this evaluation approach in their own settings. While not specific to selection, their metaphor of the medical school as a neuron situated within a complex neural network may enable medical educators to reframe the way they think about evaluating selection processes, and their outcomes.

Summary, Future Directions and Recommendations

Having reviewed the large body of research evidence it is clear the field of selection and recruitment has matured substantially since the previous Ottawa consensus statement (Prideaux et al., 2011).

Our understanding of the evidence surrounding selection methods has developed significantly. This is reflected in practice - many institutions are now using more 'evidence-based' approaches in their selection procedures in for example the UK, Netherlands, Germany, and Australia. However, many parts of the world continue to use selection methods that have little or no research evidence supporting them (e.g. personal statements, references) and/or rely on prior academic attainment as the main criterion. It would, thus, seem that more work is required to translate existing research findings into practice and to understand why such a gap persists.

We propose that to provide better quality practical guidance, the field must now move beyond single-site, single method analyses and acknowledge that selection is a complex, multi-level, dynamic phenomenon. In predictive validity studies linear regression-based analyses don't account for complexity – they assume a stable, positivistic, relationship between variables. Accordingly we must expand our methodological approaches. This includes exploring data access issues and considering the opportunities and challenges of using complex big data approaches to relate selection policies to health outcomes (Gorman, 2018). We have reviewed several key topics to guide future research by suggesting the field should (a) pay greater attention to the interaction between selection methods and selection philosophy and policy making and (b) use more sophisticated evaluation approaches and theoretical frameworks to better inform those involved in selection regarding how to best deal with issues such as

the weighting and sequencing of selection methods as well as addressing diversity and workforce shortages. Better quality evidence would also help educate those directly involved in selection to navigate the increasing presence of commercially-available selection tools.

Consistent with the previous consensus statement, diversity continues to be a critically important topic given a global mismatch between access to quality, culturally sensitive patient care. Further research in an international context, including cross-cultural comparisons, is required to identify which approaches are effective in terms of attracting, recruiting and supporting under-represented groups into medical school, through to specialty training and into practice. It is also important to explore differential attainment for different groups who are successful in getting into both medical school and specialty training programs. Additionally, we propose that much greater research attention be given to explaining why there are substantial differences in culturally acceptable ways of approaching diversity and widening access issues. Currently, the responsibility of selection lies mainly on educational institutions whereas recruitment is a responsibility shared by employers, government and the broader community. Diversity and widening participation is an area ripe for inter-disciplinary research, including opportunities for health professional educators to work closely with economists and those with expertise in health policy and planning, as well as being mindful of the local or regional legal implications of selection policies and practice (Cleland et al., 2015; Alexander et al., 2017; Razack et al., 2015; Nicholson & Cleland, 2015; Greenhalgh et al., 2004; Robb et al., 2007; Southgate et al., 2015).

Whilst the field of selection has moved on considerably since the original selection consensus statement was published, the majority of research and available material continues to originate from a limited number of global regions: northern Europe, the USA and Canada, and Australasia. Research and case studies from Asia, South America, the Middle East and Africa are lacking. Our view is that health professional education generally, and selection and widening access specifically, cannot be isolated from the cultural and social structural context in which it takes place (Stevens, 2009; Stevens & Simmonds Goulbourne, 2012). The world is variant, not homogeneous (Schuwirth & van der Vleuten, 2006). As such, we make a plea for resources for research, case studies and evaluations related to selection from contexts whose voices are currently under-represented.

Building upon the evidence and discussion from our consensus statement, we make ten key recommendations.

INSERT TABLE 2 HERE

References

Alexander K, Fahey Palma T, Nicholson S, Cleland J. 2017. "Why not you?" Discourses of widening access on UK medical school websites. Medical Education 51: 598-611.

Bailey N, Mandeville K, Rhodes T, Mipando M, Muula AS. 2012. Postgraduate career intentions of medical students and recent graduates in Malawi: a qualitative interview study. BMC Medical Education, 12(1): 87. doi:10.1186/1472-6920-12-87

Bates J, Ellaway RH. 2016. Mapping the dark matter of context: A conceptual scoping review. Med Educ 50(8):807-16.

Bowes L, Thomas L, Peck L, Nathwani T. 2013 International Research on the Effectiveness of Widening Participation. Report to HEFCE and OFFA by CFE and Edge Hill University. http://www.hefce.ac.uk/media/hefce/content/pubs/indirreports/2013/WP,international,research/2013_WPeffectiveness.pdf

Burgess A, Roberts C, Clark T, Mossman K. 2014. The social validity of a national assessment centre for selection into general practice training. BMC medical education. 14(1):261.

Canada, Association of Faculties of Medicine of. 2010. The future of medical education in Canada (FMEC): a collective vision for MD education. Association of Faculties of Medicine of Canada (AFMC).

Cleland J, Dowell J, McLachlan J, Nicholson S, Patterson F. 2012. Identifying best practice in the selection of medical students (literature review and interview survey). London: General Medical Council.

Cleland JA, Nicholson S, Kelly N, Moffat M. 2015. Taking context seriously: explaining widening access policy enactments in UK medical schools. Medical Education. 49:25-35.

Cleland JA, Patterson F, Hanson M. (under review). The wicked problem of selection, recruitment and widening access. Medical Education.

Colleges, American Association of Medical. 2018. AAMC holistic review project achieving improved learning and workforce outcomes through admissions.

Cooke M, Irby DM, O'Brien B. 2010. A summary of educating physicians: a call for reform of medical school and residency. John Wiley & Sons.

Cooksy LJ, Gill P and Kelly PA. 2001. The program logic model as an integrative framework for a multimethod evaluation. Evaluation and Program Planning, 24(2), 119–128.

Crosby FJ, Iyer A, Clayton S, Downing RA. 2003. Affirmative action. Psychological data and the policy debates. Academic Medicine. 58(2):93-115.

Dalkin SM, Greenhalgh J, Jones D, Cunningham B, Lhussier M. 2015. What's in a mechanism? Development of a key concept in realist evaluation. Implementation Science. 10(1):49.

Dalley B, Podawiltz A, Castro R et al. 2009.. The Joint Admissions Medical Program: a statewide approach to expanding medical education and career opportunities for disadvantaged students. Academic Medicine. 84(10):1373–1382.

DeLisa JA, Lindenthal JJ. 2016. Learning from physicians with disabilities and their patients. AMA Journal of Ethics. 18(10):1002.

Dore KL, Roberts C, Wright, S. 2017. Widening perspectives: reframing the way we research selection. Adv Health Sci Educ Theory Pract. 22(2):565-572. doi:10.1007/s10459-016-9730-5

The Edinburgh Declaration. 1988. Medical Education. 22(5):481-482. doi:10.1111/j.1365-2923.1988.tb00788.x

Edwards JC, Maldonado Jr FG, & Engelgau GR. 2000. Beyond affirmative action: one school's experiences with a race-neutral admission process. Academic Medicine. 75(8):806-815.

Ellaway RH, Pusic MV, Galbraith RM, Cameron T. 2014. Developing the role of big data and analytics in health professional education. Medical Teacher. 36(3):216-222.

Eva KW. 2018. From old town to new town: the state of the science 30 years after the Edinburgh declaration. Medical Education. 52(1):1-2. doi:10.1111/medu.13488

Farmer J, Kenny A, McKinstry C, Huysmans RD. 2015. A scoping review of the association between rural medical education and rural practice location. Human Resources for Health 13(1):27.

Ferguson E, James D, Madeley L. 2002. Factors associated with success in medical school: systematic review of the literature. BMJ. 324(7343):952-957. doi:10.1136/bmj.324.7343.952.

Ferguson E, Lievens F. 2017. Future directions in personality, occupational and medical selection: myths, misunderstandings, measurement, and suggestions. Advances in Health Science Education. 22:387–399

Fernando N, Cleland J, McKenzie H. 2008. Identifying the factors that determine feedback given to undergraduate medical students following formative mini-CEX assessments. Medical Education 42(1): 89-95.

Finnerty EP, Chauvin S, Bonaminio G, Andrews M, Carroll RG, Pangaro LN. 2010. Flexner revisited: the role and value of the basic sciences in medical education. Academic Medicine. 85(2):349-355.

Flexner A, Pritchet H, Henry S. 1910. Medical education in the United States and Canada bulletin number four (The Flexner Report). New York (NY): The Carnegie Foundation for the Advancement of Teaching.

Foo J, Rivers G, Ilic D, Evans DJR, Walsh K, Haines T, Paynter S, Morgan P, Lincke K, Lambrou H, Nethercote A, Maloney S. 2017. The economic cost of failure in clinical education: A multi-perspective analysis. Med Educ 51(7):740-754.

Frazer RA. 2005. Affirmative action policy in medical school admissions. Journal of Health Care Poor Underserved. 16(1):12-18.

Frank JR, Danoff D. 2007. The CanMEDS initiative: Implementing an outcomes-based framework of physician competencies. Med Teach. 29(7):642-

Frenk J, Chen L, Bhutta Z. A., Cohen, J., Crisp, N., Evans, T., Zurayk, H. 2010. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. The Lancet, 376(9756), 1923–1958.

Gale T, Parker S. 2013. Widening participation in Australian higher education. Report to the Higher Education Funding Council of England (HEFCE) and the Office for Fair Access (OFFA), England. Leicester, UK: CFE (Research and Consulting) Ltd.

Garces LM, Mickey-Pabello D. 2015. Racial diversity in the medical profession: The impact of affirmative action bans on underrepresented student of color matriculation in medical schools. The Journal of Higher Education. 86(2):264-94

Girotti J, Park YS, Tekian A. 2015. Ensuring a Fair and Equitable Selection of Students to Serve Society's Healthcare Needs. Medical Education. 49:84-92.

Gorard S, Smith E, May H, Thomas L, Adnett N, Slack K. 2006. Review of widening participation research: Addressing barriers to participation in higher education. Bristol: HEFCE.

Gorman D. 2018. Matching the production of doctors with national needs. Medical Education. Jan 1;52(1):103-13.

Greenhalgh T, Russell J, Boynton P, Lefford F, Chopra N, Dunkley L. 2006. "We were treated like adults"-development of a pre-medicine summer school for 16 year olds from deprived socioeconomic backgrounds: action research study. BMJ. 332(7544):762-7.

Greenhalgh T, Seyan K, Boynton PM. 2004. "Not a university type": focus group study of social class, ethnic, and sex differences in school pupils' perceptions about medical school. BMJ 328:1541-1544.

Griffin B. 2018. Coaching issues. In Patterson & Zibarras (eds). Selection and recruitment in the healthcare prfoessions. Palgrave: UK (in press).

Griffin B, Harding DW, Wilson IG, Yeomans ND. 2008. Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. Med J Aust 189(5): 270-3.

Griffin B, Hu W. 2015. The interaction of socio-economic status and gender in widening participation in medicine. Med Educ. 49(1):103-13.

Hanson MD, Pang C, McLeod J, Springall E, Kulasegaram K, Eva K. 2018. Engaging patiens in health professions education admissions: A scoping review of best practices, policies and research. Proceedings of the 18th Association for Medical Education (AMEE) conference; 2018 Aug 25-29; Basel, Switzerland.

Hawthorne, L. 2015. The impact of skilled migration on foreign qualification recognition reform in Australia. Canadian Public Policy, 41(1 Suppl):S173-S187. doi:10.3138/cpp.2015-027.

Hilton S, Lewis K. 2004. Opening doors to medicine - widening participation must sit alongside top up fees. British Medical Journal 328:1508-1509.

Hodges BD, Lingard L. 2012. The question of competence: reconsidering medical education in the twenty-first century (the culture and politics of health care work). In: Hodges L, Lingard BD, editors. Cornell University Press.

Hofstee WKB. 1983. The case for compromise in educational selection and grading. In: Anderson SB, Helmick JS, eds. On educational testing. San-Francisco: Josey-Bass. 109-27.

Humphreys JS, Wakerman J, Wells R, Kuipers P, Jones JA, Entwhistle P. 2008.“Beyond workforce”: A systematic solution for health service provision in small rural and remote communities. MJA 188(8):S77.

Jorm C, Roberts C. 2017. Using complexity theory to guide medical school evaluations. Academic Medicine. doi:10.1097/acm.0000000000001828.

Kehoe A, McLachlan J, Metcalf J, Forrest S, Carter M, Illing J. 2016. Supporting international medical graduates’ transition to their host-country: realist synthesis. Medical Education. 50(10):1015-1032. doi:10.1111/medu.13071.

Knorr M, Hissbach J. 2014. Multiple mini-interviews: Same concept, different approaches. Medical Education. 48(12):1157-75.

Kirch DG. 2012. Transforming admissions: the gateway to medicine. Jama, 308(21):2250-2251.

Kizito S, Mukunya D, Nakitende J, Nambasa S, Nampogo A, Kalyesubula R, . . . Sewankambo N. 2015. Career intentions of final year medical students in Uganda after graduating: the burden of brain drain. BMC Medical Education. 15(1):122. doi:10.1186/s12909-015-0396-0

Kreiter CD. 2002. The use of constrained optimization to facilitate admission decisions. Academic Medicine. 77:148–151.

Kumar K, Roberts C, Bartle E, Eley DS. 2018. Testing for medical school selection: what are prospective doctors' experiences and perceptions of the GAMSAT and what are the consequences of testing? Advances in Health Sciences Education. doi:10.1007/s10459-018-9811-8.

Kumwenda B, Cleland J, Greatrix R, MacKenzie R, Prescott G. 2017. Do graduates increase diversity of the UK medical student body? A national cohort study. Submitted for publication.

Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. 2016. Widening access in selection using situational judgement tests: evidence from the UKCAT. Medical Education. 50(6):624-36.

Mahon KE, Henderson MK, Kirch DG. 2013. Selecting tomorrow's physicians: the key to the future health care workforce. Academic Medicine. 88(12):1806-1811. doi:10.1097/acm.0000000000000023

Mathers J, Parry J. 2009. Why are there so few working-class applicants to medical schools? Learning from the success stories. Medical Education. 43:219–228.

Mathers J, Sitch A, & Parry J. (2016). Population-based longitudinal analyses of offer likelihood in UK medical schools: 1996–2012. Medical Education, 50, 612–623. doi: 10.1111/medu.12981

Mayer-Schonberger V, Cukier K. 2013. Big data: A revolution that will transform how we live, work and think. Boston, MA: Houghton Mifflin Harcourt.

McHarg J, Mattick K, Knight LV. 2007. Why people apply to medical school: implications for widening participation activities. Medical Education. 41:815–821.

McGaghie WC. 2002. Assessing readiness for medical education: evolution of the medical college admission test. Jama. 288(9):1085-1090.

McManus IC. 2002. Medical school applications - a critical situation. British Medical Journal. 325:786-787.

McManus I, Dewberry C, Nicholson S, Dowell J. The UKCAT-12 Study: Education attainment, aptitude test performance, demographoc and socioeconomic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools. BMC Med 11:244.

Moreau K, Reiter H, Eva KW. 2006. Comparison of aboriginal and nonaboriginal applicants for admissions on the Multiple Mini-Interview using aboriginal and nonaboriginal interviewers. Teaching and Learning in Medicine. 18(1):58-61.

Nicholson S, Cleland J. 2015. Reframing research on Widening Participation in Medical Education. In: Cleland J, Durning S, editors. Researching Medical Education. 1st ed. Oxford: Wiley-Blackwell; p.231-42.

Nicholson S, Cleland J. 2017. "It's making contacts": Notions of social capital and implications for widening access to medical education. Advances in Health Sciences Education. 22(2): 477-490.

Niessen ASM, Meijer RR. 2017. On the use of broadened admission criteria in higher education. Perspectives on Psychological Science. 12(3):436-448.

Patterson F, Cleland J, Cousans, F. 2017. Selection methods in healthcare professions: where are we now and where next? Advances in Health Sciences Education. 22(2):229-242. doi:10.1007/s10459-017-9752-7

Patterson F, Ferguson E, Thomas S. 2008. Using job analyses to identify core and specific competencies for three secondary care specialties: Implications for selection and recruitment. Medical Education 42: 1195-1204.

Patterson F, Knight A, Dowell J, Nicholson S, Cleland JA. 2016a. How effective are selection methods in medical education? A systematic review. Medical Education. 50: 36-60.

Patterson F, Lievens F, Kerrin M, Zibarras L, Carette B. 2012. Designing Selection Systems for Medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. International Journal of Selection and Assessment. 20(4):486-96.

Patterson F, Prescott-Clements L, Zibarras L, Edwards H, Kerrin M, Cousans F. 2016b. Recruiting for values in healthcare: a preliminary review of the evidence. Advances in Health Sciences Education. 21(4):859-881. doi:10.1007/s10459-014-9579-4

Patterson R, Price J. 2017. Widening participation in medicine: What, why and how? MedEdPublish. 6.

Patterson F, Tavabie A, Denney M, Kerrin M, Ashowrth V, Koczwara A, & MacLeod S. 2013. A new competency model for general practice: Implications for selection, training and careers. British Journal of General Practice 63(610):331-338.

Pawson R, Tilley N. 1997. Realistic evaluation. Sage.

Prideaux D, Roberts, C, Eva K, Centeno A, McCrorie, P, McManus C, Patterson F, Powis D, Tekian A, Wilkinson D. 2011. Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference. Medical Teacher 33(3):215-23.

Prober CG, Kolars JC, First LR, Melnick DE. 2016. A plea to reassess the role of United States Medical Licensing Examination Step 1 scores in residency selection. Academic Medicine. 91(1): 12-15. doi:10.1097/acm.0000000000000855

Razack S, Hodges B, Steinert Y, Maguire M. 2015. Seeking inclusion in an exclusive process: discourses of medical school student selection. Medical Education. 49(1):36–47.

Reiter H, Eva K. 2017. Vive la différence: The freedom and inherent responsibilities when designing and implementing miltiple mini-interviews. Academic Medicine. doi:10.1097/ACM.0000000000002042

Reiter H, Roberts C. 2018.Situational judgement tests in assessment. In Downing, S. M., & Yudkowsky, R. (2018). Assessment in health professions education. Routledge.

Reiter HI, Lockyer J, Ziola B, Courneya CA, Eva K, Canadian Multiple Mini-Interview Research Alliance (CaMMIRA). 2012. Should efforts in favor of medical student diversity be focused during admissions or farther upstream? Academic Medicine. 87(4):443-8.

Rittel HWJ, Webber MM. Dilemmas in a General Theory of Planning Dilemmas in a General Theory of Planning. Policy Sci. 1973;4(2):155–69.

Robb N, Dunkley L, Boynton P, Greenhalgh T. 2007. Looking for a better future: identity construction in socio-economically deprived 16-year olds considering a career in medicine. Social Science & Medicine. 65(4):738–754.

Roberts C, Khanna P, Rigby L, Bartle E, Llewellyn A, Gustavs J, Newton L, Newcombe J P, Davies M, Thistlethwaite J and Lynam J. (2017). Utility of selection methods for specialist medical training: A BEME (best evidence medical education) systematic review: BEME guide no. 45. Medical teacher, 40: 1, 3-19. DOI: 10.1080/0142159X.2017.1367375.

Saha S, Guiton G, Wimmers PF, Wilkerson L. 2008. Student body racial and ethnic composition and diversity-related outcomes in US medical schools. J Am Med Assoc, 300(10):1135-1145.

Schuwirth LW, van der Vleuten CP. 2006. Challenges for educationalists. BMJ 333:544–546.

Seyan K, Greenhalgh T, Dorling D. 2004. The standardized admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex. British Medical Journal. 328(7455):1545–1546.

Siu E, Reiter HI. 2009. Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. Advances in Health Sciences Education. 14(5):759. doi:10.1007/s10459-009-9160-8

Sianou-Kyrgiou E, Tsiplakides I. 2011. Similar performance, but different choices: social class and higher education choice in Greece. Studies in Higher Education. 36(1):89–102.

Siyam A, Zurn P, Rø OC, Gedik G, Ronquillo K, Co CJ, Vaillancourt-Laflamme C, dela Rosa J, Perfilieva G, & Dal Poz MR. 2013. Monitoring the implementation of the WHO global code of practice on the international recruitment of health personnel. Bulletin of the World Health Organization, 91:816-823.

Smith SD. 2008. The global workforce shortages and the migration of medical professions: The Australian policy response. Australia and New Zealand Health Policy 5(7).

Southgate E, Kelly BJ, Symonds IM. 2015. Disadvantage and the 'capacity to aspire' to medical school. Medical Education. 49:73–83. doi: 10.1111/medu.12540.

Stegers-Jager KM. 2018. Lessons learned from 15 years of non-grades-based selection for medical school. Med Educ. 52(1): 86-95.

Stevens FCJ. 2009. Innovations in medical education: European convergence politics and culture. In: Brosnan C, Turner BS, editors. Handbook of the sociology of medical education. London: Routledge. pp 177–184.

Stevens FCJ, Simmonds Goulbourne JC. 2012. Globalization and the modernization of medical education. Medical Teacher. 34: e684–e689.

Suciu ŞM, Popescu CA, Ciumageanu MD, Buzoianu AD. 2017. Physician migration at its roots: a study on the emigration preferences and plans among medical students in Romania. Human Resources for Health. 15(1):6. doi:10.1186/s12960-017-0181-8

Tekian A. (1998). Minority students, affirmative action and the admission process: A survey of 15 medical schools. Academic Medicine, 73: 986–992

Tekian A, Boulet J. 2015. A longitudinal study of the characteristics and performances of medical students and graduates from the Arab countries. BMC Medical Education. 15(1):200. doi:10.1186/s12909-015-0482-3

Tiffin PA, Dowell JS, McLachlan JC. 2012. Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort. BMJ. 344:e1805.

Tiffin PA, Mwandigha LM, Paton LW, Hesselgreaves H, McLachlan JC, Finn GM, Kasim AS. 2016. Predictive validity of the UKCAT for medical school undergraduate performance: A national prospective cohort study. BMC Medicine 14:140.

UK Clinical Aptitude Test (UKCAT). https://www.ukcat.ac.uk/

Undergraduate Medicine and Health Sciences Admissions Test (UMAT). ACER; Australia https://umat.acer.edu.au/

Whitla DK, Orfield G, Silen W, Teperow C, Howard C, Reede J. 2003. Educational benefits of diversity in medical school: A survey of students. Acad Med. 78(5):460-466.

Wong G, Greenhalgh T, Westhorp G, Pawson R. 2012. Realist methods in medical education research: what are they and what can they contribute? Medical Education, 46(1), 89-96.

Woolf, K., Potts, H. W., & McManus, I. C. 2011. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. BMJ, *342*, d901.

Wouters A. 2018. Effects of medical school selection on student motivation: A PhD thesis report. Perspectives on Medical Education. 7(1): 54-57.

Wouters A, Croiset G, Isik U, Kusurkar RA. 2017. Motivation of Dutch high school students from various backgrounds for applying to study medicine: A qualitative study. BMJ. 7(5):e014779.

Wouters A, Croiset G, Schrisema NR, Cohen-Schotanus J, Spaai GWG, Hulsman RL, Kusurkar RA. 2017. A multi-site study on medical school selection, performance, motivation and engagement. Adv in Health Sci Educ. 22:447-462.

The World Health Organization: The World Health Report 2006 – Working together for health. Geneva. 2006.

Xu G, Fields SK, Laine C, Veloski JJ, Barzansky B, Martini CJM. The relationship between the race/ethnicity of generalist Physicians and their care for underserved populations. Am J Public Health 1997;87(5):817-822.

Young ME, Razack S, Hanson MD, Slade S, Varpio L, Dore KL, McKnight D. 2012. Calling for a broader conceptualisation of diversity: surface and deep diversity in four Canadian medical schools. Academic Medicine. 87(11):1501–1510.

Table 1. Summary of the evidence and implications for different selection methods (Patterson et al, 2016; Roberts et al, 2017).

Selection method	Research evidence and implications
Academic achievement	There is a high level of consensus regarding predictive validity, but concerns that the discriminatory power is diminishing as increasing numbers of students achieve top high school grades. Moreover, comparability across different types of schools and school systems is frequently questioned. Academic records can negatively impact widening access if school systems are socially selective.
Aptitude tests	Different tests are located on a continuum between pure ability and pure knowledge tests. Reliability tends to be favourable. The knowledge parts especially in the natural sciences predict study performance. Evidence is mixed on the fairness of aptitude tests, specifically regarding predictive validity, so each tool requires evaluation in its own right. Knowledge tests are also often used for postgraduate selection and demonstrate in this context favourable reliability and validity.
Personal statements & CVs	Candidate acceptability is high, but susceptibility to coaching and plagiarism is also high. There is very little evidence for predictive validity. For postgraduate recruitment CVs are often used as part of an interview but usually in a non-standardised format.
References & Letters of recommendation	Little research supporting validity or reliability and high costs for scoring. However, use of references remains widespread and candidate reactions are positive. References might be used to flag problematic applicants.
Situational judgement tests	Improved validity over other high-volume selection tools for non-academic characteristics (e.g. personality tests) and can be mapped to organisational values. Although SJTs can be relatively costly to design, they can be machine-marked

(SJTs)	and delivered online, producing cost savings in high volume selection. Flexibility in format from text-based to multi-media item presentations. Susceptibility to coaching can be minimised through appropriate design.
Personality assessment	Some personality traits have been linked to in-training performance although depending on the personality tool used the evidence is mixed. Where there is a high risk for susceptibility to faking or coaching, personality assessment might be used to drive more focused questioning at interviews (rather than as a standalone instrument without verification).
Interviews/MMIs	Traditional unstructured interviews perform poorly, whereas structured interviews based on a role analysis, with standardised questions, trained interviewers and appropriate scoring, can be reliable and valid methods. MMIs are the most structured type of interviews. They typically comprise six or more interview stations, which broadens the sampling of performance to enhance reliability. MMIs are relatively expensive to design and implement but can offer favourable validity and positive candidate reactions. All types of interviews create the opportunity for a live interaction with applicants which makes them resource-intensive to deliver, but offers other advantages such as enabling a more personal connection between applicant and program. Candidates prefer interviews to other methods.
Selection centres (SCs) using work samples & simulations	Multi-station SCs are relatively expensive to design and implement as they involve a range of simulations (e.g., group exercises, in-tray tasks, presentations, interactive exercises with role players). They offer a similar multi-sampling approach as MMIs. SCs have been used in postgraduate selection and further evidence of the predictive validity of SCs in undergraduate medical selection is required.

Table 2. Recommendations

1. Use **validated taxonomies of desirable behaviours** that indicate success as a healthcare practitioner to judge the quality of selection, which are contextualised and relevant across stages of training.
2. Develop **more sophisticated research designs and methodologies to evaluate the validity of selection practices**, using multi-site high quality data sets, that take differences in context and content of selection practices into account.
3. Consider the opportunities and challenges of using **complex big data approaches to evaluate selection policies and practices linked to health outcomes**.
4. Explore the ways in which **social accountability** agendas of universities support **social inclusion**, address **workforce issues**, and navigate **government and institutional policy** issues while including the **patient perspective.**
5. Methods for selection must ensure **fairness** and **accountability** for all candidates. Research is required to explain why there is significant **variation** in differing contexts with respect to approaches to address **diversity and widening access.** In addition, more robust evidence is required to evaluate the effectiveness of **outreach, targeting strategies, preparation programmes** and other means to **widening access**.
6. Research is required to better understand the role that selection and recruitment practices might play, in influencing the **high migration rates** of health profession graduates from many developing countries to those in the developed world.
7. The opportunities and challenges of enacting **large scale, nationally coordinated** approaches to selection should be compared and contrasted with **local arrangements** for selection and recruitment in different contexts.
8. Consider working with economists in policy planning that uses **econometric evaluation and cost benefit analysis** of the infrastructure required for delivering selection processes, selection methods, decision-making and reporting in a range of contexts internationally.
9. Use **interdisciplinary theoretical frameworks** to facilitate the future development of both selection policy and practice using appropriate methodologies which are sensitive to local contextual priorities.
10. Investigate systematic approaches that can ensure a **greater translation of selection research evidence into policies and practice**.