Title	Medical student selection: a mixed methods study
Author(s)	Kelly, Maureen Eleanor
Publication Date	2015-09-29
Item record	http://hdl.handle.net/10379/5278

Medical Student Selection: A Mixed Methods Study

A thesis submitted to the School of Medicine, National University of Ireland Galway in fulfilment of the requirements for the degree of Doctor of Philosophy

By

Maureen Eleanor Kelly

M.B., B.Ch, B.A.O., DCH, DObs, M. Med Ed., MICGP, FRCGP

Research Supervisor

Professor Andrew W Murphy

Graduate Research Committee

Professor Fidelma Dunne (Co-Supervisor)

Professor Martin O'Donnell

Professor Andrew Murphy

School of Medicine

Discipline of General Practice

NUI Galway

Date May 2015

Table of Contents

List of Abbreviations

AAMC	Association of American Medical Colleges
ACER	Australian Council for Educational Research
BMAT	Bio-Medical Admissions Test
CAO	Central Applications Office
CI	Confidence Interval
CY1	University College Cork, First Year Medicine
CY2	University College Cork, Second Year Medicine
EU	European Union
GAMSAT	Graduate Medical School Admissions Test
GFY	National University of Ireland, Galway Foundation Year
GMC	General Medical Council, UK.
GMed1	National University of Ireland, Galway, First Year Medicine
GPA	Grade Point Average
GY2	National University of Ireland, Galway, Second Year Medicine
HPAT-Ireland	Health Professions Admission Test-Ireland
HRB	Health Research Board
IELTS	International English Testing System
IMGs	International Medical Graduates
LCE	Leaving Certificate Examination
MCAT	Medical College Admission Test
MCQ	Multiple Choice Questions
MMI	Multiple Mini Interview
NUI Galway	National University of Ireland, Galway
OECD	Organisation for Economic Co-operation and Development
OSCE	Objective Structured Clinical Examination
SEC	Socioeconomic Class
Trad I	Traditional Interview
UCC	University College Cork
UKCAT	UK Clinical Aptitude Test
UMAT	Undergraduate Medicine and Health Sciences Admission Test
USMLE	United States Medical Licensing Examination

List of Tables

List of Figures

Declaration

This work is submitted to fulfil the requirement of the degree of Doctor of Philosophy, at the National University of Ireland, Galway. No part of this thesis has been previously submitted at this or at any other university. Apart from due acknowledgement, it is entirely my own work.

Signed: _____ Date: <u>3rd May 2015</u>

Maureen E Kelly

Acknowledgments

Dedication

This thesis is dedicated to Paul, Róisín, Eoin, Niall and Molly- the meaning in my life.

Abstract

Selection to medicine is widely considered the single most important assessment of a doctor's medical career. It is a key deciding factor in the composition of the medical workforce.   Selectors have a social responsibility to apply best evidence to the design, use and evaluation of selection tools. To be defensible, selection tools must be able to predict future performance, be acceptable to stakeholders and be feasible, practical and affordable.

This thesis uses a multi-phase mixed methods research design to establish the predictive validity of and stakeholder perspectives on two selection tools; the Health Professions Admission Test Ireland (HPAT-Ireland) and Multiple Mini Interview (MMI). It synthesises the research evidence describing stakeholder views of selection tools for medicine (a systematic review), establishes the predictive validity of HPAT-Ireland for communication and clinical skills (an observational study), explores the views of doctors on the job-relatedness and acceptability of HPAT-Ireland (a qualitative study- informed by the principles of grounded theory tradition), demonstrates the feasibility of MMI in an Irish setting (quantitative study) and explores the predictive validity and stakeholder acceptability of MMI in an internationally diverse student population (a mixed methods - explanatory sequential study).

The findings from this programme of research indicate that HPAT-Ireland moderately predicts communication and clinical skills in Year 2 of the undergraduate course. Stakeholder perceptions of the acceptability and job relatedness of HPAT-Ireland are reasonably good, however not uniformly so across its subsections. Concerns exist regarding its potential for negative impact on socioeconomic diversity. MMI was perceived as an authentic assessment with high levels of job relatedness, by both assessors and candidates. However cultural issues and English language proficiency are considered important potential barriers to students from different backgrounds. While MMI is feasible in an Irish setting this thesis did not demonstrate evidence for its ability to predict medical student future performance.

Chapter 1: Introduction

Figure 1 Class of 1924 Yale Medical School

Photograph reproduced by kind permission of Medical Historical Library, Cushing/Whitney Medical Library, Yale University- See Appendix Chapter 1.

1.1 Background

Medicine is an attractive career choice for many people. Currently over 2,370 medical schools operate worldwide (Foundation for Advancement of International Medical Education and Research, 2014). Despite this large number of institutions, the international experience is that the number of applicants to medical schools far exceeds the number of places available. Hence some mechanism for deciding who should be offered a place is a necessity. This is not a new phenomenon. Figure 1 is a photograph of Yale Medical School graduating class of 1924. In 1920 when they applied to medicine, Yale had already earned a reputation as one of the finest medical schools in North America, attracting two hundred applicants for the fifty six places available. According to the Medical Historical Library, the Dean could "*afford to be selective*". Offers of places to

certain groups, including women, Italians, Catholics, Jews and African Americans were "*deliberately limited*" (Cushing/Whitney Medical Library, 2010). Yale was not alone; such selection practices were commonplace at a time when one of the main criteria for selection to medicine was the ability to pay tuition fees (Reiter and Eva, 2011). Many both within and without the medical profession were unhappy with this. Urged on by the Flexner Report (1910) and changes in societal norms medical schools began to seriously engage with devising more equitable, transparent and acceptable selection processes. By 1946 authors in the British Medical Journal were actively debating the issues of selection (Barlett, 1946, Smyth, 1946a, Wilkie, 1946). It quickly became apparent that the seemingly straight forward task of fairly and transparently selecting a small number of candidates from a large pool of potentially suitable applicants was beset with difficulties. Given the complexities involved one author concluded that the challenges presented by selection "*were formidable*" (Smyth, 1946b).

Today we recognise that the purpose of an effective selection tool is to rank applicants in order of their suitability to the practice of medicine and their likelihood of making good doctors (Powis, 1994). Selection tools operate in the predictive paradigm – they detect trainability (Patterson and Ferguson, 2007). A key consideration therefore in designing selection tools is: What should the end product of this training look like? The intuitive answer is a good doctor – but "*A good doctor will be different things to different people at different times*" (Tonks, 2002). There are many facets to being a good doctor and there is little consensus within the profession about what these are (Hurwitz and Vass, 2002, Rizo and Jadad, 2002, Tonks, 2002). One paper cites eighty-seven positive attributes of a good doctor (Albanese et al., 2003). When the British Medical Journal asked its readership what qualities were needed to make a good doctor they narrowed the list down to a mere seventy (Tonks, 2002).

Best practice in the design and development of selection tools recommends that first a multi-source job analysis be carried out to identify the skills, attributes, abilities and characteristics for on the job performance (Patterson and Ferguson, 2010). This provides for the development of a competency model and creation of a person specification. Appropriate selection tools are then developed and

mapped against both the job analysis and person specification. For many reasons, in practice these critical steps are not always undertaken with respect to selection in medicine. This leads to a lack of clarity about what medical schools are looking for in successful applicants (Kreiter and Axelson, 2013). There is a good degree of consensus that "*a good physician needs much more than "book smarts," and thus other relevant variables must be considered in the admissions process*" (Kulatunga Moruzi and Norman, 2002). However agreeing on a list of such variables is proving difficult (Albanese et al., 2003, Bardes et al., 2009, Benbassat and Baumal, 2007, Norman, 2010). General mental ability, critical thinking, problem solving, empathy, integrity, communication skills and psychological robustness are some of the most commonly cited qualities. Designing selection tools that measure across these facets in a reliable and valid way is difficult (Salvatori, 2001).

1.2 The Importance of Selection

Each year in the UK there are over 19,000 applicants to medicine and approximately 42,000 in the USA (Brown and Lilford, 2008).   As each applicant is entitled to apply to more than one medical school, the number of applications per year is substantially higher. In the UK alone, the figure is over 40,000 applications per year (Ferguson et al., 2002). In excess of 3,000 Irish school leavers annually put medicine as their first preference college course (Central Applications Office, 2012a). Selectors have a moral duty to apply best evidence principles to the selection of these and future applicants (Norman, 2004). In fact, the annual selection of medical students is considered by many as one of the central activities of a medical school faculty (Veloski et al., 2000). Selection is conceptualized as the first *assessment* in the medical education and training pathway (Cleland et al., 2012, Prideaux et al., 2011). Moreover, it is increasingly viewed as "*the most important assessment decision of a doctor's medical career*" (Harden, 2014).

Selection is indisputably the key "*determinant of who will practice medicine*" (Reiter and Eva, 2011) as once selected most entrants graduate (Eva and Reiter, 2004, Prideaux et al., 2011, Searle and McHarg, 2003). Therefore, the composition and calibre of the future medical workforce is significantly

dependent on the tools used to select students (British Medical Association, 2009, Lakhan, 2003). Selection decisions therefore, are truly high stakes (Albanese et al., 2003, Roberts et al., 2009).

The median cost to the student per year to attend medical school in the USA is $54,000 in public and $74 000 in private medical schools (Dunleavy et al., 2013). While the estimated cost to train a medical doctor in the UK is approximately £200,000 (€270,000) (Steele, 2011). Selecting those most likely to progress through medical school unhindered by failing examinations therefore has important implications for minimising student debt, efficient use of educational resources and prompt transit into the workforce (Dunleavy et al., 2013). Medical Schools have an added burden of duty educating poorly chosen applicants. Much greater costs however are borne by society and individual patients if applicants are selected who go on to underperform as doctors (Marrin et al., 2004, Reiter and Eva, 2005).

The potential negative consequences of poorly designed selection tools serve to further highlight its importance. These include the possibility of deterring good applicants from applying because of perceived or real biases in the system (Patterson et al., 2011), inadvertently de-selecting applicants such as potential future general practitioners or surgeons (Poole and Shulruf, 2013) and selecting a cohort of doctors who are insufficiently socially, ethnically, culturally and linguistically diverse to best serve the public (Carrasquillo and Lee-Rey, 2008, Cleland et al., 2012, Saha et al., 2008). Finally, the importance of selection is emphasized by research from Holland and Austria, countries which have recently moved from open or lottery based access to medicine to restricted access via selection. This has confirmed that selection represents more efficient use of tax payers' money as evidenced by reduced drop-out rates, reduced time to course completion-hence quicker transit into the workforce and results in better clinical clerkship grades at medical school (Reibnegger et al., 2010, Urlings-Strop et al., 2009, Urlings-Strop et al., 2011).

1.3 How Are Good Selection Tools Recognised?

1.3.1 Evaluative Standards

There are a growing number of reports that provide guidance on the expected standards of selection procedures (Cleland et al., 2012, Medical Schools Council, 2014a, Prideaux et al., 2011).   According to a General Medical Council commissioned report, identifying best practice in the selection of medical students, there are fourteen recognized evaluative standards for selection tools (Cleland et al., 2012). See Table 1.

Table 1 Evaluative Standards for Selection Procedures

1.	Establishing reliability and validity of the selection tool
2.	Positive employee /student reactions
3.	Ensuring ease of interpretation
4.	Ensuring generality of use
5.	Minimising costs and maximising value
6.	Practicality and administrative convenience
7.	Legality
8.	Stakeholder acceptance
9.	Expertise required for analysis and interpretation of information generated by the selection tools
10.	Utility
11.	Fairness
12.	Educational impact/value
13.	Generates appropriate feedback
14.	Procedures are in place for ongoing validation, evaluation and renewal of assessment tools

(Reproduced by kind permission of the author, see Chapter 1 Appendix 1).

The ideal tool would rate well across all standards. At the most basic level selection tools need to be feasible. On a practical level, feasibility includes cost, time, human resources and workability. Once feasibility is established much of the debate and research on the evaluation of selection tools focuses a shortened list of four key evaluative standards: validity, reliability, fairness, and stakeholder

acceptability. Table 2 is an adaptation of the fourteen evaluative standards, which maps many of them to the shortened list of commonly applied standards. In all cases selection must comply with legal requirements. Some of the original standards map to more than one of the shortlist of key standards. Validity and reliability are essential to the interpretation of scores on any selection tool. They are distinct yet interdependent concepts, and so are often considered separately. No selection tool is perfect; validity and reliability are always a matter of degree (Cook and Beckman, 2006).   They are highly dependent on the context within which the selection tool is applied and the content of the test (Eva et al., 2009).

Detailed Evaluative Standard		Short list of key standards
Establishing reliability and validity of the selection tool	→	Reliability
Procedures are in place for ongoing validation, evaluation and renewal of assessment tools	→	Validity (including predictive, political & consequential)
Expertise required for analysis and interpretation of information generated by the selection tools		
Utility		
Educational impact/value		
Fairness	→	Fairness
Positive employee /student reactions	→	Stakeholder Acceptability
Stakeholder acceptance		
Utility		
Generates appropriate feedback		
Minimising costs and maximising value		Feasibility
Practicality and administrative convenience	→	
Ensuring ease of interpretation		
Expertise required for analysis and interpretation of information generated by the selection tools		
Legality / Ensuring generality of use		Not mapped to shortlisted standards

Table 2 Key Evaluative Standards for Selection Procedures

(Adapted by M Kelly)

1.3.2 Validity

A selection tool is neither valid nor invalid, rather its validity is judged by the weight of evidence to support the interpretation of scores for a given purpose at a particular time (American Educational Research Association et al., 1999, Downing, 2003).   Validity is not a property of the selection tool itself but of the tool's scores and their interpretation, i.e. validity is a property of inference (Cook and Beckman, 2006, Streiner and Norman, 2008)   Predictive validity describes how well candidate scores on a selection tool predict some future outcome. It is typically expressed as a correlation between the scores on selection tools achieved by successful applicants and the grades they subsequently achieve on assessments during their medical school education and in some cases beyond, into higher professional training and licensure examinations (Cleland et al., 2012). Strengths of correlations are often compared using Cohen's effect size interpretations (small ≥.10, medium ≥.30, large ≥.50) (Cohen, 1992).   The predictive validity of a selection tool is helpful if it fulfils the following four criteria: statistically significant, practically relevant, consistent across multiple studies and adds incremental validity over and beyond other predictors (Siu and Reiter, 2009). Incremental validity is a measure of the added value any additional selection tool brings to the predictive power of the pre-existing selection process (Siu and Reiter, 2009).

Predictive validity studies are problematic because of three recognised complications (Cleland et al., 2012, McManus et al., 2013a, Patterson and Ferguson, 2010). Firstly, applicants who are successful in gaining a place medical school necessarily have performed well on the selection tool. Hence the range of scores of those admitted is narrower than the range of the entire applicant pool. This is known as range restriction. Secondly, many applicants attain the maximum attainable score on a selection tool and this "ceiling" effect is known as right-censorship. Both restriction of range and right censorship attenuate the correlation between selection and outcome scores. Thirdly, deciding the most appropriate outcome measure by which to validate the selection tool is not always obvious. Debate exists as to whether performance in medical school is an appropriate outcome measure by which to validate selection tools, as opposed to

later performance as a graduated doctor. This is known as the criterion problem (Cleland et al., 2012).

Traditionally the main criticism of predictive validity is the view that "*it is by no means certain that examination performance is an accurate reflection of true clinical performance*" (Leinster, 2013). However, findings from a longitudinal programme of research from Canada challenge this viewpoint. Studying primary care physicians, Tamblyn et al (1998) established that every standard deviation increase in doctors' licensing examination scores reflected significantly better clinical performance in terms of appropriate referral, safe prescribing and mammography screening rates. In addition follow up studies determined that licensing examination scores continued to predict quality patient care markers up to 7 years in practice and also predicted the numbers of complaints made about participating physicians to the medical regulatory authorities (Tamblyn et al., 2007, Tamblyn et al., 2002).

Likewise in the United States, Norcini et al (2002) demonstrated that holding all other variables constant post graduate examination success was associated with a 19% reduction in patient mortality. Similarly, there is evidence that in the majority of instances academic achievement in medical school predicts performance on higher professional training examinations (Hamdy et al., 2006, McManus et al., 2013b). Taken together, these studies provide evidence of a legitimate thread between medical school assessment, higher professional examinations and markers of excellence in clinical performance.

1.3.3 Reliability

Reliability refers to the reproducibility of assessment data or scores, over time or occasions (Downing, 2004).   It provides an estimate of the amount of random measurement error in the assessment (Downing, 2004). In terms of written selection tools, internal consistency is important. It takes account of a number of conditions. "Test-retest" measures the extent to which an applicant would score the same if she took the exact same test on a separate occasion. This is more commonly measured by splitting the test in half and comparing the scores on one half of the test with scores on the other half.   "Parallel test" describes the extent

to which the applicant would achieve the same score had a different but equivalent version of the tool been used. With respect to selection tools that involve raters or human judgements (for example scoring personal statements) inter-rater reliability is a key consideration (Cook and Beckman, 2006).

The reliability of a selection tool governs the maximum attainable validity (Kreiter and Axelson, 2013, Streiner and Norman, 2008). Poor reliability, in either the selection tool or the outcome measure, negatively impacts on predictive validity measurements (Patterson and Ferguson, 2010). A selection tool that is not reliable cannot support valid interpretations. Furthermore even a reliable test may be invalid, as it may be reliably testing something other than the construct of interest (Crossley et al., 2002), and could in fact be "reliably wrong" (Patterson and Ferguson, 2012).   The biggest threat to reliability is insufficient sampling (Schuwirth and van der Vleuten, 2010). Like validity, reliability is not a property of the instrument but of the score (Cook and Beckman, 2006). This is highly relevant to selection where the same instrument used in different settings or contexts, may demonstrate very different reliability. For example, a tool that enjoys good levels of reliability for selecting to undergraduate medical school may not prove as reliable in a postgraduate setting where the pool of applicants to a particular sub-speciality may be a far more homogenous group.

1.3.4 Fairness

The fairness of a selection tool is dependent on three main considerations (Patterson and Ferguson, 2010). Firstly, the tool should have objective and valid criteria that are based on a thorough job analysis. Secondly, it should involve accurate and standardised assessment by trained personnel. And thirdly, outcomes should be continually monitored.

Despite what it might look like on the surface, unfairness or selection tool bias cannot be assumed simply because members of different sub-groups obtain different test scores (Patterson and Ferguson, 2007). Selection tool bias occurs "*when deficiencies in a test itself or the manner in which it is used result in different meanings for scores earned by members of different identifiable subgroups*" (American Educational Research Association et al., 1999, Davis et al.,

2013). For instance, if the questions in a selection tool drew on experiences that were more familiar to one group of applicants than another and were unrelated to the construct being measured by the tool, this would give rise to bias (Davis et al., 2013). If the tool was shown not to have adequate job relatedness then it would be unfair to make selection decisions based on this measure, particularly when certain groups were disadvantaged by the test (Patterson and Ferguson, 2007). Likewise if the selection tool was not shown to be an important predictor of future performance or if it was found to under predict the future performance of certain groups of applicants, then this could represent bias. With an unbiased selection tool, students from different groups, achieving the same score, should go on to perform at the same level in their in-course examinations. If however one sub-group was found to score higher on their examinations than predicted by their selection tool score, then this may be evidence that the tool was biased (Sackett et al., 2008). Fairness of selection tools is therefore closely related to both their job relatedness and predictive validity (Patterson and Ferguson, 2007).

1.3.5 Stakeholder Acceptability

Acceptability is a key component of the utility of any assessment tool (Van Der Vleuten, 1996). In assessment for selection, acceptability assumes even greater importance due to the numerous stakeholder groups and their differing agendas and expectations. Best practice in the design, development and continued use of selection tools is an iterative process informed by regular feedback from stakeholders (Patterson and Ferguson, 2010). Increasingly, stakeholder views are recognised as an important and highly influential measure of the utility and effectiveness of a selection tool. They are so influential that the term political validity has been coined to describe "the *extent to which various stakeholders and stakeholder groups consider the tool(s) to be appropriate and acceptable for use in selection*" (Cleland et al., 2012). Political validity is an important consideration in widening access to medical schools (Prideaux et al., 2011).

In some situations, there appears to be a trade-off between stakeholder views and other evaluative measures of selection tools, such as predictive validity and reliability. For example, there is plenty of evidence that personal statements, letters of reference and traditional interview have limited predictive validity and

are prone to bias (Benbassat and Baumal, 2007, Ferguson et al., 2002, O'Flynn, 2010, Salvatori, 2001, Siu and Reiter, 2009). Paradoxically, these tools continue to be widely used internationally. It has been argued that this can in part be explained by these tools serving some other political agenda for which they achieve stakeholder approval (Patterson et al., 2012a). It is crucial therefore that stakeholder views are explored, understood and communicated effectively in order to increase the likelihood that selection processes can be developed to meet both political agendas and conform to best evidence practice.

Despite their significance, stakeholder views of selection tools and processes are relatively under-explored and under-represented in the literature with respect to medical student selection (Cleland et al., 2012). There are a number of potential reasons why this may be the case. Firstly, collating stakeholder views does not form part of the routine collection of data, such as recording undergraduate or postgraduate assessment results used in predictive validation studies, and hence requires additional resources and planning. Secondly, exploring stakeholder views in a meaningful way often entails in-depth, time consuming research methods such as qualitative or mixed methods. Thirdly, stakeholders are a heterogeneous and wide-reaching group. It can include for example medical students, applicants and potential applicants, Medical School Admissions Committees, Medical Faculty, the profession itself, patient groups, school career guidance teachers and society. Even within these groups, there are subgroups whose individual voices need to be heard for example applicants for whom English is not a first language (Bardes et al., 2009). The medical profession itself has a great stake in the selection process. Ultimately, its members will be working in the medical environment, sharing patient care and clinical decision making with the successful applicants once they graduate. The particular stake that future co-workers have in the outcomes of selection decisions has been emphasised in the field of work psychology (Gilliland and Cherry, 2000). The usefulness of establishing the views of doctors, not directly involved in the selection process, have been highlighted as they *"represent the broader population of medical professionals"* (Ginsburg et al., 2004, Murphy et al., 2008).

Chapter 1: Introduction

In the field of medicine, unlike that of psychology, there has been very limited use of theoretical models to conceptualise stakeholder reactions to selection tools. Over the past fifty years, work psychologists have developed organisational justice theories to describe perceptions of fairness in organisational processes (Colquitt et al., 2005). Gilliland (1993, 1994) first proposed using them as a way to comprehend applicants' reactions to and perceptions of selection tools.

Patterson et al (2011) established their usefulness in understanding applicant reactions to selection in the medical profession. Organisational justice theories can be categorised as distributive, procedural and interactional justice (Greenberg and Colquitt, 2005).

Distributive justice refers to the perceived fairness of organisational outcome distributions (Chambers, 2002). In the context of selection this relates to the fairness of the selection outcome - such as medical school places, in terms of equal opportunity and equity (Patterson et al., 2011). From a distributive justice perspective selection is viewed to be fair when everyone receives the same opportunities.

Procedural justice relates to the perceived fairness of the procedures used to arrive at the resource allocation decisions (Chambers, 2002). With respect to selection, procedural justice concerns the perceived fairness of the selection tool in terms of job relevance and characteristics of the test (Patterson et al., 2011). From a procedural justice perspective selection is viewed more positively when the methods used in selection are connected with the job, when the purpose of the methods is explained to applicants and when applicants receive feedback on why they were not successful (Gilliland, 1994, Truxillo et al., 2004). Perceptions of fairness of the selection process also relate to measures of reliability and validity. There is evidence that applicants' perceptions of procedural justice are higher for selection tools which demonstrate good predictive validity (Fodchuk and Sidebotham, 2005).

Lastly interactional justice describes the interpersonal treatment people receive as organisational procedures are enacted (Chambers, 2002). In terms of selection it refers to how applicants are met during the selection process and

includes the information applicants are given as well as the manner in which it is conveyed (Bies and Moag, 1986, Kanerva et al., 2010). The fairness of the communication is a very influential determinant of how interactional justice is perceived (Erdogan, 2003, Kanerva et al., 2010).

1.3.6 Widening Diversity

Widening diversity is not listed as one of the evaluative standards for selection tools to medicine. Yet it is increasingly recognised as important in selection to higher education in general and medicine specifically (Higher Education Authority, 2008, Milburn, 2009). One of the principal motivations is a growing acceptance of the need to widen the demography of the medical workforce to better mirror that of the general public (American Medical Association, 2014 , British Medical Association, 2009). According to the Ottawa Consensus statement on selection, medical schools need to be aware of the communities that they serve and ensure that these communities are represented amongst their entrants (Prideaux et al., 2011). There are widespread mismatches between the socio-demographic variables of medical students and the population at large (Garlick and Brown, 2008). For example only approximately 5% of UK medical students have parents from non-professional backgrounds (Tiffin et al., 2012). While in the United States, racial and ethnic minorities comprise 26% of the population, but only 6% of the medical workforce (American Medical Student Association, 2014).

Widening access to medicine refers to a policy of increasing the numbers of people from groups that traditionally have lower participation in medicine (Medical Schools Council, 2014a). These include "*students from disadvantaged backgrounds, mature students, students from ethnic and cultural groups and disabled students …"* (Cleland et al., 2012). There are a number of recognised benefits of widening access to medicine. Diversity in the medical workforce is linked with better access and quality health care for underserved populations (Lakhan, 2003). Some patients prefer attending doctors from their own cultural, racial, ethnic and linguistic background (Carrasquillo and Lee-Rey, 2008). Underrepresented and minority medical students have been shown to be substantially more likely to plan to serve the underserved (Saha et al., 2008).

Furthermore, white students who attend medical school with higher proportions of racial diversity rate themselves as better prepared than students at less diverse schools to care for racial and ethnic minority patients and had stronger attitudes about inadequate access to health care (Saha et al., 2008). Students report that enhanced diversity within the medical school impacts positively on their medical education (Whitla et al., 2003). Finally, widening access to medicine helps to improve social mobility and equality (Cleland et al., 2012). Therefore it is important to recognise and evaluate the impact that selection tools have on the widening diversity agenda.

1.4 Selection Tools for Medicine

1.4.1 Categories of Selection Tools

Selection tools for medicine fall into eight broad categories: academic record, tests of general mental ability including aptitude tests; personal statements/ essays/auto biographical summaries; references; situational judgement tests; personality assessment/ emotional intelligence; interviews / multiple mini interviews and selection centres (Cleland et al., 2012). Each category has its own strengths and weaknesses and best practice recommends using a combination of tools (Patterson and Ferguson, 2007). A recent General Medical Council commissioned review provides a description and comparison of each of these groups of selection tools with respect to their validity, reliability, impact on diversity and stakeholder acceptability (Cleland et al., 2012). It demonstrated that the evidence to support aptitude testing, multiple mini interviews, situational judgement tests and selection centres was stronger than for traditional interviews, references and autobiographical submissions (Cleland et al., 2012). This is broadly in keeping with conclusions from elsewhere (Benbassat and Baumal, 2007, Ferguson et al., 2002, Salvatori, 2001, Siu and Reiter, 2009). The triad of academic record, aptitude test and interview are commonly found in selection systems for medicine in USA, Canada, UK, Australia and New Zealand. The latter two tools are the particular focus of this thesis.

1.4.2 Academic Record

Academic record has long been the mainstay of selection to medical school and most admission policies internationally still rely heavily on evidence of academic attainment for selection purposes (Patterson and Ferguson, 2010). Until very recently, selection to medicine for Irish school leavers was based exclusively on evidence of academic record. Conversations regarding selection to medicine often start with the question what was wrong with selection based on academic record alone? Prior academic achievement has been identified as a robust and dependable predictor of future academic performance not only within medicine, but all professional disciplines (Salvatori, 2001). Academic record has repeatedly been shown to have one of the strongest predictive validity coefficients when compared to other measures routinely used for selection to medicine (Cleland et al., 2012, Ferguson et al., 2002, Kreiter and Kreiter, 2007, Kulatunga Moruzi and Norman, 2002, McManus et al., 2013b, Norman, 2004, Salvatori, 2001, Simpson et al., 2014).

In a recent landmark publication McManus et al (2013b) conceptualised the contribution of past academic achievement as the "academic backbone", with each level of achievement providing scaffolding for further learning. In their meta-analysis of five cohort studies, with almost 12,000 participants, in virtually all cases the last level of academic achievement predicted the next, throughout medical school and, importantly, into higher professional training.

However, the evidence clearly shows that academic record is not without its shortcomings. Selection on the basis of academic performance alone has been criticized as "misguided" (Leinster, 2013, Rosenfeld et al., 2008, Searle and McHarg, 2003). Despite its strong predictive validity, academic record accounts for only 23% of the variance in medical student examination performance and 6% of the variance at postgraduate level (Ferguson et al., 2002). When calculating percentage variance in performance from the total applicant pool as opposed to just those who were selected to medicine (so called construct predictive validity), the amount of variance accounted for by past academic record rises to 65% (McManus et al., 2013a). Hence, at a minimum, approximately one third of

variance in medical student performance remains unaccounted for by previous academic record.

Secondly, grade inflation means that it is increasingly difficult to distinguish between applicants based on academic scores alone (James et al., 2010, McManus et al., 2013c, Parry et al., 2006, Wood, 1999). An example is the UK based school leaving test General Certificate of Education Advanced Level (A-level) where the numbers of applicants achieving the maximum three A level grades became a significant challenge to medical school admissions committees (McManus et al., 2013a, Parry et al., 2006). Thirdly academic record is associated with patterns of social exclusion whereby students from professional backgrounds, fee paying, private or selective schooling and white ethnicity achieve higher scores compared to the general population (Archer et al., 2005, James et al., 2010, McManus et al., 2005, Powis et al., 2007a, Schwartz, 2004, Tiffin et al., 2014a). This is in direct contrast to widening participation initiatives.

Finally, selection based on academic record alone takes no account of whether or not applicants possess the personal qualities or attributes deemed desirable for the practice of medicine (Bore et al., 2005, Fernando et al., 2009, James et al., 2010, Powis, 2003). For decades Powis has been arguing that selection to medicine must include at least some assessment of personal qualities both on the grounds of reducing unprofessional behaviours that give rise to patient complaints and dissatisfaction with doctors as well as decreasing the incidence of burnout and stress amongst doctors who are ill suited to the career (Powis, 1994, Powis, 1998, Powis, 2003, Powis et al., 2007a, Powis, 2010, Powis, 2014). Furthermore, medical accreditation authorities internationally place increasingly strong emphasis on the importance of qualities including professionalism, communication, advocacy, ethical practice (American Board of Medical Specialities, 2014 , Frank, 2005, General Medical Council, 2009, Medical Council, 2010a). This has heightened awareness of the need to consider these attributes at the selection stage. As one author concluded, whilst *"no one is advocating the abandonment"* of academic record as a core aspect of selection to medicine there is clearly an established need for the inclusion of adjunct selection tools (Nicholson, 2005).

1.4.3 Aptitude Tests

Aptitude tests are amongst the commonest tests of general mental ability used for the selection to medicine. They are widely used across the United States, Canada, the UK, Australia and New Zealand. They can be designed to measure across a broad range of abilities including verbal, numerical and in some cases knowledge based components (Cleland et al., 20112). It is feasible to use these tools for large scale testing. The Medical College Admission Test (MCAT) is one of the longest standing and most researched such test available. There is good evidence that it is predictive of successful applicants' future performance, both in medical school and licensing examinations (Donnon et al., 2007, Ferguson et al., 2002, Julian, 2005, Krieter and Krieter, 2007, Salvatori 2001, Siu and Reiter, 2009). However, newer aptitude tests introduced over approximately the past fifteen years, have not been as well researched (Cleland et al., 2012).   There is limited, even conflicting evidence to support their use, particularly with respect to predictive validity (Lynch et al., 2009, Poole et al., 2012b, Wilkinson et al., 2011, Wright and Bradley, 2010, Yates and James, 2010). The stakeholder acceptability of aptitude tests is also unclear. Concerns have been expressed over potential fairness and possible bias (Davis et al., 2013, Puddey and Mercer, 2013, Tiffin et al., 2014a); the job relevance has been queried (Dhar et al., 2012); they can be viewed as an obstacle by minority applicants (Henry, 2006, Patterson et al., 2009). Questions have been raised about their susceptibility to coaching, although this appears to be largely unfounded (Griffin et al., 2012, Griffin et al., 2008, Wilkinson and Wilkinson, 2013). There have been calls for more research to establish the utility of aptitude tests in general (Cleland et al., 2012, Prideaux et al., 2011, Salvatori, 2001).

1.4.4 Interview

Interviews are used by virtually all medical schools in North America, UK, Australia and New Zealand. One of the most striking features of the interview in selection is that their face validity is high and they enjoy widespread stakeholder acceptability (Cleland et al., 2012). There is good evidence that admissions committees place substantial weight on the interview data when deciding offers of places (Monroe et al., 2013). However, the traditional panel interview has

been widely criticised for variable (and sometimes quite poor) levels of reliability and limited evidence for predictive validity (Benbassat and Baumal, 2007, Ferguson et al., 2002, Goho and Blackman, 2006, Kreiter et al., 2004, Salvatori, 2001, Siu and Reiter, 2009).

One large meta-analysis of the selection interview for health related professions concluded that it was only a weak predictor in terms of future academic and clinical performances (Goho and Blackman, 2006). A review of studies of medical student selection found that in the majority of cases the admission interview did not predict future performance (Benbassat and Baumal, 2007). Structured interviews however have been shown to have better reliability profiles (Albanese et al., 2003).

Context specificity also means that the scores on a single interview may have poor generalisability (Eva et al., 2004a). Additionally, applicants' scores on the traditional panel interview are often influenced by construct irrelevant factors. These include interviewers' leniency or severity, applicants' characteristics such as gender, age, race or physical characteristics, the effect of similarities and dissimilarities between interviewer and interviewees and the tendency to make generalisations based on initial impressions (a phenomenon known as the halo effect) (Goho and Blackman, 2006).

The mismatch between the high face validity and relatively poorer levels of reliability and predictive validity has been identified elsewhere and represents the gap that can exist between political validity and other metrics used to evaluate selection tools (Patterson et al., 2012a).

Multiple Mini interview (MMI) is a particular type of structured interview increasingly used in selection. It was developed by Kevin Eva and colleagues in an effort to devise an interview process that had better psychometric properties than the traditional panel interview (Eva et al., 2004a). It is a process whereby applicants are examined in a timed circuit similar to that used in an Objective Structured Clinical Examination. The typical MMI consists of ten stations, each of eight minutes duration with usually one assessor (Pau et al., 2013). Some stations utilise role players. The main theoretical justification for MMI is that multiple

observations helps to avoid context specificity of single encounters and provides better, more generalizable measure of the applicant's abilities than traditional panel interviews (Eva et al., 2004a, Roberts et al., 2008, Roberts et al., 2009).   The past ten years have seen an explosion in the use of MMI internationally (Pau et al., 2013).

1.5 The Irish Context

Chapter 3 gives a detailed account of selection to medicine in the Irish context. This paragraph summarises key aspects as they inform the rationale for this thesis. Medicine in Ireland is predominately an undergraduate degree with students entering directly from second level school. The selection process however depends on the country of origin of the applicant. Traditionally, for applicants from within the European Union, selection was based exclusively on second level school exit examination scores. In 2006, a government commissioned review of medical education and training in Ireland recommended a move away from selection based on academic grades alone (Fottrell, 2006). This led directly to the introduction of an aptitude test known as the Health Professions Admission Test-Ireland (HPAT-Ireland) (Australian Council for Educational Research, 2014a). It was first used for the selection of medical students in 2009. Since this time, medical school places are now offered to EU school leavers, on a combination of the academic record and applicant's performance on HPAT-Ireland. This process is administered via a central application system. Selection to medicine of EU applicants is strictly regulated. The number of places available is determined and capped at government level. All medical schools apply the same criteria and weighting of admissions data according to nationally agreed guidelines   (Central Applications Office (CAO), 2014). The introduction of HPAT-Ireland has proved controversial and its use has been publically criticised (See Chapter 3).

Irish medical schools also offer a substantial number of places to Non-EU students each year. Non-EU students are fee paying and contribute significant revenue to medical schools and their constituent universities (Fottrell, 2006). Many of these students come for Malaysia, United States and Canada. Applicants

from outside of the EU are not required to sit HPAT-Ireland. The selection criteria for this group are not uniform across the medical schools. Depending on the school involved these students must satisfy academic requirements, undergo structured interviews, sit English language proficiency tests, in some cases provide evidence of suitability to a career in medicine by their life experiences (volunteering, team activities etc.) and may be required to take a test of general mental ability or aptitude such as the Medical College Admission Test (MCAT).

At the outset of this thesis there were virtually no studies exploring the selection of medical students in an Irish context. Similarly, with the exception of the Fottrell Report 2006, there was very little critical debate within the medical profession regarding the relative merits or evidence base to support the various selection tools used, either for EU or Non-EU applicants.

1.6 Statement of Problem

Medical student selection is a complex issue of importance both within the medical profession and public domain. It is a powerful determinant of the future medical workforce with significant implications for provision of optimum care to patients.

Medicine is considered an esteemed and privileged profession. It continues to attract an abundance of excellent applicants, the numbers of which far exceed the places available for training. The purpose of selection is to rank order these applicants in as valid, reliable, fair, feasible and acceptable a manner as possible. Public and professional expectations demand that selection tools should be able to accurately predict those who will make the best future doctors. The evaluative standards that good selection systems must uphold recognise that selection to medicine does not occur in a social vacuum. There are many important groups with a stake in the selection process. Hence balancing the views of these groups against other metrics of selection tools such as predictive validity is an important consideration in the political validity of selection tools (Patterson et al., 2012a).

Aptitude tests are widely employed in the selection of medical students internationally. However, their use is controversial and they are not uniformly

endorsed. Questions remain to be answered about both their predictive validity and stakeholder acceptability. HPAT-Ireland is one of the newest aptitude tests for medicine to become available. At the outset of this thesis neither the predictive validity, nor stakeholder acceptability of HPAT-Ireland were known.

Multiple Mini Interview is rapidly gaining popularity. The need to establish the predictive validity and acceptability of MMI in different educational and cultural contexts has been highlighted (Husbands and Dowell, 2013, Pau et al., 2013). MMI is considered a labour intensive and resource demanding tool. At the outset of this thesis there were very few published papers describing the feasibility, predictive validity and acceptability of MMI outside of North America.

1.7 Research Aim and Objectives

The aims of this thesis are two-fold:

1. To establish the predictive validity of HPAT-Ireland and explore its stakeholder acceptability.
2. To establish the feasibility and predictive validity of MMI and explore its stakeholder acceptability.

The specific objectives are to

1. To establish the predictive validity of HPAT-Ireland with respect to student performance on tests of communication and clinical skills in the early undergraduate years.
2. To explore the perspectives of doctors, a key stakeholder group, on the job-relatedness of HPAT-Ireland and its acceptability as a selection tool.
3. To establish the feasibility of running a MMI in an Irish setting
4. To establish the predictive validity and fairness of MMI in an internationally diverse student population
5. To explore the perspectives of two stakeholder groups, medical students and MMI interviewers, on the acceptability of MMI, with a particular focus on the selection of international students.

1.8 Setting of Study

The study takes place in one of Ireland's six medical schools; the National University of Ireland, Galway (http://www.nuigalway.ie/). Originally known as Queen's College Galway, it opened its doors to 63 students for the first time in the academic year 1849/ 50, with three Faculties: Arts; Medicine and Law and two additional Schools: Agriculture and Engineering. Today it has over 17,000 registered students from 92 different countries, 2,400 members of staff and over 90,000 alumni. NUI Galway medical degree is entirely an undergraduate programme. It is of five or six year's duration depending on the scores in relevant science subjects acquired by the medical school applicant in their second level school. On average 200 students are enrolled annually. There are a number of entry routes to pursue an undergraduate medical degree in Ireland and these are discussed further in Chapter 3. This thesis also involved significant collaboration with two other medical schools at Dundee University and University College, Cork.

1.9 Overview of Thesis

Chapter 2 reviews the relevant international literature on selection of medical students. Chapter 3 describes medical student selection in the Irish context. Chapter 4 discusses key methodological considerations. The thesis comprises four discrete research studies. Figure 2 describes diagrammatically how these studies connect to each other and the overall thesis. They are presented as chapters that relate to the resultant papers that have been published in the peer reviewed academic press. Chapter 5 reports on a quantitative study that established the predictive validity of HPAT-Ireland for communication and clinical skills in the early undergraduate medical course (Study 1). This study was a two year follow-up of students in the National University of Ireland Galway (NUI Galway) and University College Cork Medical Schools. Chapter 6 reports on stakeholder views of HPAT-Ireland. It comprises a qualitative study that describes the perceptions of doctors on HPAT-Ireland, their views on it job relatedness and acceptability as a selection tool (Study 2). Chapter 7 describes the feasibility of running a Multiple Mini Interview (MMI) in an Irish setting (Study 3). Chapter 8 explores by means of a mixed methods study, the use of MMI in an internationally diverse student population in terms of fairness with respect to age, gender, socioeconomic group,

and origin; predictive validity and acceptability to stakeholders (medical students from within and without the EU and MMI Assessors) (Study 4). Chapter 9 draws the findings of all these research papers together and discusses the conclusions and implications and suggestions for future research. A description of the dissemination of research findings, innovations arising, awards and fellowships earned during the course of the PhD are included in the Appendix.
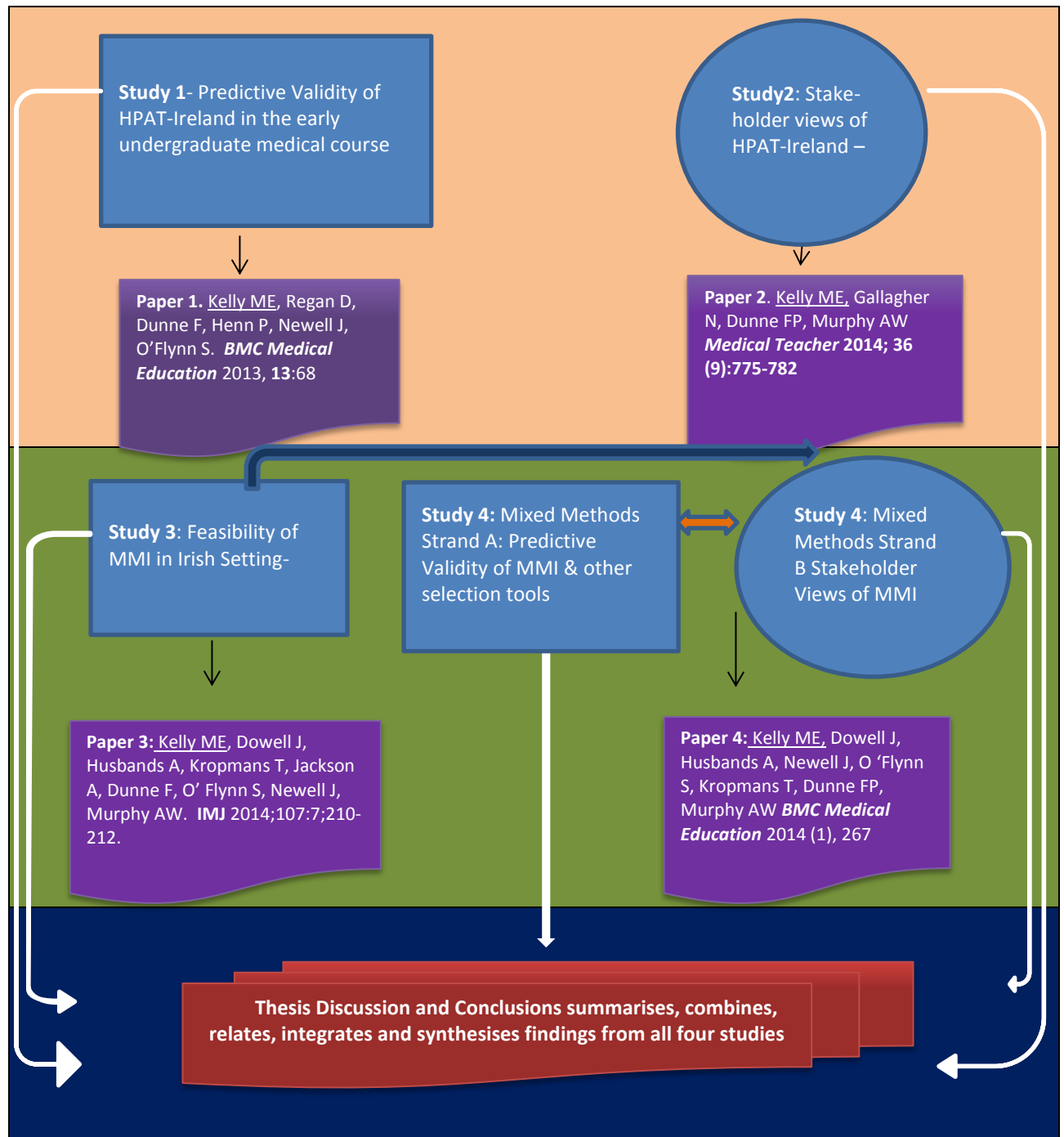
Figure 2 Multiphase Mixed Methods Research Design Diagram

Legend: **Phase 1: HPAT-Ireland Studies** background denotes Phase 1: HPAT-Ireland Studies. *Green background denotes* **Phase 2: MMI Studies**. *Quantitative studies/strands are shown in boxes. Qualitative studies/strands in circles. Dark blue arrow indicates that findings from Study 3 informed the methods of Study 4, Strand 2. Orange arrows indicate that findings from both strands of Study 4 were merged at the interpretation stage. Purple background denotes published papers. The navy background denotes* **PhD thesis discussion and conclusions** *-with the white arrows indicating that findings from all four studies were merged and integrated at this stage.*

Chapter 2: Review of Literature

Chapter 2: Review of Literature

2.1 Introduction

The purpose of this chapter is to appraise the international literature relevant to this thesis. It is presented in three discrete parts. It entails a review of the literature on aptitude tests and multiple mini interview in the selection of medical students. It describes their background, predictive validity, feasibility, reliability, issues of fairness and impact on diversity. The last section is an in-depth exploration and critical synthesis of the literature on stakeholder views of selection. This represents a more detailed systematic review (Moher et al., 2009). Unlike other aspects of selection, such as predictive validity, there has been no review of stakeholders' views published to date. The intention of this section is to address this gap in the literature. Stakeholders' views of aptitude tests and multiple mini interview are included in this section.

According to Jesson and Lacey (2006) a critical review of the literature demonstrates "*awareness of the current state of knowledge in the subject area; a synthesis of resources showing strengths and limitations, omissions and biases and how the research fits into this wider context*". Focused literature searches were conducted at key points throughout the study as evidenced by the background and conclusions sections of the four published papers. This chapter is based on a final literature search completed over six months between July and December 2014. Repetition of reviewed literature between this and subsequent chapters has been kept to a minimum. Medical student selection is a rapidly evolving research area, with a growing number of annual publications. It is important to note that some of insights and perspectives cited in this literature review were not available at the outset of this thesis.

Data searching was informed by the best evidence medical education guides for systematic searching of the literature (Haig and Dozier, 2003a, Haig and Dozier, 2003b). Critical review of identified articles was further informed by published guidelines (Greenhalgh, 2014, Haig and Dozier, 2003a, Haig and Dozier, 2003b, Jesson and Lacey, 2006, Sharma et al., 2015). Endnote Reference Manager Version 6 was utilised (Thomas Reuters http://endnote.com/) for the purposes of collating citations, abstracts and full text articles, categorising them

into groups to facilitate the critical review and to keep an audit trail. Separate searches were conducted for the different sections of the literature review.

The literature examining selection tools for medicine is extensive. This review focuses on current thinking and draws on papers published within the past 15 years. This is justifiable given that most of the aptitude tests used for the selection to medicine, and under review in this chapter, were first introduced within this time frame. Earlier work, relating to older aptitude tests, has been summarised by review documents during this time frame. Likewise, Multiple Mini Interview was first described in 2004; hence publications relating to it fall within this time scale also.

Five electronic databases: PubMed, SCOPUS, OVID Medline, PsycINFO and ERIC were searched for relevant published literature from January 2000 to December 2014. Terms were mapped to MESH terms or the appropriate term from the controlled thesaurus of the various databases. In addition text word searches were used for key words. An example search is included in Chapter 2 Appendix 1. Identified papers were imported into EndNote (http://endnote.com/) and duplicate papers were removed electronically and manually. Additional papers were identified from the reference lists of included papers and from my personal library which was built up throughout the course of this thesis.

2. 2 Aptitude Tests

2.2.1 Background

Aptitude tests measure the ability of an applicant to develop skills or acquire knowledge (Cleland et al., 2012). Amongst the most commonly used tests are the Medical College Admission Test (MCAT), the BioMedical Admissions Test (BMAT), the Graduate Medical School Admissions Test (GAMSAT), the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and the UK Clinical Aptitude Test (UKCAT) (Admissions Testing Service, 2014, Association of American Medical Colleges, 2014, Australian Council for Educational Research, 2014b, Australian Council for Educational Research, 2014c, UKCAT Consortium, 2014a).

Chapter 2: Review of Literature

The Health Professions Admission Test –Ireland (HPAT-Ireland) is one of the newest aptitude test to be used (Australian Council for Educational Research, 2014a).

Aptitude tests offer an opportunity to test some of the qualities thought to be of importance at the selection stage (Nicholson, 2005). In doing so, they broaden the scope of areas assessed beyond that tested by traditional measures of academic attainment. Theoretically at least then, aptitude tests may moderate some of the inequities associated with selection based on academic attainment alone and this is seen by many as one of the main justifications for their introduction and use (Tiffin et al., 2012). In general, these tools measure candidates' decision making, critical thinking, logical reasoning and processing skills - known as fluid intelligence (McManus et al., 2005, Patterson and Ferguson, 2010). Usually they do not have a specific curriculum and the emphasis is not on learned knowledge. However some also include a measure of scientific knowledge (for example the MCAT, GAMSAT and BMAT) termed crystallized intelligence (McManus et al., 2005, Patterson and Ferguson, 2010). Medical schools tend not to be directly responsible for the design and delivery of aptitude tests; rather they are contracted out to various testing agencies. Hence, they can be delivered without much resource implications for individual medical schools. Other advantages of aptitude tests are that they are standardized (Cleland et al., 2012) and can be taken by very large numbers of applicants annually (Dore et al., 2009). This makes them suitable for use very early in the selection process when the numbers of applicants are high.

The evidence supporting aptitude tests however, is conflicting and their use is controversial (Cleland et al., 2012, McManus et al., 2005, Siu and Reiter, 2009). This section will review the literature with respect to the six aptitude tests named above. It focuses on their predictive validity, as this dominates the published literature with respect to aptitude tests. Each tool will be described in turn and its predictive validity reviewed. MCAT, GAMSAT and BMAT are presented after each other as these three have similar test structure. Following this UKCAT, UMAT and HPAT-Ireland are presented. Then, the review will address the feasibility and

reliability of aptitude tests, followed by issues of fairness and impact on diversity. Stakeholder views of aptitude tests will be detailed later in this chapter.

2.2.2 Medical College Admission Test (MCAT)

The MCAT is the longest established, most widely employed and researched aptitude test used for selection to medicine. It is overseen by the Association of American Medical Colleges (AAMC) (2014). It has undergone four major revisions since being introduced in 1928 (Callahan et al., 2010). The principal motivations for its original development were the large attrition rate from medicine in the early 20th century (up to 50% in some institutions) and reforms to medical education subsequent to the Flexner report (Flexner, 1910, Mc Gaghie, 2002). In recent times, a key advantage of the MCAT is that it provides a standardised test sat by all applicants, which aids comparisons; unlike the grade point average, which is often dependant on extraneous factors such as the institution attended, course of study and choice of major subject (Davis et al., 2013, Kreiter and Axelson, 2013). MCAT is now used by virtually all medical schools in North America (Monroe et al., 2013).

Most of the recent published research relates to the post 1991 version (substantial changes are being introduced again in 2015) and it is this research that is considered here. This version of the MCAT had four sections: Verbal Reasoning, Biological Sciences, Physical Sciences and Writing Sample. According to the AAMC (2014) the Biological and Physical Sciences sections test applicants' introductory-level knowledge of biology/organic chemistry and physics/ general (inorganic) chemistry concepts respectively. Each of these sections has 52 multiple choice questions. In addition, they both measure data evaluation; quantitative reasoning skills; problem solving and critical thinking. The Verbal Reasoning section comprises 40 multiple-choice questions and tests "ability to understand, evaluate, and apply information and arguments presented in prose text". The now deleted Written Sample tested applicants' ability to develop a central idea, synthesise concepts and present arguments in a coherent and logical manner. Applicants had to write two 30 minute essays. These were scored by two trained reviewers and converted to an alphabetical scale ranging from a low of J to a high of T (Hojat et al., 2000). Each of the three multiple-choice sections is

scored 1–15. The total score is calculated as the sum of the three section scores and ranges from 3 to 45 (Davis et al., 2013). The writing sample was withdrawn due to evidence that its predictive validity was poor and reports from medical schools to indicate that it was only used in selection decisions for very few applicants (Association of American Medical Colleges, 2013 , Donnon et al., 2007).

Predictive Validity of MCAT

It is widely accepted that the MCAT has demonstrated acceptable predictive validity for future performance in medical school and licensing examinations (Albanese et al., 2003, Kreiter and Kreiter, 2007, Salvatori, 2001, Siu and Reiter, 2009, Veloski et al., 2000). The strength of this predictive validity has remained relatively stable over the history of MCAT and its various forms (Callahan et al., 2010).

MCAT has consistently displayed incremental validity over undergraduate total and science GPA. Julian (2005) followed up two cohorts from 14 medical schools and found that the best predictive model for cumulative medical school results was achieved by combining undergraduate GPA (uGPA) with MCAT. The correlation coefficients for uGPA was 0.4, for MCAT alone was 0.44, while for combined uGPA and MCAT it was 0.53 (all coefficients uncorrected for range restriction). Furthermore MCAT was a stronger predictor of USMLE Steps I, II, III results than uGPA. Little incremental validity was gained by combining the two measures (uGPA ranged 0.29 to 0.39, MCAT 0.49 to 0.61, combined uGPA and MCAT 0.52 to 0.64, all coefficients uncorrected for range restriction). A real strength of this study was that it was multi-institutional.

In one of the largest single MCAT studies, data on 840,000 candidates who sat the MCAT over a ten year period were analysed (Collin et al., 2009). Undergraduate GPA was conceptualised as measuring *general achievement*, MCAT as measuring *aptitude for medicine* and the authors postulated that both of these would explain significant variability in *competence in medicine* as measured by USMLEs (Steps 1-3). Latent variable path modelling confirmed this theory (comparative fit index 0.932, p<0.001). This type of statistical modelling is used to evaluate presumed casual relationships between variables (Columbia University,

2015). There were direct paths between general achievement, aptitude for medicine and competence in medicine, confirming their causal relationship. The model indicated that aptitude had incremental validity over achievement alone (path coefficient aptitude for medicine 0.55, p<0.001; path coefficient general achievement 0.22, p<0.001), explaining over 30% of the variance in competence in medicine.

In a meta-analysis of 23 peer reviewed publications, between 1991-2006, Donnon et al (2007) concluded that the strength of correlations between MCAT (including MCAT subsections) and both medical school assessments and licensing examinations ranged from small to medium.   The correlation between total MCAT scores and basic science/preclinical assessments in medical school was r=0.39 (adjusted for range restriction r=0.43) and r=0.34 for clinical clerkships (adjusted for range restriction r=0.39). The Biological Sciences section correlated most strongly with both the basic sciences (r=0.32, adjusted r=0.40) and the clinical clerkship grades (r=0.12, adjusted r= 0.15), while the Writing Sample was not a significant predictor of either. The correlation coefficients for the Physical Sciences and the Verbal Reasoning sections with pre-clinical grades range from r=0.00 to 0.29, which the authors deemed small, applying Cohen's classification. Likewise their correlations with clinical clerkships were small ranging from 0.06 to 0.14. This systematic review confirmed the findings of earlier papers that total MCAT correlated more strongly with the USMLEs. It had a large predictive validity coefficient effect size for the first step and medium for both steps 2 and 3 (Step 1: r= 0.60, adjusted r=0.66; Step 2: r=0.38, adjusted r=0.43; and Step 3 r=0.43, adjusted r=0.48). Again the Biological Science Section showed the stronger correlations (r ranged 0.11-0.48 unadjusted to 0.14-0.58 adjusted). Overall Total MCAT explains between 19-44% of the variance in the USMLE licensing examinations Steps 1, 2, 3 and between 15-19% of the variance in medical school assessments.

MCAT has also been shown to significantly predict the likelihood of a student having *unimpeded progress* through medical school (Dunleavy et al., 2013). Unimpeded progress was defined as not being dismissed or withdrawing from the course, passing each step of the USMLE examinations first time and graduating

within 5 years of matriculation. It has important implications regarding the curtailment of student debt, efficient use of educational resources and quicker transit time into the workforce. In their study of students in 119 medical schools, Dunleavy et al (2013) found that whereas undergraduate grade point average alone predicted unimpeded progress in 64% of schools, using MCAT and undergraduate GPA together increased this prediction to 90%. This finding was mirrored in a large five year follow up study (n>84,000) which found that lower MCAT scores were significantly independently associated with suboptimal outcomes for medical training (Andriole and Jeffe, 2010). For example, medical graduates who scored 18-20 on MCAT, compared to those who scored above 29 (out of possible maximum of 45) were more likely to withdraw or be dismissed from the course (adjusted odds ratio 11.08, p<0.001) and more likely to graduate without passing USMLE part I or Part II (adjusted odds ratio 13.06, p<0.001).

MCAT Verbal Reasoning has been shown to predict communication skills as measured by a validated OSCE in Part II of the Medical Council of Canada's Licencing Examinations (unadjusted r = 0.43, p<0.05) (Kulatunga Moruzi and Norman, 2002). Although this was a small study (n=97) in just one medical school. A separate larger study (n=597) found that Verbal Reasoning also significantly predicted clinical skills in the Canadian medical council licensing examinations (beta 2.41, p<0.001) (Violato and Donnon, 2005). MCAT and grade point average together accounted for over 11% of the variance in clinical skills score. Similarly Verbal Reasoning is the MCAT subsection most predictive of Step III of the USMLE – a stage of the licensing examination that is heavily weighted to test application of clinical reasoning skills (beta coefficient 0.28, p<0.05) (Veloski et al., 2000).

Taken together these are important finding as they confirm that it is possible for written aptitude tests – often considered to test predominately in the cognitive domain- to predict communication and clinical skills which are critical to future performance as doctors.

In summary, MCAT has a consistently good track record of small to medium correlations with medical school assessments, which provide meaningful incremental validity over academic record alone. It has an even stronger

relationship with medical licensing examinations, entailing medium to large correlations. Although it has a stronger relationship with the earlier steps of the exam, MCAT explains a significant proportion of the variance in both the knowledge and clinical steps of the exam.

2.2.3 Graduate Medical School Admissions Test (GAMSAT)

GAMSAT is a selection tool designed for graduate entry admission programmes and used in twenty five medical schools in Australia, Ireland, Poland and the UK (Australian Council for Educational Research, 2014b). It was introduced in 1995 (Australian Council for Educational Research, 2013a). Like UMAT and HPAT-Ireland it is designed and delivered by the Australian Council for Educational Research (2012a).

There are three sections: (1) reasoning in humanities and social sciences; (2) written communication and (3) reasoning in biological and physical sciences with a total test time of 5 ½ hours. Section 1 generally contains written passages relating to socio-cultural personal and interpersonal issues that require verbal processing, conceptual and logical thinking to answer. Section 2 contains two 30 minute writing tasks and is marked on the quality of the thinking and control of language used to convey it. Section 3 is a multiple choice paper assessing chemistry (40%), biology (40%) and physics (20%) (Australian Council for Educational Research, 2013). GAMSAT score is calculated on a 1:1:2 ratio (Puddey and Mercer, 2014). Used since the 1990's, there are surprisingly few peer reviewed publications on GAMSAT.

Predictive Validity of GAMSAT

The predictive validity of GAMSAT is contested and overall findings have been mixed. GAMSAT has a poor record with respect to predicting medical student outcomes on assessments of clinical abilities. In two separate studies, Groves et al (2003, 2007) used previously validated tools to assess medical students' ability in clinical reasoning and diagnostic thinking (Clinical Reasoning Problems and Diagnostic Thinking Inventory). The first study followed students (n=290) for two years and found no significant correlations (Groves et al., 2003). A follow-up study conducted in two Australian Medical Schools (n=189) again found no

significant correlations with total GAMSAT or GAMSAT subsections and clinical reasoning problems. This time, GAMSAT negatively predicted diagnostic thinking inventory in one of the institutions (r=-0.31, p<0.01). Likewise Quinlinvan et al (2010) found no relationship between GAMSAT and clinical assessments in their medical course (n=233). However none of the above studies provided any measure of the reliability of the clinical assessments, and this may have compromised the correlations.

Elsewhere Coates (2008) cites that Donnelly (2006) found overall no overall relationship between GAMSAT and competency based assessments (Donnelly, 2006). With respect to overall course assessments and knowledge based tests during medical school, the findings are somewhat more encouraging. Coates (2008) compared admission data on 351 students from six universities, with performance in first year examinations. Correlations between assessment outcomes and GAMSAT scores (total and section scores) were listed separately for over eighty combinations. Of these, 23 reached statistical significance. The majority involved GAMSAT Section 3 (unadjusted r values ranging from 0.24 to 0.57 average r=0.36). On regression, GAMSAT Section 3 was the only section of GAMSAT that made a significant contribution to the prediction of Year 1 grades (beta ranged 0.27-0.38, depending on the model). A shortcoming of this study was that First Year assessments were assigned an alphabetical code, but no information was provided about the type of course, knowledge versus clinical domains or the assessment measurement used.

A study of eight cohorts in one medical school (n=421) found that total GAMSAT correlated significantly with academic performance from Years 2 to 5 (of a 6 year programme), stronger in the earlier years and again largely attributable to the GAMSAT Section 3 (r ranged 0.12 to 0.28, p values generally <0.001) (Puddey and Mercer, 2014). Multiple regression revealed GAMSAT contributed significantly to the prediction of overall weighted average mark for the medical degree programme (beta 0.35, p <0.001). Overall, GAMSAT was seen to predict the knowledge based units more strongly than the clinical ones. However, these findings were challenged by Wilkinson et al (2008), who found that GAMSAT added very little unique contribution to the prediction of annual assessment

outcomes, only significantly contributing to the Year 1 and Year 4 written assessments (partial correlation coefficient 0.11, 0.14 respectively, p values <0.01).

Bodger et al (2011) examined, amongst other things, the predictive validity of GAMSAT in a graduate entry medical class in Wales. They followed up 102 students, from two cohorts, over the first two years of the medical degree programme. Factor analysis did not identify any significant relationship between GAMSAT scores and subsequent assessment performance. Like many studies though, this paper does not provide any measurement of the reliability of the in course assessments.

In summary, there is some evidence that GAMSAT contributes to the prediction of medical school grades, more so in the earlier years and this appears to be mostly attributable to GAMSAT Section 3. To date there does not appear to be evidence that GAMSAT predicts clinical outcomes.

2.2.4 Biomedical Admissions Test (BMAT)

The BMAT is an aptitude test designed and administered by the Admission Testing Service, which is part of Cambridge Assessment (Admissions Testing Service, 2014). It is used by seven medical schools in the UK, Singapore and the Netherlands. Individual medical schools vary in the extent to which it contributes to the selection process. The BMAT came into use in 2003 and was derived from a previously used version known as the Medical and Veterinary Admissions Test (MVAT) (Emery and Bell, 2009).

It lasts two hours and consists of three sections. Section 1 measures aptitude and skills in problem solving, data and graphical interpretation and inference and Section 2 measure scientific knowledge and applications relating to biology, chemistry, physics and maths. Both of these sections are either multiple choice or numerical answers and are marked objectively and reported as a numeric score on a scale of 1-9 (Emery et al., 2011). In Section 3 applicants choose from a menu of three short essay questions, and are assessed on the clarity or argument and written communication skills (Emery and Bell, 2009). Scores on the essay are

marked holistically and scored on a scale of 0-15. Scores on the individual sections are not intended to be combined (Emery et al., 2011).

Predictive Validity of BMAT

The predictive validity of BMAT has been the subject of controversy and debate in the peer reviewed literature. Emery and Bell (2009) reported that the BMAT Section scores of four cohorts of students (average n per cohort = 251) correlated with Years 1 and 2 examination outcomes to a significant degree (uncorrected r values ranged from 0.13 to 0.26, p<0.05), with Section 2, Scientific knowledge, correlating more strongly. The latter section also predicted the likelihood of achieving a first class honours grade (odds ratio First Year 1.7, Second Year 1.5, p<0.05). However, the findings in this paper were sharply criticised by McManus et al (2011a) for lacking statistical rigour and presenting insufficient data particularly with respect to reliability of BMAT. There have been further calls for more transparency in the reporting of BMAT performance (Ramachandran and Venkatesh, 2014).

In summary, there is very little peer–reviewed evidence regarding BMAT, with one study reporting that it significantly correlated with medical school assessments in the first two years of the course.

2.2.5 UK-Clinical Aptitude Test (UKCAT)

The UKCAT was introduced in 2006 to the selection process for medical and dental students in the UK (UKCAT Consortium, 2007a). It is developed and administered by Pearson VUE in conjunction with a group of medical and dental schools known as the UKCAT Consortium (Pearson VUE, 2014, UKCAT Consortium, 2007a). It comprises four cognitive sections; verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis that are tested by a multiple choice format. A non-cognitive section was included in 2007 and was trialled for a number of years then replaced with a situational judgement test (SJT) in 2012. The SJT was included as a live part of the examination only from 2013. As a result, at the time of this literature review, data regarding this section had not appeared in peer reviewed publications with respect to UKCAT.

According to Lynch et al (2009)   "*Verbal reasoning is designed to test ability to think logically about written information and to arrive at a reasoned conclusion; Quantitative Reasoning assesses the ability to solve numerical problems; Abstract Reasoning tests the ability to infer relationships from information by convergent and divergent thinking, and Decision Analysis measures the ability to deal with various forms of information, to infer relationships, to make informed judgements and to decide on an appropriate response in situations of complexity and ambiguity*".

The main motivations for introducing the UKCAT were to provide an objective and fair means of distinguishing between high achieving applicants, to redress some of the socioeconomic biases associated with selection based on academic record and to support widening diversity (Adam et al., 2011, McManus et al., 2013c, UKCAT Consortium, 2007a).   A further aim was to establish the predictive validity of these tools (Adam et al., 2011). The amount to which UKCAT contributes to the selection process is at the discretion of individual medical schools. It has been used variously to determine offers of interviews or places to borderline applicants only, as a threshold whereby applicants must attain a certain UKCAT score to proceed to next stage of selection, as a rescue mechanism for applicants who can then use it to compensate for lower scores on some other selection tool or as a factor whereby every applicant's UKCAT score contributes to their overall ranking (Adam et al., 2011).

Predictive Validity of UKCAT

Reports on the predictive validity of UKCAT are emerging as successive cohorts selected since its introduction are making their way through the medical degree programme and on to graduation (see Table 3 for summary).   Early papers explored the relationship between students' performance in the initial years of the undergraduate course and scores on selection tools including UKCAT.

Lynch et al (2009) reported on First Year medical students from Aberdeen and Dundee (n=341) and found no significant correlations between scores on UKCAT, UCAS and interview and subsequent performance on either knowledge tests or OSCE performance. Conversely Wright and Bradley (2010) found that UKCAT total

score was predictive of student performance in knowledge based examinations in Years 1 and 2 of the programme (beta range from 0.01 to 0.02, all p values <0.05), but not for OSCE assessments. This paper did not report a correlation coefficient hence direct comparisons with other studies cannot be made.

A study in Hull York Medical school (n=146) found that UKCAT total scores significantly correlated with overall assessment outcomes in Years 1 and 2 (r=0.28, 0.32 p<0.001), but not with OSCE scores (Adam et al., 2012). Additionally every UKCAT subsection score correlated significantly with overall assessment outcomes in at least one of the two years (r ranged 0.21 to 0.29, p<0.05 to <0.001). UKCAT only weakly correlated with student professionalism as measured by tutor formative assessments of students' professional and interpersonal skills, (using a modified peer assessment of professionalism inventory). Limitations of this study included that multiple subsections of the tutor assessment scales were included individually, without adjusting for multiple statistical testing (risking a type 1 error) and there was no regression analysis presented.

A two year follow-up study from Nottingham (n=204) Yates and James (2010), reported significant correlations between total UKCAT and specific modules in Years 1 and 2. Both total UKCAT and its subsections, in particular Verbal Reasoning, correlated significantly with outcomes on a number of assessments (r ranged between 0.19 and 0.31, p<0.01), although none significantly correlated with OSCE performance. Regression analysis demonstrated that UKCAT total contributed a small amount to prediction of module outcomes, largely due to Verbal Reasoning Section (beta ranged 0.17 to 0.28, p values < 0.05). The authors were unconvinced that UKCAT was of any additional value in the selection process. However, they did not compare the performance of UKCAT with any of the other selection tools employed in that institution for example the interview or personal statement hence estimates of incremental validity were not actually possible.

Husbands and Dowell (2013) followed up two cohorts of medical students in one institution and found that UKCAT correlated with Year 1 Semester 1 written and OSCE assessments for the 2009 cohort (r=0.25, p<0.01, r=0.18, p<0.03

respectively), but not with Year 1 Semester 2 assessments, nor any Year 2 assessment for this cohort. Regression analysis demonstrated that UKCAT explained 6% of the variance in Year 1 Semester 1 written (r^2=0.06, beta 0.25, p=0.004) and UKCAT and MMI together explained 7% of the variance in Year 1 Semester 1 OSCE. (r^2=0.07, beta 0.18, p=0.004). No significant correlations at all were found for the 2010 cohort.

A publication by McManus et al (2013c) followed up three cohorts of students in twelve UK medical schools (n=4811) and established the predictive validity of UKCAT for First Year assessments. This was a milestone publication because, in contrast with the previous publications it was an amalgamation of twelve large medical schools sufficiently varied to be generally representative of the larger pool of medical schools, it was adequately powered to detect important predictor/predicted relationships, it included and adjusted for exhaustive details on educational attainment and demographic background factors and provided detailed explanation and justification for statistical methods employed throughout the study. The principle findings were that UKCAT correlated to a small but highly significant degree with overall Year 1 score (r=0.15, p<0.001), this was somewhat stronger for mature students (r=0.25, p<0.001) and that it correlated more strongly for theory over skills based assessments (r 0.16 and 0.08, all p values <0.001 respectively). UKCAT total score provided incremental validity over educational attainment for the prediction of overall First Year score but to a very small degree (beta 0.057). All UKCAT subsections correlated significantly with overall mark and to a greater degree with theory as compared to skills marks (see Table 3).

Table 3 Summary Table Predictive Validity UKCAT

Reference	Predicted Variable / Number of students (n)	Correlation Coefficients	Regression Analysis
Adams et al 2012	Yrs 1&2 Tutor assessments of students' interpersonal skills & professional behaviours. Annual total examination scores & categorised into recall & knowledge, communication &clinical skills (with OSCE), data evaluation (n=146)	UKCAT total scores significantly predicted overall assessment scores in Yr 1 & 2 (r=0.28, 0.32 p<0.001), and predicted data evaluation the best (r 0.39 (yr 1) , 0.34 (Yr2) p<0.001), but did not significantly predict OSCE scores. All UKCAT subsections predicted data evaluation in both years of study (r ranged 0.18 to 0.39, p<0.05) with VR strongest. No UKCAT subsection correlated with communication OSCE but DA was significant in Yr 1 for practical skills (r= 0.20 p<0.05). Only one significant correlation between UKCAT total and any tutor assessment (Yr 2 "contributes work for group" r=0.19, p<0.05). UKCAT AR correlated with four tutor assessment items (r ranged 0.18 to 0.25, p<0.05), DA with two (r=.20, .22, P<0.05) and QR and VR negatively with one item each (r -.19, -.22, p<0.05)	Not reported
Husbands & Dowell 2013	First and Second Year, Two cohorts (n=140, 150). Written assessments, OSCE scores in Semester 1 and 2. Two year follow up	2009 cohort: UKCAT total score significantly correlated with Year 1 Semester 1 written (r=0.25, p<0.01) and OSCE (r=0.18, p<0.03), but not with Year 1 Semester 2 and not with any Year 2 assessment. There were no significant correlations between UKCAT and Year 1 assessments in the 2010 cohort.	In the 2009 cohort UKCAT explained 6% of the variance in Year 1 Semester 1 written (r^2=0.06, beta 0.25, p=0.004) and UKCAT and MMI together explained 7% of the variance in Year 1 Semester 1 OSCE. (r^2=0.07, beta 0.18, p=0.004)
Husbands et al 2014	Senior Medical Years , 2 schools (n=154) Written assessment OSCE scores	In Aberdeen UKCAT correlated significantly with the Yr 4 Written – r=0.24 p<0.05, Yr 4 OSCE r=0.36, P<0.01, Yr 5 OSCE r=0.29, p<0.01 In Dundee the UKCAT correlated with Yr 4 Written r=0.34, p<0.01 but not OSCE	Aberdeen: UKCAT contributed significantly to the prediction of Yr 4 written (beta 0.24, p=0.02), Yr 4 OSCE (beta 0.36, p=0.01), Yr 5 OSCE (beta 0.29, p=0.01). In Dundee UKCAT contributed significantly to the prediction of Yr 4 written (beta 0.34, p=0.02)
Lynch et al 2009	First year knowledge and OSCE scores (n=341)	UKCAT total NS / UKCAT subsection scores NS	Not undertaken as no significant correlations
McManus et al 2013 (c)	First Year Medical School Performance Measure (pass 1st attempt/ pass after repeat/ repeat yr/ dropout), overall mark & Theory mark, skills mark. Three cohorts of students attending 12 UK medical schools (n=4811)	Total UKCAT correlated significantly with overall First Year Mark r=0.15, p<0.001, with skills mark (r=0.08, p<0.001) and theory mark (r=0.16, p<0.001). All Subsections correlated significantly with overall mark: VR (r=0.12), QR (r=.08) DC (r=.09), AR (r=.08) all p values <0.001. Subsection correlated stronger with the theory than the skills mark (r theory: 0.05 to 0.18, r skills 0.03 to 0.06) and were significant in all but one case (VR did not significantly correlate with skills mark).	UKCAT total score provided incremental validity over educational attainment for the prediction of overall First Year score but to a very small degree (beta 0.057).

Key: NS: Not significant, VR: Verbal reasoning, Q: Quantitative reasoning, AR: Abstract reasoning, DA: Decision analysis, Yr(s):

Table 3 Continued

Reference	Predicted Variable / Number of students (n	Correlation Coefficients	Regression Analysis
McManus et al 2013 (a)		Construct level predictive validity of UKCAT (and another aptitude measure AH5) was 0.181	
Sartania et al (2014)	Full Medical Programme Educational performance measure score end of Yr4 Honours/ commendation end of Year 5 (final ranking) (n=189)	Not reported	UKCAT total score made a significant unique contribution to the prediction of variance in student outcomes for both the Educational performance measure (beta 0.21, p<0.005) and honours and commendation (beta .25 p<0.001) and was the only admission tool to be independently predictive of both outcomes.
Wright & Bradley 2010	First 2 years of course (two cohorts n=307) knowledge scores, OSCE scores	Not reported	UKCAT total contributed significantly to the prediction of 7 out of the 8 Assessments included in the model (beta ranged 0.01 to 0.02, all p values <0.05) (used stepwise regression- personal statement & interview were not deemed statically relevant and not entered into model)
Yates & James 2010	First two years results in 4 Modules : The cell/ the person/ the community/ personal & professional development/OSCE(n=204)	Total UKCAT and the Cell r=.21 (p= 0.003) Total UKCAT and the community r=.23 (p= 0.001) UKCAT VR and the cell r=.19 (p=0.008) and the community r= .32, p<0.001 UKCAT QR and the cell r=.24 (p<0.001). NS correlation with OSCE	Total UKCAT and the community beta 0.18, p=0.012 and the cell (beta =0.17, p=0.03). UKCAT VR and the cell (beta 0.19 p=0.017), the person (beta 0.20, p=0.012), and the community (beta 0.28 p<0.001). UMCAT QR and the cell (beta 0.22, p<0.005) UKCAT DA and personal professional development (beta -.21 p=0.008)
Yates & James 2013 (follow up of 2010 study)	In preclinical phase (First 2.5yrs) : average annual mark + overall preclinical average Clinical phase (Senior 2.5 yrs) knowledge, skills and weighted average annual mark (n=204)	In the preclinical yrs only UKCAT VR correlated overall weighted average (r=0.18, p<0.01) UKCAT total correlated with 7 out of the 9 assessment outcomes in the clinical phase (r ranged 0.17 to 0.26, p's <0.01) UKCAT VR correlated significantly with every assessment outcome in clinical phase except one (r ranged from 0.19 to 0.28, all p values <0.01) UKCAT QR correlated significantly with overall final year average, final year knowledge and first clinical yr knowledge (r's between 0.17 to .20, p<0.01). UKCAT AR,UKCAT DA= NS at any stage	UKCAT total contributed significantly to the prediction of knowledge in the clinical; years (beta 0.18 to 0.21, p<0.05), one assessment of clinical skills (phase 2 beta 0.18, p=0.02) weighted average mark phase 2 (beta 0.2, p=0.008). UKCAT VR contributed to a small but significant degree to the prediction of knowledge in Yrs 4 and 5 (beta 0.23, 0.21 both p's <0.01) and weighted average marks (beta 0.15 to 0.20, p<0.05), UKCAT QR predicted knowledge Yr 5 (beta 0.19, p=0.014). But once student's prior learning was entered into the model the effect of UKCAT total or subsection scores were no longer significant..

Key: NS: Not significant, VR: Verbal reasoning, Q: Quantitative reasoning, AR: Abstract reasoning, DA: Decision analysis, Yr(s): Year(s)

The more recent publications have followed up students throughout their entire undergraduate degree. In a follow up to their earlier paper, Yates and James (2013) found that UKCAT correlated more in the clinical years than in the earlier preclinical phase (five years follow up, n=204). Total UKCAT (r ranged 0.17 to 0.26, p values <0.01) and Verbal Reasoning Section (r ranged from 0.19 to 0.28, all p values <0.01) both correlated significantly with assessment outcomes in clinical phase (see Table 3). In an interesting move, the authors included outcomes of learning throughout the medical course as a predictor of future assessment performance, within the hierarchal regression model. Once this was done it became evident that prior learning was extremely influential (beta ranged 0.16 to 0.61, almost all p values <0.001) removing any significance of UKCAT to the model. Again the authors conclude that UKCAT had little to offer in terms of independent prediction of medical school assessments.

Sartania et al (2014) reached the opposite conclusion in their follow-up of medical students in Glasgow (n=189). Their principal outcome measures were a) Educational Performance Measure and b) Honours and Commendation based on aggregate exam performance to the end of Year 4 and students' final medical degree ranking respectively. UKCAT total score made a significant unique contribution to the prediction of variance in student outcomes for both of these measures (beta 0.21, p<0.005 and beta .25 p<0.001, respectively) and was the only admission tool to be independently predictive of final student ranking. Even when the regression model was altered to adjust for socio-demographic variables (gender, age, ethnicity, deprivation) and mutually adjusted for the other admission tools UKCAT maintained its significant independent prediction of assessment outcomes explaining approximately 6% of the variance in performance. Similarly Husbands et al (2014b) report that UKCAT was correlated significantly with written assessments in the senior clinical years in both Aberdeen and Dundee medical schools (r=0.24 p<0.05, r=0.34, p<0.01) and Year 4 and 5 OSCEs in the former (r=0.36, r=0.29, both p's <0.01). While neither interview nor UCAS correlated with any outcome measure in either school. On regression UKCAT score explained 6-13% of the variance assessment performance in Aberdeen and 11% of the variance in Dundee.

In summary, there is conflicting evidence for the predictive validity of UKCAT from a number of small studies, with short follow up. However the results from a large high quality study of 12 medical schools provides good evidence that UKCAT significantly predicts student performance, at least in Year 1 of the medical degree programme. The level of this prediction is small, explaining about 2% of the variance in Year 1 assessments. The Verbal Reasoning section appears to have the best correlation. There is also emerging evidence that UKCAT may become more predictive in the senior clinical years, contributing between 6-13% to the explanation of variance in the overall assessment outcomes.

2.2.6 Undergraduate Medicine and Health Sciences Admission Test (UMAT)

The UMAT is an admission test designed and administered by the Australian Council for Educational Research (ACER) (2014c) on behalf of the UMAT Consortium universities. It has been used to varying degrees for selection to medicine in Australia and New Zealand since the late 1990's (Mercer and Puddey, 2011). Currently it is required for selection to 11 medical schools. It has three sections. UMAT Section 1 measures Logical Reasoning and Problem Solving, Section 2 Understanding People and Section 3 Non-verbal Reasoning. The exam lasts approximately three hours; questions are of a multiple choice format with a single best answer.   According to ACER (2014c) each section tests the following: Section 1 presents a passage of text or graphical information. It is designed to assess comprehension, data evaluation, ability to draw conclusions and generate hypotheses. Section 2 tests the ability to understand people, infer their thoughts and feelings, and identify intentions. It presents a dialogue or passage of text describing personal situations. Section 3 is based on patterns and sequences of shapes and tests ability to problem solve in a non-verbal context.   The result from each of the three sections is averaged to provide a total UMAT score (Griffin et al., 2008). The UMAT is designed for applicants to undergraduate medical courses and the majority of test takers are either school leavers or students studying first year health sciences in university.

Predictive Validity UMAT

The predictive validity of UMAT is unclear with conflicting findings being reported in the peer reviewed literature (see Table 4 for summary). There is some

evidence that the overall UMAT score is predictive of student performance in the early years of the medical course as measured by their grade point average based on combined assessment scores. Wilkinson et al (2011) report a correlation coefficient of r=0.15 (p<0.005) between overall UMAT and mean cumulative GPA earned during the first four years of the medical degree programme, in University of Queensland (n=339).

But when this was examined year on year the relationship only was significant in Year 1 of the programme (r=0.24, p<0.005). Edwards et al (2013) compare findings across three anonymised institutions (n=650). They report that total UMAT correlated significantly with medical school assessments in each of the first three years in Institution A, in two out of the first two years in Institution B and only in the first year in Institution C (r ranged from 0.19 to 0.41, average significant r =0.32, all p values <0.05, corrected for restriction of range). UMAT contributed some unique explanation of variability in outcome measures but this differed substantially across the three institutions, and across years (beta coefficients ranged from: UMAT Section 1 (-0.02 to 0.25), UMAT Section 2 (0.00 to 0.25) UMAT Section 3 (0.03 to -0.09)). Simpson et al (2014) also found that total UMAT significantly correlated with overall assessment score in the first two years of study in their medical school (r=0.15, p<0.01, 2 cohorts of students, n=318). The significant correlation lay with knowledge, as opposed to clinical assessments (r= 0.12, p<0.05).

In a national study, Poole at al (2012b) compared findings from the two medical schools in New Zealand (n=1346). They found that in Auckland UMAT explained a small amount of the variance in each year of medical school averaging 5%. Correspondingly, in Otago Medical School UMAT explained 0.4% in Year 1 rising to 4% in Year 6. A combination of GPA and UMAT provided a slightly better explanation of variance than either alone, and this appeared to be more meaningful in the senior years.

However, these findings are challenged by others. Mercer and Puddey (2011), in an eleven year follow-up study of 1174 students report that total UMAT did not significantly correlate with end of year average marks at any stage throughout the six years of the medical degree programme.

Table 4 Summary of Predictive Validity Findings for UMAT

Reference	Predicted Variable / Number of students (n)	Significant Correlation Coefficients	Regression Analysis
Edwards et al 2013	Students' annual grade point average for the early years in medical school in three separate institutions: **A** Years 1-3, **B** Years 1-4, **C** Years 1-2 (n=650)	**Total UMAT** correlated with medical school assessments every year in Institution A, in the first two years in Institution B and in the first year in Institution C (r ranged from 0.19 to 0.41) **UMAT Section 1** correlated significantly in each of the three schools, for at least one year of the programme, (r ranged from 0.16 to 0.48) **UMAT Section 2:** correlated significantly in Institution A and B at least in two years (r ranged 0.17 to 0.43) **UMAT Section 3** correlated in Institution C in Year 2 (r=0.18), but not in the others. (All r's corrected for restriction of range, all p values <0.05)	**Section 1** contributed significantly to the amount of explained variance in assessment outcomes only in Institution A but across all 3 years (beta 0.21 to 0.25, p<0.05).. **Section 2** contributed significantly in Institution A and B, but only in the first year in the latter (beta 0.14 to 0.25, p<0.01). **Section 3** contributed significantly but only negatively in Yr 2 Institution A and in 3 out of the 4 yrs in Institution B (beta all negative -0.11 to -0.19, p<0.05).
Mercer & Puddey 2011	Annual weighted average exam scores. - Years 1-6. Scores on Knowledge based and Clinically based units. (n=1174)	**UMAT total** NS for any assessment outcome. **UMAT Section** 1 correlated with Year 5 (r=0.09, p<0.05) **UMAT Section 2** – NS **UMAT Section 3** with Year 5 (r=-.09, P<0.05)	**UMAT section** scores NS for annual overall scores. **UMAT Section 1** was predictive of knowledge in Years 4-6 and Clinical skills in Years 4-6 (beta 0.10, 0.08, p<0.05). **UMAT Section 3** negatively predictive of clinical skills Yr 1-3 (beta -.08, p=0.004). **UMAT Section 2** NS
Poole et al 2012	Assessments Yrs2-6 in two institutions. (n=1346)	Not reported	In Auckland **UMAT average total score** explained 5% in assessment outcomes Year 2 rising to 8% in Year 5 (10% in Year 5 written, 3% of the Year5 clinical) and dropping to 6% of the 6th yr. In Otago UMAT explained 0.4% in Year 2, rose to 6% in Year 2 settling at 4% in Year 6.
Shulruf et al 2012	Assessment outcomes Years 2-4 (n324)	Not reported	**UMAT Sections 1 and 3** predicted Year 2 assessments (both Bs=0.03, p<0.01). **UMAT Section 2** NS
Simpson et al 2014	Average mark for Years 1&2, 3&4, 5&6 and overall programme. Clinical Skills/ Knowledge based units (n=318)	**UMAT Total Score** correlated significantly with average mark for Year 1&2 (r=0.15, p<0.05), knowledge Years 1&2(r=0.12, p=0.03) and Clinical skills Years 5&6 (r=-0.12, p=0.04).	**UMAT subsection scores** did not significantly contribute to any outcome measure
Wilkinson et al 2011	Mean GPA during first four years of medical school and GPA each year (cumulative) (n=339)	**Mean overall UMAT** correlated with mean GPA (r=0.15, p=0.005). **UMAT Section 1** correlated with overall mean GPA r=0.14 P<0.01 and Year 1 GPA (r=0.24, p<0.005). No other Section of UMAT was significantly correlated with either mean overall GPA or GPA in any year in the course	**UMAT Section 1** significantly correlated with mean GPA (B= 0.01, p < 0.05) and Year 1 GPA (B = 0.02, p<0.005). **UMAT Section 2, 3**, NS

Key NS: Not significant

UMAT sub-section scores correlated weakly but significantly in just two instances: UMAT Section 1 with Year 1 assessments (r= .09, p<0.05) and UMAT Section 3 correlated negatively with Year 5 assessments (r=-0.09, p<0.05). On regression, subsection scores were not found to significantly contribute to the variance in any end of year overall assessment. Shulruf et al (2012), in their study of three cohorts of students in one medical school (n=324), report that UMAT Sections 1 and 3 significantly predict assessment outcomes in Year 2 (B= 0.03, 0.03 respectively, p<0.01) but not in the years 3 or 4. It was notable that in all the above studies, previous academic record was shown to be a very strong predictor of performance in medical school. Even when UMAT and its subsections were significant the contribution was dwarfed in comparison with academic record.

In summary, the predictive validity of UMAT overall is contested. The evidence shows that it contributes to the explanation of variability in student assessment scores to a limited degree. UMAT Section 1 has the most consistent record, with most studies finding that it correlates with at least one outcome measure, commonly knowledge based assessments (Edwards et al., 2013, Mercer and Puddey, 2011, Poole et al., 2012b, Shulruf et al., 2012, Wilkinson et al., 2011), but also some clinical assessments (Mercer and Puddey, 2011, Poole et al., 2012b). By contrast UMAT Section 2 does not appear to demonstrate significant correlation with assessment scores (Mercer and Puddey, 2011, Shulruf et al., 2012, Simpson et al., 2014, Wilkinson et al., 2011). Likewise, UMAT Section 3 has a mixed record, with studies varying from small positive (Edwards et al., 2013, Shulruf et al., 2012) to negative correlations (Mercer and Puddey, 2011). The largest UMAT study (Poole et al 2012) however suggests that even though the correlation is limited, it does add incremental validity over academic record alone and suggests that this may become more meaningful in the senior clinical years.

2.2.7 Health Professions Admission Test (HPAT)–Ireland

The main impetus for the introduction of HPAT-Ireland was the publication of a Government initiated report which recommended that medical student selection, in Ireland, should no longer be based on academic grades alone (Fottrell, 2006). This is further discussed in the next chapter.

Chapter 2: Review of Literature

Like GAMSAT and UMAT, HPAT-Ireland is designed and independently delivered by the Australian Council for Educational Research (Australian Council for Educational Research, 2014a). ACER also designs HPAT-Ulster, a test which was recently introduced for selection to the allied health professions in University of Ulster, Northern Ireland. As it is not used for selection to medicine it will not be considered further in this thesis.

HPAT-Ireland is a multiple choice test. In terms of intelligences tested it largely focuses on fluid intelligence.   Like UMAT there are three sections, which are described by ACER as follows (Australian Council for Educational Research, 2014a): Section 1: Logical reasoning and problem solving consists of 44 multiple choice questions based on a passage of text or a diagram. Applicants are required to analyse and logically reason through the information presented. Section 2: Interpersonal Understanding consists of 36 multiple choice questions based on a scenario representing specific interpersonal situations. Applicants have to identify, understand, and infer the thoughts, feelings and behaviours of the people represented in the situations. Section 3: Non-Verbal Reasoning consists of 30 multiple choice questions based on recognition of patterns and sequences of shapes, to test applicants' abstract and non-verbal reasoning.

At the outset of this thesis, there were two peer reviewed publications and one editorial published on HPAT-Ireland. Quinn et al (2010) ran a mini version of HPAT-Ireland and recorded no significant difference in scores between groups of consultant surgeons (n= 9), non-consultant hospital doctors, undergraduate medical students (n= 105), 1st Year Graduate Entry Programme students(n=60) and Final Year Graduate Entry Programme students (n=18). Halpenny et al (2010) looked at the validity HPAT-Ireland, by administering a mini version of the test to a group comprising 27 established doctors and 29 final year medical students. They report a significant correlation between candidates' medical school results and their HPAT score (r: 0.314, p = 0.018, Spearman Rank). However, findings from both of these studies cannot be generalised as in both cases they used a modified scaled down version of HPAT-Ireland utilising just twelve and twenty-two questions respectively rather than the full test. However, they were important as they highlighted the need for regular and careful research to evaluate the utility of HPAT-Ireland.

2.2.8 Summary of Predictive Validity of Aptitude Tests

The evidence for predictive validity of aptitude tests varies according to the version of the tool used. The MCAT has the longest and most consistent record of moderately good predictive validity. It explains between 15-19% of the variance in medical school assessments and 19-44% of the variance for Steps I-III of the USMLE. There is good evidence that UKCAT predicts first year medical school assessments to a very small but significant degree, explaining 2% of the variance. Emerging evidence from a small number of other studies suggests that it may explain between 6-13% of the variance in the senior medical school years. The evidence with respect to GAMSAT, BMAT and UMAT is limited and conflicting. For the most part the predictive validity of these tests is small, occurs early in the medical degree programme and favours knowledge based assessments over clinical ones. At the outset of this thesis, there was no evidence of the predictive validity of HPAT-Ireland for medical students selected by this tool.

2.2.9 Feasibility of Aptitude Tests

One of the main advantages of aptitude tests is the relative ease with which they can be delivered to large groups. The feasibility of aptitude tests is rarely mentioned in the peer reviewed literature. Most likely, this reflects the fact that unlike some other selection tools, such as interviews, letters of reference or personal statements, medical schools rarely have direct responsibility for delivering or scoring aptitude tests. As the delivery and scoring of aptitude test is outsourced to assessment agencies, the issue of feasibility does not arise for medical schools at large. Feasibility issues include the numbers of applicants the test can accommodate, the number of days on which it is administered annually, the number and location of suitable venues, the management and administration of the test centres, appointments, student registration, student identification, marking of assessments, results databases, release of results, handling re-checks, issuing of results and liaising with applicants, medical schools and university admissions.

The UKCAT Consortium is an exception in that they have reported on the feasibility and cost effectiveness of UKCAT annually since it was first used in 2006. The UKCAT is taken by an average of 20,000 applicants annually (ranging from 18, 542 to 25, 679) (UKCAT Consortium, 2007b, UKCAT Consortium, 2014a). It is

delivered in 150 local testing centres in the UK, and is available in 65 other countries (UKCAT Consortium, 2007b). The test is held on numerous occasions between July and October annually. For security reasons a number of different equivalent versions of the test are available in any one academic year. Minor feasibility issues included confusion regarding the use of the on line calculator (UKCAT Consortium, 2007b, UKCAT Consortium, 2012a), delays in reporting results to medical schools (UKCAT Consortium, 2010, UKCAT Consortium, 2012a), and late registration or difficulties experienced by applicants in booking a suitable appointment (UKCAT Consortium, 2007b, UKCAT Consortium, 2013). More serious issues include attempted breaches of test security (UKCAT Consortium, 2014a), item marking errors (UKCAT Consortium, 2008) and candidate identification number errors (UKCAT Consortium, 2010). The costs related to UKCAT are borne predominately by the applicant, although the medical schools did invest in the set-up of the test (UKCAT Consortium, 2007a). Approximately 5% of applicants avail of a bursary to cover the costs, on grounds of financial hardship. In summary the feasibility UKCAT is well established, reviewed annually and reported to the public.

Information on the feasibility of GAMSAT, UMAT and HPAT-Ireland, detailed here, was provided on foot of my request to ACER (Vale, 2015). GAMSAT is held twice per year, once in March and once in September. Approximately 13000 candidates sit the test each year. It is held in a number of cities throughout Ireland, Australia, New Zealand, UK, Singapore and USA. UMAT is offered in 26 centres throughout Australia, New Zealand and also in London, Singapore and Washington DC. Approximately 14000 applicants take this test annually. While HPAT –Ireland is taken by 2600 – 2800 applicants each year. The test is administered on one day per year in five test centres through-out the country.

In summary, there is little published data in the medical education literature describing the feasibility of designing, developing, administering and financing aptitude tests for selection to medicine. The exception is UKCAT which makes its annual reports, addressing such issues, freely available on the internet.

2.2.10 Reliability of Aptitude Tests

Like feasibility, the reliability of the six aptitude tests under consideration, is rarely discussed in the peer reviewed literature. This is a limitation of publications in the field as conclusions regarding predictive validity assume acceptable levels of test reliability, which may or may not be the case. The likely cause is that data regarding test reliability lies with the test developers and may not be publically available. This may reflect concerns regarding potential legal challenges from unhappy applicants. The widespread introduction and use of aptitude tests without adequate published evidence of their reliability has been criticised (McManus et al., 2005).   There have been calls for more transparency in making reliability data openly available (McManus et al., 2011b).

Threats to the reliability of aptitude tests include short test time, inadequate numbers of questions, irrelevant, ambiguous or poorly worded questions (Cook and Beckman, 2006, Downing, 2004). Davis et al (2013) provide insight into the steps required to develop reliable questions. They describe the rigor adopted by MCAT developers in an effort to ensure that their test items do not contain hidden biases against different ethnic groups. MCAT writers, reviewers and editors adhere to strict guidelines to check that each test item is construct relevant. Items are then subjected to a bias and sensitivity review by a diverse panel. Following this questions are piloted as non-scoring items and applicant responses analysed before they are finally included in the live examination.

The UKCAT Consortium openly publishes test reliability data each year in the annual and technical reports (UKCAT Consortium, 2007a, UKCAT Consortium, 2007b, UKCAT Consortium, 2008, UKCAT Consortium, 2010, UKCAT Consortium, 2011, UKCAT Consortium, 2012a, UKCAT Consortium, 2013, UKCAT Consortium, 2014a). Cronbach alpha's for overall test reliability range from 0.83-0.92 Three subsections have consistently acceptable levels of reliability (Verbal Reasoning 0.64-0.74, Quantitative reasoning 0.60-0.76, Abstract reasoning 0.75-0.86), while Decision Analysis has lower readings (0.53-0.61). Estimates of the standard error of measurement lie between 90-103 points out of a total possible 3,600 points. Significant correlations between the four subsections are in the range of 0.32 to 0.45, indicating that although there is some cross over the individual subsections are largely testing discrete domains.

ACER does not make information relating to the psychometric properties of UMAT, GAMSAT and HPAT-Ireland publically available (Vale, 2015). They do however state that Rasch methodology is used to measure and analyse the performance of the tests and they undergo independent review by an external auditor annually.

In summary, there is a paucity of publicly available data on the reliability of aptitude tests in general, which has implications for any conclusions regarding their predictive validity. The UKCAT, however, demonstrates moderate to good reliability in a context of openness and transparency.

2.2.11 Aptitude Tests: Issues of Fairness and Impact on Diversity

Despite their widespread use aptitude tests are not uniformly endorsed and concerns have been raised about issues of fairness and their possible impact on student diversity.

There is evidence that standardised aptitude tests across many disciplines show patterns of racial and ethnic differences (Cleland et al., 2012). Davis et al (2013) established that there were significant and large differences in the mean MCAT scores of white versus Latino applicants (standardised mean difference =0.8) and white versus black applicants (standardised mean difference =1.0), with white applicants achieving approximately 5 points higher total MCAT score (out of total possible maximum 45). Similar findings are reported by Andriole and Jeffe (2010). Reiter et al (2012) report lower MCAT scores amongst aboriginal applicants ($p<0.05$). Non-white ethnicity and having English as a second language are both independently associated with lower UKCAT total and subsection scores (all p values <0.001) (Tiffin et al., 2014a). Similar findings are reported by others (James et al., 2010, Yates and James, 2010). Aboriginal or Torres Strait Islander applicants achieved UMAT scores that were lower than white applicants by almost 10 percentiles ($p < 0.001$) (Puddey and Mercer, 2013). While using language spoken at home as a surrogate for ethnicity showed that applicants speaking languages other than English scored on average 14 percentiles lower on total UMAT ($p<0.001$) (Puddey and Mercer, 2013).

Socioeconomic factors are also associated with applicant performance on aptitude tests. In their study of all Australian applicants who sat the UMAT

between 2000 and 2012, (n=118,085) Puddey and Mercer (2013) found that UMAT scores were consistently and significantly associated with a variety of markers of socioeconomic advantage. Total UMAT scores for applicants in the highest two socioeconomic deciles were 13 percentiles higher than those in the lowest two deciles (p < 0.001). Applicants who attended state schools had UMAT total scores almost 4 percentiles lower than those who attended fee paying schools (p<0.001). Similarly, being from a non-professional background, or attending a state school (excluding academically selective state grammar schools) were independent predictors of lower UKCAT total and subsection scores (all p values < 0.001) (Tiffin et al., 2014a). James et al (2010) examined the issue from the opposite perspective and found that higher socioeconomic class and attending a private school were independent predictors of attaining a UCKCAT score above the 30th centile (odds ratio 1.34, 1.91 respectively, both p<0.001). Attending an independent school (largely corresponds to private schools) independently and significantly predicts BMAT Section 1 and 2 scores, while being from a state run further education school or college negatively predicted BMAT section scores (Emery et al., 2011). By contrast, one study reports no significant association between income status and MCAT scores (Reiter et al., 2012).

There is a pattern of gender difference that indicates males out-perform females on most of the aptitude tests under consideration in this review. For example, the UKCAT 12 study found that male sex correlated with UKCAT scores (r=0.061, p<0.001, n=4,742) (McManus et al., 2013C). Similar findings have been reported by others (James et al., 2010, Lambe et al., 2012, Tiffin et al., 2014a, Tiffin et al., 2012). Females are reported to perform lower on GAMSAT (p<0.001) (Puddey and Mercer, 2014). Males outperform females on all sections of the MCAT except the writing section (n= 839,710, p<0.05) (Collin et al., 2009). Males achieve higher UMAT total scores than females (male scores higher by 3.6 percentile, p<0.001, n= 118, 085) (Puddey and Mercer, 2013). Males score higher than females on BMAT Sections 1 and 2 (Emery et al., 2011).

Increasing age is also associated with poorer performance on aptitude tests for medicine. One study found UMAT scores decreased linearly with age, with older applicants (those above 30yrs) scoring almost 23 percentiles lower than those ages 16 or less (p < 0.001) (Puddey and Mercer, 2013). Similar findings are

reported on UKCAT with increasing age correlating negatively with total score (r=-0.06, p<0.001) (McManus et al., 2013c).

Taken together these findings prompt the question: Are aptitude tests inherently biased against certain groups of applicants? Bias in any form of selection tool is highly undesirable and would be considered indefensible. Ironically, some of the groups listed above are the very ones that widening diversity initiatives are endeavouring to attract into the applicant pool. Of concern in this regard, is that many of the patterns of differences in scores on the various aptitude tests mirror those found with tests of previous academic record. White ethnicity, professional or managerial socioeconomic class and private or independent schooling have all been shown to independently predict higher academic grades at the selection stage (James et al., 2010, Powis et al., 2007a, Powis et al., 2007b, Tiffin et al., 2014a). While not the sole reason, this pattern has been held partially responsible for the fact that only approximately 5% of medical school entrants are from non-professional backgrounds (Schwartz, 2004).

It is legitimate, therefore, to consider if selection decisions, that include academic record and aptitude tests, have the potential to doubly advantage some applicants to the detriment of others. While this is a genuine concern, there is some emerging evidence to indicate that the corollary may be the case. Wright and Bradley (2010) demonstrated that UKCAT is less sensitive to the effects of private or independent schooling than A levels or personal statements.   Tiffin et al (2012) determined that making stronger use of UKCAT results in the selection process significantly reduced the disadvantage faced by applicants from certain socio-demographic groups in securing an offer of a place. They examined the relationship between six socio-demographic variables and the outcome of receiving an unconditional offer of a place. For medical schools with strong use of UKCAT (as a threshold to get to the next round of selection) the only two predictive variables were UKCAT (odds ratio 8.59, p<0.001) and academic record (odds ratio 1.63, p<0.001). For medical schools with relatively weak use of UKCAT (for example just considered in borderline applications) the strongest predictors of an offer of a place was academic record (odds ratio 2.63, p<0.001) and UKCAT (odds ratio 1.23 p<0.001). But the model also indicated that in the latter case, non-white ethnicity, male sex, attendance at a state or non-grammar school and

socioeconomic class 4 or 5 were all significantly and independently associated with reduced odds of receiving an offer (odds ratios ranged from 0.51-0.62, all p values <0.005). However, the model did not take into account the input that tools such as personal statements and letter of reference made to the conditional offer of a place. Both of these tools have been criticised for susceptibility to socio-demographic influences (Wright and Bradley, 2010). It is possible that schools with stronger use of UKCAT were relying less on these tools and that this may have influenced the findings.

Concerns that socioeconomic bias could also be extended via the effect of commercial coaching on aptitude tests have led to several studies exploring the evidence for this. Three studies with respect to UMAT indicate that although raw scores are marginally higher for those who received coaching, when controlled for previous academic record, sex and age and possible repeat effect, these differences were no longer significant (Griffin et al., 2012, Griffin et al., 2008, Wilkinson and Wilkinson, 2013). The exception is for Subsection 3, non –verbal reasoning which shows slight gain but not enough to affect overall score. Likewise, there is no evidence that attending at a commercial preparatory course significantly improves UKCAT scores, although applicants who spend more time preparing tended to do better (Lambe et al., 2012).

There have been some efforts in the literature to tease out these possibilities and examine the data more closely for evidence of bias. Davis et al (2013) inspected the MCAT for differential predictive validity with respect to applicant ethnicity. Using logistic regression analysis they compared the predicted and observed success rates of graduating on time and passing USMLE Step I separately for White, Latino and Black medical graduates. The results offered no evidence of under prediction of either Black or Latino applicants, rather the MCAT slightly over-predicted both graduation on time and passing Step 1 of the USMLE but only by a small margin (2.2% fewer black and 1.6% fewer Latino students passed USMLE I first time than were predicted by MCAT). This is in keeping with aptitude tests in general (Sackett et al., 2008). In a systematic review Ferguson et al (2002) conclude that there is good evidence that cognitive tests such as MCAT have significant predictive power for ethnic minorities.

The UKCAT Consortium reports annually on differential item function analysis used to check UKCAT for evidence of disadvantaging applicants according to age, gender or ethnicity. Results of this analysis indicate that on average less than 20 questions out of the total pool of questions used each year show any evidence of differential item function. The patterns vary such, that there is not sufficient evidence to demonstrate that any one group is particularly disadvantaged (UKCAT Consortium, 2007a, UKCAT Consortium, 2007b, UKCAT Consortium, 2008, UKCAT Consortium, 2010, UKCAT Consortium, 2011, UKCAT Consortium, 2012a). These however are not peer reviewed data.

In summary, one of the concerns regarding aptitude tests is the evidence of differential performance of applicants according to certain socio-demographic groups. Applicants who are young, white, male, from higher socioeconomic groups and who attended private or selective schooling tend to perform better. However, many of these factors also impact on applicants' academic attainment and other tools such as personal statements.   There is some emerging evidence that aptitude tests may be less sensitive to these factors than either of these. Hence, at present, the use of aptitude tests is supported on the grounds of widening diversity (Cleland et al., 2012).   There is a need, though for much more research before test users can be confident of these findings. Even perceptions of bias can have profound negative consequences, potentially deterring applicants from applying and this is something that also requires further exploration.

2.3 Multiple Mini Interview (MMI)

2.3.1 Background

This section will review the literature with respect to the use of Multiple Mini Interview (MMI) for selection in medicine. It will explain what is meant by MMI; describe the feasibility, reliability and predictive validity of MMIs. It will then discuss issues of fairness and impact on diversity. Stakeholder acceptability of MMI will be detailed in the next section of the literature review.

The development of MMI was motivated by a desire to test for applicants' non-academic personal attributes, in a way that had better reliability and predictive validity than panel interviews (Eva et al., 2004a). MMI is a process

whereby applicants are examined in a timed circuit similar to an Objective Structured Clinical Examination. It was first described by Eva et al (2004a) and used in McMaster University, Canada. The typical MMI consists of ten stations, each of eight minutes duration with one assessor (Pau et al., 2013). Applicants are provided with written details outlining the tasks to be completed at the station. Stations can either be one to one, where the applicant and the assessor discuss certain issues such as an ethical dilemma, or they entail an interaction or conversation with a role-player observed by the assessor (Dowell et al., 2012).

The use of MMI is growing internationally. For example, in addition to widespread use in medical schools in North America, the UK and Australia it has been applied to selection for family practice training (Hofmeister et al., 2009); emergency medicine residencies (Hopson et al., 2012);physical medicine (Finlayson and Townson, 2011); dental students (McAndrew and Ellis, 2012); nursing students (Perkins et al., 2013); veterinary medicine (Raghavan et al., 2013a); health sciences (Yen et al., 2011) and pharmacy (Stowe et al., 2014).

The MMI is a selection process, not a test (Eva et al., 2009). The content varies, depending on the mission, selection demands and requirements of the specific institution using it (Zaidi et al., 2014). Dowell lists over 25 constructs tested by MMI in various institutions, including communication skills, critical thinking, ethical decision making, empathy, advocacy, dealing with ambiguity, Integrity, preparation & motivation and teamwork (Cleland et al., 2012). While these constructs are often viewed as non-cognitive, Roberts et al (2009) have shown that at least some of these fit well into a cognitive conceptual framework of "*entry level reasoning skills in professionalism*".

2.3.2 Feasibility of MMI

The feasibility of MMI, on the one hand has been established by its growing use. Nevertheless, this cannot be assumed. It is dependent on local resources including the availability of human resources, infrastructure, financial support and expertise in station development (Pau et al., 2013). Other authors have commented on the need for researchers to establish the feasibility and psychometric properties of MMI in different contexts. For example in a population of school leavers in the UK (Husbands and Dowell, 2013), for the

selection of international medical graduates to residency programmes in Canada (Dore et al., 2010), and in Germany, where interview is not a routine part of selection to medicine (Hissbach et al., 2014). Pau et al (2013) recommend that more research is required to explore the use of MMI in different cultural contexts.

MMIs are labour intensive, costly, administratively and logistically demanding to host (Axelson and Kreiter, 2009, Reiter et al., 2006, Uijtdehaage et al., 2011). This limits the amount of applicants that can be assessed to the hundreds rather than thousands (Kulasegaram et al., 2010). Hence it is not likely to be useful as a first stage selection tool for most institutions. However resource analysis comparing MMIs to traditional panel interviews have found that MMIs can be conducted with acceptable reliability indices, using less interviewers, over shorter durations of time than standard panel interviews (Brownell et al., 2007, Eva et al., 2004a, Finlayson and Townson, 2011, Rosenfeld et al., 2008). Therefore for these institutions MMI may represent significant savings.

Faculty time for station development is a separate issue. This can vary significantly between institutions. One German Medical School reported 40 staff hours per station while in the home of MMI, at McMaster University, station development required 3 hours of staff time per station (Hissbach et al., 2014, Rosenfeld et al., 2006). The shorter time frame in the latter may reflect their confidence and familiarity with the tool.

In Germany the costs per applicant in the set-up year were over $1,000 per student, while in the second year this was reduced to approximately $400 (Hissbach et al., 2014). Tiller et al (2013) experimented with internet based MMI (iMMI) using Skype, as an alternative to members of Faculty hosting MMIs in different countries to select international students. The iMMI proved feasible and represented an 84% reduction in administrative costs alone (from AUS $458 per applicant cost the Medical School to mount an international MMI in person, to AUS $35 per applicant for the iMMI).

Other hidden costs include physical location, with MMI requiring substantially more rooms than panel interviews; catering for MMI interviewers and role-players and documentation costs. Using one rater per station is a more cost effective model as long as facilities exist to accommodate 10 or more stations in

order to maintain high levels of reliability (Hissbach et al., 2014). Pau et al (2013) established that the average MMI has 10 stations, each lasting eight minutes and is rated by one assessor. A controlled experiment of optimum station duration found that while candidates' mean station scores at five minutes, were lower than their scores at eight minutes (3.50 and 3.70 respectively, $p<0.01$, $n=175$), the pairs of scores were highly correlated and there was virtually no change in candidate overall ranking, or MMI reliability (Dodson et al., 2009). However the reliability of shortened station duration requires confirmation in other studies and this practice has not been widely adopted to date.

In summary, the feasibility of MMI cannot be assumed and is dependent on the local availability of resources, infrastructure and expertise. For systems already using panel interviews, MMI can provide a feasible alternative and may even represent cost savings to the institution. However introducing MMI de novo, with the creation of a station bank, requires significant investment of time, human resources and money.

2.3.3 Reliability of MMI

The psychometric properties of MMI are context dependent and do not necessarily transfer between different populations and settings (Eva et al., 2009, Uijtdehaage et al., 2011). Reliability depends on issues such as the number of stations, the total test time, the similarities and differences between raters, the different domains being assessed and other setting variables (Hofmeister et al., 2009). There are a number of aspects of reliability: within station inter-item reliability, inter-rater reliability for stations with two or more assessors, inter-station reliability and overall reliability.

A recent systematic review established that MMI used in medical school settings generally show reasonably acceptable levels (Pau et al., 2013). Overall reliability coefficients range from 0.69-to 0.98 averaging 0.75 (Reiter et al., 2007). In their original paper describing the admissions OSCE, Eva et al (2004a) reported an overall test generalisability coefficient of 0.65 and 0.81 for two separate 12 stations MMIs with one rater per station. Since then, the reliability of MMI has been established in a variety of selection and geographical contexts. For example, estimates of 0.69-0.73 have been recorded for undergraduate medical student

selection in the UK and 0.6– 0.8 for recent use of MMI in Arab countries (Ahmed et al., 2014, Dowell et al., 2012, El Says et al., 2013, O'Brien et al., 2011). In the postgraduate training/residency setting some studies are reporting coefficient averaging 0.70 (Campagna-Vaillancourt et al., 2014, Hofmeister et al., 2009). However lower reliabilities have been reported elsewhere – generally where fewer stations have been used. For example 0.54 and 0.55 for an eight and four station MMI respectively (Finlayson and Townson, 2011, Lubarsky and Young, 2013).   An innovative internet based MMI, where applicants were interviewed via skype reported a marginally better reliability coefficient with the internet as opposed to in person MMI (0.76, 0.70) (Tiller et al., 2013).

Eva and Macala (2014) experimented with three forms of MMI: situational judgement type stations, behavioural interviewing type stations and free form stations. They report better reliability coefficients from the stations utilising a behavioural interview format, where applicants were asked to think of an event from their own lives that entailed a particular professionalism challenge and describe what she/ he did, rather than what they would do, or an unstructured conversation on the topic. However, this was a very small study of volunteers (n=41), with just four stations per MMI format.

There are a number of possible sources of error in MMI. The generalisability of an applicant's score depends on a complex set of possible interactions between and within the applicant, the MMI assessor and the station (Dore et al., 2010, Hofmeister et al., 2009, Roberts et al., 2008). In an ideal world, differences in applicants' scores should reflect true differences in their abilities and should not be influenced by which assessor they met, the station they sat, the time of day or circuit of the MMI and so forth (Roberts et al., 2010). Many authors now use generalisability theory to check for the largest source of error in order to minimise it. Several studies report that the candidate by station interaction accounts for a large proportion of variance, typically accounting for around 30% but ranging from 28%-84% (Ahmed et al., 2014, Dodson et al., 2009, Dore et al., 2010, Hissbach et al., 2014, Hofmeister et al., 2009, Till et al., 2013, Uijtdehaage et al., 2011).   This highlights the importance of context specificity. One way to increase the reliability is to increase the number of stations (Eva et al., 2004b).   For example, Dore et al (2010) report on the use of MMIS in three different residency

programmes. Using generalisability theory, they were able to demonstrate that the reliability of the MMIs would have risen from 0.55-0.72 to 0.64-0.79 if they had used 10 stations as opposed to seven.

Dodson et al (2009) established that reliability estimated at five minutes into the station was virtually the same as when the station was allowed to run the full eight minutes (5 minute score: applicant by station variance component 0.42, generalisability co-efficient 0.75: 8 minute score applicant by station variance component 0.42 generalisability co-efficient 0.78). Should this finding be confirmed in other studies the reduced station duration would have positive implications for feasibility.

Other sources of unwanted variance are the rater and the rater-applicant interaction (Eva et al., 2004b). Assessor leniency or stringency is an example of interviewer related error. It describes a tendency for the assessor to preferentially use the top or the bottom of the rating scale and can account for approximately 9% of variance (Roberts et al., 2010). Use of fair scores, as calculated by Rasch modelling, has been suggested as one way to deal with compensate for this. A fair score is the score an applicant would have received if they had met the average assessor and the station was of average difficulty (Roberts et al., 2010). However, only two papers were identified that reported on this method. If fair scores had been applied, the selection outcome would have been different for approximately 10% of applicants (Roberts et al., 2010, Till et al., 2013). It is more common to try to make improvements by providing thorough assessor training on the use of the entire marking scale (Hissbach et al., 2014). For example Uijtdehaage et al (2011) witnessed an improvement from 0.58 to 0.71 in part by instructing assessors to mark applicants *relative* to the applicant pool.

In summary, the reliability of MMI depends on a complex interaction of applicant, assessor, and station. In general, reliability of MMI is acceptable to good and better than traditional interview. However, the level of variance attributable to factors beyond the capability of the applicant is still a cause of concern. Sophisticated techniques, including generalisability and Rasch modelling, are used to identify and minimise the sources of these errors. A

criticism of reported reliability for MMIs is that not all authors provide it and many do not state which reliability coefficient they are quoting. This can give rise to confusion.

2.3.4 Predictive Validity of MMI

Due to its novelty, evidence for the predictive validity of MMI is only emerging in recent years (See Table 5 for summary of identified studies). To date, the evidence shows that MMI significantly correlates with medical student OSCE performances in the early years of the medical degree programme. The strength of this relationship ranges from small to medium (uncorrected r 0.19 to 0.32), rising, when corrected for range restriction (r from 0.24 to 0.50) (Eva et al., 2004c, Husbands and Dowell, 2013, Reiter et al., 2007). Regression analysis confirms that MMI contributes significantly to the explanation of variance in OSCE scores (beta ranges from 0.18-0.44). Concerns that this may just reflect students' strengths in assessment technique (MMI and OSCE being largely identical formats) are somewhat allayed by evidence that MMI is also significantly associated with scores on written assessments of knowledge. Admittedly these levels are lower, but MMI are not designed for this purpose (r 0.18-0.27, rising to 0.26 to 0.33 when corrected for range restriction) (Eva et al., 2004c, Husbands and Dowell, 2013).

In comparison to other selection tools, MMI was the only selection variable predictive of OSCE results in one study (Eva et al., 2004c) while neither UKCAT nor UCAS were found to consistently predict performance to the same degree as MMI in the other (Husbands and Dowell, 2013). Criticisms of these studies includes that they are each based in just one medical school, with small numbers of participants, followed up for a relatively short duration of one to two years.

In the senior clinical years one study found that MMI was the only admission variable significantly correlated with clinical clerkship director ratings and measures students' clinical skills and professional behaviour (corrected r values 0.57, 0.51 p<.001 respectively) (Reiter et al., 2007). There have been suggestions that MMI might be useful in excluding students whose professionalism might go on to give cause for concern (Eva et al., 2004c, Reiter et al., 2007). However, this were based on retrospective analysis of the MMI scores of just two students

whose behaviour was identified as a cause of concern, hence such assertions cannot be substantiated without further study.

Several papers have looked at correlations between MMI and performance in the Canadian Medical Licensing Exam Parts I and II (Part I must be taken within a year of graduation). The findings are somewhat mixed. The best quality evidence comes from Eva et al (2012) who examined the scores of 751 graduated doctors who had been offered a place at McMaster medical school following selection which included an admission MMI. These were categorised into those who accepted a place at McMaster and were trained there (n=472) and those who declined a place in McMaster and were trained elsewhere (n=279). Doctors who were selected using MMI scored significantly higher on both Part I and Part II (average 15-20 points difference, $p<0.01$). Performance on MMI independently and significantly predicted outcomes on total scores and every subsection score for Canadian Medical Licensing Exam Part I and Part II (beta values ranged 0.10 to 0.23, all p values <0.01) and was the only selection measure to do so. Elsewhere MMI has correlated significantly with the number of OSCE stations passed in Part II ($r=0.35-0.43$, $p<0.05$) (Eva et al., 2004c), the ethical and population health (beta > 0.4, $p < 0.01$) and the clinical decisions making subsections of Part I (beta 0.35, $P < 0.05$) (Reiter et al., 2007).

Two studies were identified which reported otherwise. Hofmeister et al (2009) used the MMI for selecting international medical graduates to a training programme in Family Practice. They found no significant correlation with Part I or Part II of the licensing examination, or an additional mandatory examination for international graduates (Canada Evaluating Examination). Hopson et al (2014) mounted a research MMI with new interns on an Emergency Medicine training programme and found MMI did not correlate with either USMLE Step 1 or 2. These papers need to be interpreted with caution as they are technically concurrent rather than predictive validation studies as candidates sat the licensing/membership examinations either before the MMI or in the same short time frame, both had small numbers (n=-71), were set in one location and in the case of the latter, the MMI was not used for actual selection.

In summary, there is emerging evidence that MMI has predictive validity for some important outcomes in both undergraduate and licensing examinations, particularly in the domain of communication/clinical skills and professionalism. The strength of the relationship overall is medium to large. However, these findings are based for the most part on a limited though growing number of small studies, in single institutions and require further replication and substantiation.

Table 5 Summary of Predictive Validity Findings for MMI

Reference	Predicted Variable / Students/ Trainees (n=number)	Significant Correlation Coefficients	Regression Analysis
Eva et al 2004	Mean Personal Progress Inventory scores – a knowledge MCQ taken 4 times over the study. Mean OSCE score -10 station OSCE taken twice. 2nd Yr students (n=45)	MMI predicted first OSCE (r=0.32, p<0.1), Second OCSCE (r=0.23, p=NS),MMI correlated with scores on 3rd and 4th PPI r=0.27, 0.24 (p<0.1), but not the first or second PPI. GPA correlated with 3rd and 4th PPI (r=0.26, 0.28, p<0.1) Essay with 1st and 2nd PPI (r=0.29, 0.25, p<0.1). Traditional Interview was NS all figures uncorrected	MMI was the only admission factor predictive of OSCE (beta 0.44, p<0.05) and both GPA and autobiographical submission were predictive of PPI (beta 0.54, 0.45, p<0.05 respectively)
Eva et al 2009	Part II Canadian Medical Licensing Exam –Number of OSCE stations passed/ average total OSCE score. Two cohorts: (1) Postgraduate residents (n=22) and (2)graduates who sat MMI at time of selection (n=34)	Cohort (1) MMI correlated with number of OSCE stations passed r=0.43 , p< 0.05, Total OSCE score r= 0.36, p= 0.10 NS. Cohort (2) : MMI correlated with number of OSCE stations passed r= 0.35 p< 0.05 Total OSCE score r= 0.19 NS. MMI was the only admission variable that significantly correlated with Part II.	Not reported Note -The post graduate cohort was not technically predictive as many of them sat the MMI after their licencing exam.
Eva et al 2012	Part I and II of the Canadian Medical Licensing Exam Subjects = medical school graduates, who either had or had not been offered a place at McMaster Medical School (n= 751)	Doctors who were selected using MMI scored significantly higher on both Part I and Part II (average 15-20 points difference) (p<0.01).	MMI independently and significantly predicted outcomes on total scores and every subsection score for Part I and Part II (betas ranged 0.10 to 0.23, all p values <0.01).
Hopson et al 2014	Association between MMI and other measures of selection to Emergency Medicine (EM) training. EM interns in their first month of training (n=71)	EM clerkship in third and fourth-year medical school predicted MMI (F[1, 66] = 4.18; p < 0.05). MMI did not correlate with either USMLE Step 1, (r = 0.04, p = NS) or USMLE Step 2 (r=0.13, p=NS) . Applicants' lowest global rank on a letter of recommendation correlated with MMI performance (r=0.36, p<0.05).	Note not really predictive as MMI and other selection tests taken in short time frame
Hofmeister et al 2009	Family practice selection OSCE / Canada Evaluating Examination/ Canadian licensing exam Part I, Part II, International Applicants to Family Medicine (n=71)	MMI average station score did not correlate significantly with OSCE total or clinical score but did with OSCE Communication score (r=0.38, p<0.01) , but did not correlate significantly with the Canada Evaluating exam or Part I or Part II	Not reported
Husbands & Dowell 2013	Two cohorts (n=418 in total) 2009 cohort 2 yr follow up 2010 Cohort 1 yr follow up. Exams and cohorts analysed separately. Yr 1 Semester 1 OSCE / written; Semester 2 OSCE/ written. Yr 2 Semester 1 OSCE / Semester 1 Written	MMI correlated significantly with two out of the five written assessments (r 0.18, 0.23 rising to 0.26 and 0.33 when corrected for range, p<0.05). MMI correlated significantly with 4 out of the 5 OSCE assessments (r 0.19 to 0.35 rising to 0.24 to 0.50 when corrected for range, p<0.05). UKCAT only significant in Sem 1 , Yr1 (r 0.18-0.25 rising when corrected to 0.24-0.33) UCAS was NS with any outcome	MMI contributed significantly to the explanation of variance in OSCE scores in both years and cohorts (beta ranged 0.18-0.34,most ps<0.0001) . And also to written examinations (beta 0.21, 0.26, p<0.05). Overall MMI accounted for between 6-15% of the variance in examination scores
Reiter et al 2007	Med-school measures : Mean Personal Progress Inventory (PPI) scores , Mean OSCE score; Final Year Clinical Clerkship encounter cards; Final Year End Clinical Clerkship Director rating Part I Canadian Medical Licensing Exam (9 sub-scores and 1 total score, including MCQ /clinical decision making/ Legal, Ethical and Organisational Aspects of Medicine &Population Health/ Ethical, Legal / Organisational Aspects (n=45)	MMI was the only admission variable significantly correlated with Clerkship director ratings and clinical encounter cards (r 0.57, 0.51 p<.001 respectively) GPA correlated with PPI (r 0.33, p<0.05) . Personal interview, autobiographical summary or simulated tutorial were NS except tutorial negative for PPI (r=-.28, p<0.1) GPA and simulated tutorial correlated with total Part I score (r=.26, -.28, P<0.01) . MMI correlated with CLEO and PHELO (r=0.39, 0.37, P<0.001).	Results provided for regression only of Part I licencing exam. Only the MMI was predictive of CLEO or PHELO (beta > 0.4, P < 0.01). Only GPA was predictive of MCQ (beta 0.38, P < 0.05); Only MMI was predictive of Clinical decision making (beta 0.35, P < 0.05), and MMI and GPA were equally predictive of overall test performance (standardised b> 0.3, P < 0.06).

2.3.5 MMI: Issues of Fairness and Impact on Diversity

To date, most studies that report on gender indicate that there is no evidence that MMI performance differs between the sexes (Eva et al., 2004a, Lemay et al., 2007, O'Brien et al., 2011, Uijtdehaage et al., 2011). However, one study found MMI mean score was significantly associated with female sex (p<0.02) (Jerant et al., 2012). Age did not correlate with MMI scores in two studies (Lemay et al., 2007, O'Brien et al., 2011). Conversely Dowell et al (2012) found that graduate/mature candidates achieved significantly higher mean scores than all other groups of candidates (p<0.01). Likewise, Jerant et al (2012) found applicants aged 19-21 years had significantly lower mean scores than those aged 25-39 years.

Two studies compared the MMI scores of Aboriginal and non-Aboriginal applicants (Moreau et al., 2006, Raghavan et al., 2013b). Neither found any significant association between MMI scores and Aboriginal ethnicity. In the former study half of the MMI assessors as well as half of the applicants were Aboriginal. The variance component attributable to the interviewer type by interviewee type was not significant. However, these findings relate to a sample size of just of five self-declared Aboriginal applicants and six aboriginal raters and must be interpreted cautiously. The latter study reported that applicants from a rural second level school scored significantly lower than their urban counterparts by a very slight margin (equating to approximately 1.75% lower). As Aboriginal applicants have a much higher rate of attending rural second level schooling this raises concerns about a possible negative impact. Again this finding needs corroboration in larger, multi centred studies.

Hofmeister et al (2009) describe the use of MMI for selection of international medical graduates to residency training in Canada. They had 71 applicants, from 23 different countries, speaking 25 separate first languages, although over 70% had attended medical school in English. They report no differences in mean MMI scores between students attending medical school in English and those who did not. I did not find any paper describing the use of MMI for the selection of international medical students.

Others have queried whether MMI favours applicants with certain personality profiles and how this may impact on student diversity. The findings have been conflicting. One study found no significant correlation between MMI and the five factor model of personality (Kulasegaram et al., 2010). A more recent larger study (n=868), over three cohorts of students found significant correlations existed with extraversion (r range 0.19 to 0.30), conscientiousness (r ranged 0.20 to 0.25) and agreeableness (r ranged 0.17 to 0.19, all p values <0.02) (Griffin and Wilson, 2011). The latter findings were mirrored in a further study which confirmed that extraversion was associated with higher MMI scores (n=444, r = 0.35, P < .01) (Jerant et al., 2012). The authors express concern that this may lead to diminished student diversity and negatively impact on personal and professional growth.

Uijtdehaage et al (2011) report no relationship between MMI scores and applicant self-identified disadvantaged status. This was the only study identified that reported on the relationship between socioeconomic status and MMI. In light of the widening diversity agenda in selection clearly this is an area that merits further investigation.

There is a small amount of emerging evidence that MMI may be relatively insusceptible to coaching. Findings from one small study (n=287) in an Australian medical school found that self- reported coaching had no significant impact on total MMI scores, and in fact resulted in significantly lower scores in one communication station (Griffin et al., 2008). Elsewhere, in a series of controlled experiments no significant difference in scores were detected between applicants who were informed of MMI station content in advance, and those who were not (Reiter et al., 2006). The evidence for a repeat effect is split, with one study finding no evidence (Reiter et al., 2006), while another reports that there was significant improvement on stations with identical or alternative forms, but no difference in scores on new stations (Griffin et al., 2008).

In summary, the evidence to date suggests that MMI is neutral with respect to gender and equivocal regarding age. There is some evidence that personality factors, particularly extraversion and MMI are associated with attaining higher MMI scores. This may have implications for student diversity; however this may

be dependent on the relative weighting given to MMI in the selection process. There is a dearth of evidence regarding MMI and demographic, socio-economic, cultural and ethnic background of the applicant. These all require further study especially in the context of widening diversity.

2.4 Stakeholder Views of Selection Tools for Medicine (A Systematic Review)

2.4.1 Background

Increasingly stakeholder views are recognised as an important evaluative measure of selection tools. However, this is an under-researched aspect of selection to medicine (Cleland et al., 2012).

Stakeholders represent a very diverse group, with many agendas and desired outcomes. The arguments for considering stakeholder views are compelling. For example, with respect to applicants alone, there is evidence that those who find a particular selection tool invasive or unfair, may negatively view the medical school using this tool, may dissuade others from applying and may be less likely to accept an offer or more likely to bring a law suit against a medical school (Cleland et al., 2012, Patterson et al., 2011). Furthermore, negative applicant reactions are associated with loss of competent applicants from the selection process and may deter them from reapplying (Patterson et al., 2011).

Medical students are an important stakeholder group, who have immediate experience of the selection process and unique insights into it. However, this stakeholder group should be viewed as distinct from applicants for three reasons. Medical students are a relatively small subgroup of applicants who have successfully navigated the selection process. Therefore, they may be more likely to view whatever process they personally came through in a positive light. Secondly, once they enter medical school they embark on a process of socialisation and education as medical professionals. This may lead them to have insights into the relevance of selection tools to medical practice that may be unknown to applicants. Thirdly, medical students, particularly in their senior years, are often included in the selection process from the other side- by serving

on admission panels and committees which provides a different perspective to the applicant view (Dowell et al., 2012).

A third group includes those with direct responsibility for specific aspects of the selection process generally, comprising admissions committees and deans, medical school faculty, interviewers and selectors and those involved in shortlisting. Increasingly, this group is including medical student, community members and laypersons (Marrin et al., 2004, Reiter and Eva, 2005). Working within medical council regulations, this highly influential group largely determines the weighting and relative contribution of the different selection tools to the individual institutions' selection process.

Other noteworthy stakeholders include members of the medical profession (for example physicians, surgeons and general practitioners) who will be working in clinical teams with future graduates selected by the various processes. Medical regulatory bodies that have responsibility for overseeing selection form another stakeholder group. Likewise, parents, teachers and career guidance councillors are heavily invested in the aspirations of school students interested in applying to medicine. Due to the investment of tax payers' money in the selection and training of doctors, the general public are stakeholders in the process. Arguably, one of the most important stakeholder groups is patients themselves.

To our knowledge, there has been no review of the literature drawing together the views of this disparate and wide reaching group of stakeholders. Therefore, the aim of this review is to systematically review the literature with respect to stakeholder views of selection to medical school.

2.4.2 Methods

Search Strategy

Nine electronic databases: PubMed, EMBASE, SCOPUS, OVID Medline, PsycINFO, Web of Science, ERIC, British Education Index and Australian Education Index were searched for relevant published literature from January 2000 to July 2014. The justification for this timeframe is twofold a) due to changes in society, such as demographic patterns and economic climate, stakeholder views from more than 15 years ago may no longer represent the views of the current

stakeholder pool and b) many of the selection tools in current use were not either available or widely used in medicine prior to 2000. A similar timeframe was applied in a recent review of selection tools for medicine (Cleland et al., 2012).

Relevant papers were identified using search terms for each of the four concepts "stakeholder", "views", "selection" and "medical school". Terms were mapped to MESH terms or the appropriate term from the controlled thesaurus of the various databases.   In addition, text word searches were used for key words. The search strategy was informed by discussions with a research librarian (Jane Mulligan).

"Stakeholders" were defined as those who are affected by or can affect selection processes   (Freeman, 2010). The search terms for stakeholder were deliberately cast widely to encompass as many stakeholder groups as possible. "View" was defined as an opinion or attitude. "Selection" was taken to mean any admission test or entrance assessment process that a medical school applicant would have to go through in order to be offered a place. "Medical school" was taken to include both graduate and undergraduate schools. Additionally, as there is significant overlap between some tools used for selection to medical school and selection to higher professional training (for example Multiple Mini Interview and Situational Judgement Tests are increasingly used in both settings) this search was widened to include internship and residency.   However, papers were only included if they reported on stakeholder views of issues relevant to selection to medical school. For example, papers referring to the residency match process were excluded.   Within each concept, terms were joined using the Boolean operator "OR". The four searches were then combined with the operator "AND". Language or type of publication restrictions were not applied during the searching phase. (See Chapter 2 Appendix 2 for sample search).

The databases EMBASE, ERIC, British Education Index and Australian Education Index provided access to many sources of grey literature as they include conference and research publications that may not have been published in the peer reviewed journals.   In addition, grey literature searching was facilitated by networking with other researchers at conference and by email, which enabled me to identify others undertaking higher degrees in selection and

facilitated access to unpublished PhD work. Published books of conference abstracts for the annual scientific meetings of the Association for the Study of Medical Education (ASME) and the Association for Medical Education in Europe (AMEE) for the years 2011, 2012 and 2013 were hand searched for relevant papers (the 2014 books were not available at the time of the review) in addition to the conference abstracts of the biannual Ottawa conferences for the years 2012 and 2014. The reference lists of papers included in the review were hand searched for additional relevant publications. In addition, two experts in field were contacted for any unpublished work. The sensitivity of the search was estimated by screening retrieved articles for a number of published papers known to the authors at the outset.

Inclusion and Exclusion Criteria

The inclusion criteria were: (a) Studies published between January 2000 and July 2014 (b) Studies on selection to medical school (c) Studies on selection to residency and internship programmes which described selection processes relevant to selection to medical school (d) Studies which reported on the views of at least one stakeholder group established by means of quantitative, qualitative or mixed methods research.   The exclusion criteria were: (a) Reviews or articles which were not original studies (b) Papers for which an English language translation was not available.

Study Selection and Data Extraction

Figure 3 illustrates the steps from initial identification of records, to identifying those included and excluded. Records were retrieved from the electronic search as follows: all records identified in the electronic database search (total n= 2685) and by the additional means described above (n=26) were transferred to EndNote database, duplicates were removed (by automatic deduplication and manual check) and the remaining records were inspected (n=1016).

One reviewer (MK) assessed the potential relevance of all titles and removed records if the title indicated that they did not meet the inclusion criteria. As a quality assurance measure a random sample of 80 excluded titles (>10%) were assessed independently by AW Murphy (research supervisor) and compared with

the initial sorting. There were only four citations identified by this means that caused disagreement and this was resolved by discussion. Subsequently the entire title list was re-examined by MK to confirm that they were to be excluded.

Two reviewers (MK and AWM) independently screened all the abstracts of the retained records (n=220) to identify those to be assessed on full text, based on the inclusion and exclusion criteria. There was almost complete agreement between the two reviewers at this stage, with only six abstracts requiring discussion.   Agreement was reached following discussion and of the disputed abstracts five were included for full review of paper and one was excluded.

This left a total of 106 records which were read in full by both reviewers and independently assessed for eligibility to be included in the full review. Seven records were disputed and agreement was reached by discussion, when four were excluded and three included in the full review. Subsequently 71 records were included for full review and 35 excluded. Figure 3 indicates the reasons for exclusion.

The following data were collected from each eligible record and collated in a data extraction form: author, publication year, type of publication, principal study aim, location and setting, study design, medical selection tool used, stakeholder characteristics (including identification of stakeholder, sample size, response rate, gender, age, socioeconomic group, background if provided), data collection method, overall findings and indicators of study quality. (See Chapter 2 Appendix 3 for example).

Declaration of interest

MK was co-author on four of the retained records and AWM was co-author on two (Dennehy et al., 2013 , Kelly et al., 2014a, Kelly et al., 2014b, Stevens et al., 2014). As this review includes publications up to July 2014, two papers arising from this thesis are included. This is indicated by referencing the appropriate thesis chapter.

Identification

Records identified through database search (n = 2686)

Additional records identified through other sources (n =26)

Total Records n=2712
after date limits applied
n=1726 (986 deleted)

Screening

Records after duplicates removed
(n =1017) (709 deleted)

Records excluded:
From title (n =797)

Records excluded:
From abstract
(n =114)

Records screened
(n =1017)

Eligibility

Full-text publications assessed for eligibility (n = 106)

Full-texts excluded, with reasons (n=35)
Did not report stakeholder view (n = 7)
Opinion/ personal view or editorial (n= 6)
English translation not available from author (n=3)
Outside of field of medicine (n = 5)
Not relevant to medical student selection (n=5)
Match rank process (n=6)
Full text unavailable (n=1)
Did not meet inclusion criteria C (n=2)

Included

Records included
in review (n = 71)

Figure 3 Study Search Strategy and Review Process

Figure template taken from : http://www.prisma-statement.org./statement.htm

Quality Assessment Strategy

Quality criteria were assessed using the Medical Education Research Study Quality Instrument (MERSQI), a validated ten-item checklist for rating the methodological quality of medical education research papers (Reed, 2007). This instrument has six domains, each with a maximum score of 3 points giving an overall maximum score of eighteen. The domains are: study design, sampling (number of institutions studied and response rate), type of data, validity of evaluation instrument (internal structure, content and relationship to other variables), data analysis (appropriateness of analysis, complexity of analysis) and outcomes. The MERSQI has been used by a number of recent medical education systematic reviews (Brennan and Mattick, 2013, Kothari et al., 2011, Mookherjee et al., 2013, van der Leeuw et al., 2012, Zendejas et al., 2013). It was deemed an appropriate tool for this review as it was specifically developed for medical education research and accommodates observational studies. MERSQI scores were used to compare quality and were not used for the purpose of excluding records from this review.

Samples of five records were independently scored using the instrument by the researcher and supervisor (MK and AWM) and scoring was discussed and debated. A challenge to applying MERSQI, which gave rise to debate, was that reporting stakeholder views was sometimes not the principal aim of the study, rather these findings were additionally described in the methods +/or results section of the record. After discussion, a decision was taken to apply the MERSQI to the research that reported on stakeholder views as this is the subject of this systematic review. This ensured that the reader could compare the quality of research evidence of stakeholder views of selection to medicine. This decision was supported by email advice from the author of MERSQI who confirmed that the tool has been previously applied to specific portions of research – e.g. the quantitative strand of a mixed methods study (Reed, 2014). MERSQI is not suitable for the quality assessment of qualitative studies, but as only 9 records were qualitative the decision was to simply record that a MERSQI score was not relevant to them.

Quality Assessment and Evidence Synthesis

The MERSQI ratings for the included records ranged from 3 to 10.5. The mean was 7.2 and the median 7.5. The highest possible score on this measure is 18. As a comparison a review of over 200 published peer review medical education papers determined that the mean MERSQI of published papers was 9.95 (range 5-16)(Reed, 2007). This is discussed further in the review limitations and strengths. Table 6 lists the MERSQI scores for the included records in alphabetical order.

Study Designs

A data display matrix, based on the data extraction forms, briefly summarises context, stakeholder group and main research findings of the studies included in this review is presented at the end of this review (see Table 7).

Included records comprised nine qualitative studies (Greenhalgh et al., 2004, Jayasuriya et al., 2012, Kelly et al., 2014b, Kumar et al., 2009, Mathers and Parry, 2010, Stagg and Rosenthal, 2012, Turner and Nicholson 2011, White et al., 2011, Wright, 2012). Seven were mixed methods studies (Campagna-Vaillancourt et al., 2014, Cleland et al., 2011, Goulston and Oates, 2009, Hofmeister et al., 2008, Marrin et al., 2004, Razack et al., 2009, Stevens et al., 2014). The remaining records were quantitative.

Twelve records were abstracts (Brown and Griffin, 2012, Gula et al., 2014, Husbands et al., 2014a, Jayasuriya et al., 2012, Lubarsky and Young, 2013, Niyomdecha et al., 2012, Patel et al., 2011, Patterson et al., 2013, Rich, 2011, Rodgerson et al., 2013 , Samarasekera et al., 2014, Vermeulen et al., 2012).   Two were journal short research reports (Johnson and Elam, 2001, Milne et al., 2001) (Johnson & Elam 2001, Milne et al 2001), one PhD (Wright 2012), one report (Goulston and Oates, 2009), one letter describing original research (Kaffenberger et al., 2014). The remaining were journal research papers.

Twenty-four records were from studies conducted in the UK, 12 in Canada, 12 in USA, 6 in Australia, 5 in Ireland, 2 in New Zealand, 2 in Belgium, 1 in Australia/ Canada, I in USA/Canada, and 1 each in Israel, Pakistan, Netherlands, Singapore, Thailand, Saudi Arabia.

Table 6 MERSQI Checklist for Assessing the Quality of Included Studies

Study	Study Design (max 3)	Sampling (max 3)	Type of data (max 3)	Validity of evaluation instrument (max 3)	Data Analysis (max 3)	Outcomes (max 3)	Overall Score (max 18)
Adams 2009	1	1.5+0.5	1	0+0+0	1+1	1	7
Agrawal et al 2005	1	1.5 +1	1	1+0+0	1+1	1	8.5
Brown & Griffin 2012	2	0.5+0.5	1	0+1+0	UTD+UTD	1	6
Brownell et al 2007	2	0.5+1.5	1	0+0+0	1+1	1	8
Campagna-Vaillancourtet al 2013	2	1+1.5	1	0+0+0	1+1	1	8.5^
Cleland et al 2011	1	1.5+1.5	1	0+0+0	1+2	1	9^
Christakis et al 2010	1	1+1.5	1	0+0+0	1+1	1	7.5
Daram et al 2014	1	0.5+0.5	1	0+0+0	1+1	1	6
Dennehy et al 2013	1	1+1.5	1	0+0+0	1+1	1	7.5
Dhar et al 2012	1	1+1	1	0+0+0	1+2	1	8
Dore et al 2010	2	1+0.5	1	1+0+0	1	1	7.5
Dowell et al 2012	2	0.5+1	1	0+0+0	1+1	1	7.5
El Says et al 2013	1	0.5+0.5	UTD	UTD	UTD	1	3
Eva et al 2004 (a)	2	0.5+0.5	1	0+0+0	1+1	1	7
Eva & Macala 2014	2	0.5+0.5	1	0+0+0	1+1	1	7
Gale et al 2010	2	0.5+0.5	1	0+0+0	1+1	1	7
Goulston & Oates 2009	2	0.5+0.5	1	0+0+0	UTD	UTD	4
Greenhalgh et al 2004	NR	NR	NR	NR	NR	NR	NR
Griffin et al 2008	1	1+1.5	1	0+0+0	1+1	2	8.5
Gula 2014	2	1+0.5	1	0+0+0	UTD +UTD	1	5.5
Harris & Owen 2007	1	0.5+0.5	1	0+1+0	1+1	1	7^

Key: *NR= not relevant because qualitative study, UTD –unable to determine, ^indicates mixed methods studies with the MERSQI score applied to the quantitative strand.*

Study	Study Design (max 3)	Sampling (max 3)	Type of data (max 3)	Validity of evaluation instrument (max 3)	Data Analysis (max 3)	Outcomes (max 3)	Overall Score (max 18)
Henry 2006	1	0.5+0.5	1	1+1+0	1+1	1	8
Hofmeister et al 2008	2	1+1.5	1	0+0+0	1+1	1	8.5^
Hopson et al 2014	1.5	1.5+1.5	1	0+0+0	1+2	1	10
Humphrey et al 2008	2	1+1.5	1	0+1+0	1+1	1	9.5
Husbands et al 2014 (a)	1	UTD+ 0.5	1	0+0+0	UTD+UTD	1	3.5
Jauhar et al 2008	1	1.5+1	1	1+0+0	1+1	1	8.5
Jayasuriya et al 2012	NR	NR	NR	NR	NR	NR	NR
Johnson & Elam 2001	2	0.5+1.5	1	0+0+0	1+1	1	8
Kaffenberger et al 2014-	1	1+0.5	1	0+0+0	1+1	1	6.5
Kleshinski et al 2008	2	0.5+1	1	0+0+0	1+1	1	7.5
Kelly et al 2014 (a)	2	0.5+1	1	0+0+0	1+1	1	8
Kelly et al 2014 (b)	NR	NR	NR	NR	NR	NR	NR
Kumwenda et al 2013	1	1.5+0.5	1	1+1+0	1+1	1	9
Koczwara et al 2012	1	0.5+1.5	1	1+0+0	1+1	1	8
Kumar et al 2009	NR	NR	NR	NR	NR	NR	NR
Lambe et al 2012	1	0.5 +1	1	0+0+0	1+1	1	6.5
Lievens & Sackett 2006	2	1.5+1	1	1+1+0	1+1	1	10.5
Lievens 2013	1	1.5+1	1	1+0+0	1+1	1	8.5
Lubarsky &Young 2013	1	0.5+0.5	1	0+0+0	UTD +UTD	1	4
O'Brien et al 2011	2	0.5+0.5	1	0+0+0	1+1	1	7
Marrin et al 2004	2	0.5+1	1	0+1+0	1+1	1	8.5
Mathers & Parry 2010	NR	NR	NR	NR	NR	NR	NR
Milne et al 2001	1	0.5+1.5	1	1+0+0	1+1	1	8
Mitchson 2009	1	0.5+1.5	1	0+0+0	1+1	1	7
Monroe et al 2013	1	1.5+1.5	1	0+1+0	1+1	2	10

Table 6 Continued

Study	Study Design (max 3)	Sampling (max 3)	Type of data (max 3)	Validity of evaluation instrument (max 3)	Data Analysis (max 3)	Outcomes (max 3)	Overall Score (max 18)
Niyomdecha et al 2012	1	0.5+0.5	1	0+0+0	UTD+UTD	1	4
O'Brien et al 2011	2	0.5+0.5	1	0+0+0	1+1	1	7
O'Flynn et al 2013 (a)	1	1.5+0.5	1	0+0+0	1+1	1	7
Patel et al 2011	1	0.5+0.5	1	0+0+0	1+1	1	6
Patterson et al 2009	2	0.5+0.5	1	1+0+0	1+1	1	8
Patterson et al 2011	1	1.5+1	1	1+1+0	1+1	1	9.5
Patterson et al 2013	1	UTD	1	0+0+0	UTD	1	3
Randall et al 2006	1	0.5+0.5	1	0+0+0	1+1	1	6
Razack et al 2009	2	0.5+1.5	1	0+0+0	1+1	1	8^
Rich 2011	2	0.5+0.5	1	0+0+0	UTD +UTD	1	5
Rodgerson et al 2013	1	0.5+1	1	0+0+0	UTD+UTD	1	4.5
Samarasekera et al 2014	2	0.5+0.5	1	0+0+0	UTD+UTD	1	5
Stagg & Rosenthal 2012	NR	NR	NR	NR	NR	NR	NR
Stevens et al 2013	1	1.5+1.5	1	0+0+0	1+1	1	8^
Tiller et al 2013	2	1+1	1	0+0+0	1+1	1	8
Turner & Nicholson 2011	NR	NR	NR	NR	NR	NR	NR
Uijtdehaage et al 2011	2	0.5+1.5	1	1+0+0	1+1	1	9
UKCAT Consortium 2009/2010	1	1.5+0.5	1	0+0+0	1+1	1	6
UKCAT Consortium 2011	1	1.5+0.5	1	0+0+0	1+1	1	6
Vermeulen et al 2012	1	UTD+0.5	1	0+0+0	UTD+1	1	4.5
Waheed et al 2011	2	0.5+ 0.5	1	0+0+0	1+1	1	7
Westwood et al 2008	1	0.5+1.5	1	0+0+0	1+1	1	7
White et al 2007	NR	NR	NR	NR	NR	NR	NR
Wilkinson & Wilkinson 2013	2	0.5+0.5	3	0+0+0	1+2	1.5	10.5

Key: *NR= not relevant because qualitative study, UTD –unable to determine, ^indicates mixed methods studies with the MERSQI score applied to the quantitative strand.*

Table 6 Continued

Study	Study Design (max 3)	Sampling (max 3)	Type of data (max 3)	Validity of evaluation instrument (max 3)	Data Analysis (max 3)	Outcomes (max 3)	Overall Score (max 18)
Wright 2012	NR	NR	NR	NR	NR	NR	NR
Ziv et al 2008	2	0.5+0.5	1	0+0+0	1+1	1	7

Key: *NR= not relevant because qualitative study, UTD –unable to determine, ^indicates mixed methods studies with the MERSQI score applied to the quantitative strand.*

2.4.3 Results

The following is a narrative synthesis of the records included in this review. It is presented under the heading of each stakeholder group. The research largely explored the views of three main groups: a) Applicants; b) Interviewers, Faculty and Admissions Committee Members and c) Medical Students, with any other stakeholders reported under the category d) Other Stakeholders. Many studies explored more than one group, most commonly applicants and interviewers. This was particularly common with respect to studies exploring stakeholder views of multiple mini interview, which formed the largest single subgroup.

The Views of Applicants

Applicants constituted the most researched stakeholder group. Overall 46 records were identified that included the views of applicants, with medical school applicants being by far the largest group. Also included in this stakeholder group were records exploring the views of residency/intern applicants, international applicants, mature applicants, underrepresented minority/ widening access applicants and one that explored the views of high achieving school children as potential medical school applicants (Greenhalgh et al., 2004). These are synthesised into two categories: applicants' views of selection tools and the views of underrepresented and minority applicants.

Applicants' views of selection tools

Applicants' views of the multiple mini interview (MMI) have been extensively surveyed in many different countries. The research in this category was generally of good quality (10 records with a MERSQI score over 8), achieving high response rates (9 records with response rates over 75%) and reasonable number of applicants included (9 records where n = ranged 69 -324). The evidence is based largely on post MMI exit surveys.

Applicants are overwhelmingly supportive of MMI. Dowell et al (2012) found that 94% of medical school applicants either agreed or strongly agreed that it was a fair assessment tool. Overall, mean ratings indicate that medical school applicants perceive that MMI is relatively free of gender or cultural bias, that the quality of advance information and clarity of instructions are good and that applicants have adequate opportunity to present their abilities and strengths

(Brownell et al., 2007, Dowell et al., 2012, Eva et al., 2004a, Uijtdehaage et al., 2011). Using very similar questionnaires, studies from a range of postgraduate residency and training programmes report good levels of applicant satisfaction with lack of gender or cultural bias, clarity of information and instructions and opportunity to demonstrate skills and abilities (Campagna-Vaillancourt et al., 2014, Hofmeister et al., 2008, Humphrey et al., 2008).

When compared to traditional interview, applicants indicate a preference for MMI (Campagna-Vaillancourt et al., 2014, Dore et al., 2010, Hofmeister et al., 2008, Humphrey et al., 2008, O'Brien et al., 2011, Rodgerson et al., 2013). International medical graduates are significantly more likely to hold this view (p = 0.01) (Humphrey et al., 2008). Age, gender and applicants' previous MMI experience did not impact on their opinions of MMI (Humphrey et al., 2008). A large Canadian study found that medical school applicants rated MMI as significantly fairer than standard interview (p=0.001) and more effective at evaluating non-academic aptitudes (p=0.016), with 45% listing MMI as the most enjoyable aspect of the selection process (Razack et al., 2009). The latter study also followed up on the views of unsuccessful applicants six months after interviewing. Even in this cohort of disappointed applicants, half of them commented that it was enjoyable (n=12), five felt it was good and two said it was fair. Negative comments included that it felt like acting (2) that the design was poor (2) and it was stressful (1).

In an in-depth qualitative study, Kumar et al (2009) established that applicants particularly valued the multiple assessment opportunities, the independence of interviewers, and the one to one format. Applicants were also impressed by the authenticity of the MMI stations and possible reduced susceptibility to coaching-both views supported by free text comments from other studies (Campagna-Vaillancourt et al., 2014, Rodgerson et al., 2013).

Applicants' misgivings with respect to MMI were relatively few. One paper based on selection to residency training, was less positive overall (Hopson et al., 2014). Although these applicants viewed MMI as an accurate assessment of communication skills (mean 3.3 out of 5) and problem solving skills (mean 3.3 out of 5) it would still negatively affect their decision to accept an offer of an

interview.   This paper was the only one that correlated performance on MMI with applicant preference and although the result was not significant, those with higher MMI showed increased preference for MMI while those with lower MMI score showed increased preference for traditional interview. Overall these applicants were more in favour of a combined approach of MMI and traditional interview rather than sole reliance on either type of interview format.

Other work identified that some applicants found MMI more difficult (O'Brien et al., 2011), and more stressful (Dowell et al., 2012), than standardised interview. Concerns that MMI favours highly communicative applicants have been expressed by medical school applicants (Kumar et al., 2009). Lastly, some applicants commented in free text questions that the allocated time is too short (Eva et al., 2004a, Kumar et al., 2009). When compared to ratings of other aspects of the MMI, applicant satisfaction with time was slightly lower in a number of studies, but was still acceptable (i.e. above the midway point on the Likert scale) (Brownell et al., 2007, Dore et al., 2010, Eva et al., 2004a, Uijtdehaage et al., 2011).

Applicant views of other interview techniques were also largely positive. Standardised interviews have been positively received by applicants in one Canadian medical school (Gula et al., 2014). Technological advances have made web based interviewing a possibility. International applicants' views (n=119) of internet based multiple mini interview (iMMI) were explored (Tiller et al., 2013). They report high levels of satisfaction with both the technology and overall interview process. Interestingly, even though given the choice the majority would opt for an in-person MMI, almost a third would favour the iMMI. Likewise Daram et al (2014) piloted the acceptability of web based video conferenced interviews with 16 residency applicants and reported over 80% satisfaction levels, with 87% of applicants recommending that web based interviewing should be an option at residency selection.

There is evidence that applicants see the interview as an important way to inform themselves of the values and ethos of the school or programme to which they are applying. In this regard, applicants appreciate the opportunity to ask questions about the school/ programme and to get information that would help them to make a decision to accept should they be offered a place (Christakis et al.,

2010, Milne et al., 2001). The vast majority (93%) believed that not having an interview was unacceptable (Milne et al., 2001). This is supported by an American study where the overwhelming majority of applicants felt that including questions about professionalism in the medical school interview strongly enhanced their impression of the values of the medical school and positively impacted on their interview experience (Kleshinski et al., 2008). One study reported favourable levels of applicant satisfaction with group interview, however international applicants felt they had a significantly harder time impressing interviewers than local candidates (p=0.004) (Patel et al., 2011).

Only one paper was identified that reported negative applicant reaction to interview (Jauhar et al., 2008). A national survey of trainee psychiatrists (n=123), half of whom had been successful in achieving a training scheme place of their choice in the previous year and half of whom had not, found that 45% (n=55) of them thought the interview process was invalid. Criticisms related mostly to inappropriate duration (37%), lack of feedback (68%) or unhelpful feedback (44%). Suggestions for improvement included that references should be available to the panel at the time of interview (n=103, 93%) and that these should be structured (n=77, 63%).

Applicant reaction to Situational Judgment Tests (SJTs) and Selection Centres (SCs) have also been documented and compared with Cognitive Tests. Belgium has been using SJTs as part of the selection of medical and dental students for many years. Two good quality studies report on applicant reaction to SJTs by comparing them to their views of cognitive tests (Lievens, 2013, Lievens and Sackett, 2006). In both studies, applicants rated SJTs as having significantly better face validity than cognitive tests (mean ratings of SJTs ranged from 3.19-4.2 out of 5, compared with mean ratings of cognitive tests which ranged from 2.76- 2.79 out of 5 respectively). It is worth noting that these were both national surveys, using validated survey instruments. In a small study in the UK, medical school applicants found SJTs relevant and valid (Husbands et al., 2014a).   A large UK study of foundation year doctors confirmed the same findings (Patterson et al., 2013).

In the postgraduate general practice training setting Koczwara et al (2012) also established that applicants were not in favour of cognitive ability tests. When comparing the face validity of two cognitive ability tests with a clinical problem-solving test (CPST) and a situational judgement test (SJT), the authors report that over 30% disagreed that Cognitive Ability Tests were fair, 35% did not see the content as appropriate, 54% did not see them as relevant, and 62% did not feel they had sufficient opportunity to demonstrate their ability. By contrast, figures from the overall 2009 general practice applicant pool (n=2947) showed that the CPST and SJT were regarded as relevant by (89%, 63%), appropriate (85%, 68%) and fair (85% and 53%), with CPST being particularly highly regarded.

Patterson et al (2011) used organisational justice theories to examine applicant reactions to selection for General Practice Training. This national UK selection process included an SJT, CPST (for shortlisting) and a SC. The SC comprised a simulated patient, group exercise and written exercise.   The authors focus on applicants' views of procedural justice. These concern job relatedness, formal test characteristics and interpersonal treatment. Applicants viewed the SJT as job relevant (mean Likert scores ranged from 8.1-9.9 out of 15) and perceived the CPST as even more so (11.3-12.1 out of 15, p<0.001). Perceptions of fairness were good for both formal test characteristics (10.6-11.3 out of 15) and interpersonal treatment (12.1-12.6 out of 15). With respect to the Selection Centre, applicants viewed all three tasks positively in terms of job relevance but the simulated patient was viewed most positively (11.9-12.5. p< 0.001).   Again perceptions of formal test characteristics (12.3-12.7) and interpersonal treatment (13.5- 13.7) were high.

These findings are mirrored in other studies, which report that applicants consistently consider SC type assessments to be fair, relevant, appropriate and to offer adequate opportunity to demonstrate skills and abilities (Gale et al., 2010, Randall et al., 2006, Samarasekera et al., 2014, Vermeulen et al., 2012).

Selection centres are not often used for medical student selection, however two examples were located.   MOR is the Hebrew acronym, for a medical student selection centre used in Israel (Ziv et al., 2008). This selection centre comprised 1) behavioural stations 2) a judgement and decision-making questionnaire and essay

writing task - typically entailing ethical or moral dilemmas and 3) a biographical questionnaire. Overall, 76% of applicants rated MOR as fair and 76% felt they had opportunity to express their capabilities to great extent. In Singapore, medical school applicants reported high levels of acceptability (92.4%) of a selection centre type involving focused skills assessment (evaluating domains including empathy, communication, integrity and resilience) and SJT (Samarasekera et al., 2014).

By contrast, the three studies identified by this review that report on applicants' views of the UK Clinical Aptitude Test (UKCAT) were much less positive. An online survey found that only a little over half of respondents (55% , n=787) thought that the test was fair, 86% thought that one can prepare for it, 44% found that advice on the UKCAT was confusing and only 44% agreed that it was testing attributes relevant to the practice of medicine (Lambe et al., 2012). The UKCAT consortium has published two surveys of applicants which report similar findings (UKCAT Consortium, 2010, UKCAT Consortium, 2011). They found that UKCAT is considered a difficult test, with applicants generally unconvinced of its relevance. Approximately 40% felt their college or school were not well informed about UKCAT and while 36% of applicants from independent schools rated the advice they got from school or college as good, only 18% of applicants from comprehensive schools agreed with this. The vast majority of applicants prepared using online practice tests (93%) and books (90%) and rated their usefulness highly.

One paper was identified that used a modified grounded theory to explore applicants' approach to the biographical essay (White et al., 2012). This entailed a review of 240 randomly selected essays and the emerging themes were discussed in face to face individual interviews with 20 applicants. Applicants expressed the idea that they had approached the essays as a way to ''show themselves'' and ''tell their own story'' in a subjective way which they felt was missing from other parts of the admission process.

In summary, applicants' views of specific selection tools have been widely surveyed with respect to certain selection tools, in particular MMI. They overwhelmingly support interview, and MMI specifically, as a selection tool,

consistently viewing it as free from bias, providing adequate information and instructions and offering the candidate adequate opportunity to demonstrate their abilities. There is strong emerging evidence that both SJTs and Selection Centres are equally well regarded. In particular, applicants view them as having high validity, good job relevance and again providing sufficient opportunity to demonstrate abilities. Perceptions of procedural fairness and interpersonal treatment appear to be important aspects of applicants' positive reactions to these tools. Conversely aptitude tests, in the small number of studies evaluating them, were not well supported by applicants.

The Views of Under-Represented and Minority Applicants

A number of studies explored the perceptions of under-represented and minority applicants. This is important work in light of the widening diversity agenda of medical schools internationally. For example, currently, approximately only 5% of applicants to UK medical school are from non-professional backgrounds (Tiffin et al., 2012). Uncovering the views of this group with respect to selection, may go some way to help understand this low level of engagement. Almost all of this evidence comes from in-depth qualitative studies.

In a series of focus groups, Greenhalgh et al (2004) investigated what going to medical school meant to school students from a wide variety of ethnic and social groups in London, UK. All students involved were aged 14-16 years, identified by their teachers as academically able and could be considered potential medical school applicants. Students from less affluent backgrounds had a superficial view of medicine, saw medicine as a culturally alien career choice, for "*posh*" people and requiring prohibitive levels of personal sacrifice. They were also less informed about the next steps should they want to apply, saw medicine as highly competitive, underestimated their chances of admission and thought that affluent students would have an advantage in the entry process.

An American study found that access to supports (e.g. health professions preparatory programme or school based premedical club) was limited for American Indians and Alaskan natives medical school applicants, but that rejected applicants reported having significantly less support (Patterson et al., 2009). They cited the lack of time to devote to volunteering or building a CV (as many were in

full time employment and often supporting families) as an additional barrier. The Medical Colleges Admission Test (MCAT) was viewed as considerable obstacle by almost two thirds of these applicants and in a separate study by 38% of African American and Hispanic students on a pre-medical preparatory course (Henry, 2006).

A striking commonality, between all the above papers, is that finance was perceived as substantial barrier for under-represented and minority applicants, to both applying and being selected to medical school. Rejected applicants were even more likely to view finance as an obstacle ($p<0.05$) (Patterson et al., 2009). The cost of applying, preparatory courses for MCAT, attending open days and interviews, and were all noted as substantial. In summary, under-represented and minority applicants perceived many obstacles to selection to medicine: lack of adequate supports, failing to achieve high enough MCAT scores, socioeconomic background and finance.

The Views of Interviewers, Faculty and Admissions Committee Members

A number of studies report on the views of stakeholders who are directly involved in the selection process. This included persons from a variety of backgrounds joined by a common responsibility for some discrete aspect of the selection process. It included members of faculty, residency or training programme directors, clinicians, clinical teachers, admissions deans, admissions committee members, medical students and increasingly lay or community members. These individuals served variously to shortlist written applications, letters of reference, personal statements; serve on interviews- including MMI, or assess situational judgement tests or selection centres. Research with these stakeholders is presented into two categories. The first relates to their views of the core values and qualities of selection processes. The second establishes their views of specific selection tools.

Views on Values of Selection Processes

Marrin et al (2004) conducted a study in one Canadian medical school to identify the key qualities of the selection process and their relative importance to faculty (n=70), students (n=123) and community (n=84). Using a paired comparison approach, participants had to rank order the most valued qualities of

an ideal selection process. There was no significant difference between the rankings of the three groups of stakeholders. Fairness, validity and comprehensiveness were the most valued qualities (mean z score 0.92, 0.87 and 0.44 respectively). Affordability and making a public statement about the values of the medical school were valued least (mean z score -0.8, -0.9). This information was then used to redesign the selection processes for the medical school.

Through a series of qualitative interviews with 12 community members, Stagg and Rosenthal (2012) established why they want to participate in medical student selection. Overwhelmingly, community members saw their involvement with the university in the selection of students into medical school as positive. Five themes emerged -: opportunity for professional growth; for personal growth; responsibility to represent the broader community; protecting the student and public interest and self-interest in shaping the future workforce. They believe that the university gains both financially and politically by their involvement. There was also a strong sense of accountability.

In summary, Interviewers, Faculty and Admissions Committee Members are made up of persons from a wide variety of backgrounds, both clinical and non-clinical, with a shared common responsibility for particular aspects of selection. Fairness, validity and comprehensiveness are viewed as important crucial aspects of the selection process by these stakeholders. A strong sense of social accountability motivates community members and lay persons to be involved in the selection process.

Views on Selection Tools

Monroe et al (2013) surveyed the Admissions Deans of 142 medical schools in North America, to describe the admission processes and to determine admission deans' view of the relative importance of different admission data at each stage of the selection process. All schools reported that they interviewed, with 59% conducting two interviews per candidate. The rating of selection tools differed depending on the stage of the selection process. The MCAT and Grade Point Average (GPA) were considered most important for shortlisting but less important in the decision of who to admit. At the admission stage, the interview (mean rating 4.5 out of 5) and letter of recommendation (3.7 out of 5) were more valued

than either (GPA (3.6), community service/volunteering (3.5), MCAT (3.4) or personal statement (3.3)). The survey also showed that, overall there was much higher emphasis given to non-academic characteristics than the previous survey 20 years earlier. Elsewhere, Adams (2009) established that admissions committees rank the equivalent academic record from a university background, as superior to that from a community college, and both the aforementioned as superior to that from an academic degree awarded following an online course.

A large study (including applicants , medical students, patients and doctors involved in the training of medical students -n=938) was conducted to identify views on the perceived validity of the tools used for selection of medical students in Australia – namely GPA, UMAT, GAMSAT and interview (Brown and Griffin, 2012). It found that mean level of confidence in selection tools was low for all groups, with medical professionals having lower confidence than other stakeholders. Tests of cognitive ability were viewed as the least valid and interview as the most valid. Perceived validity was associated with the extent to which the test was thought to be susceptible to bias.

In a separate survey, the Deans of Student Affairs /Medical Admissions of 144 American Medical Schools were asked for their opinions on barriers to selection of under-represented and minority applicants (n= 86 response rate 59%) (Agrawal et al., 2005). They identified low MCAT scores as a very significant barrier (90%) along with low grade point average (60%), absence of role models (77%), not enough minority faculty members (71%) and lack of financial aid (48%).

Taken together, these studies suggest that tests of cognitive ability and academic record are not held in the same regard by this stakeholder group as other selection tools, and are viewed as potential barriers to some students, possibly susceptible to bias and more appropriate for shortlisting than for selection.

A qualitative study in the UK examined the reasons selectors gave for accepting or rejecting an applicant before interview based on their UCAS application – comprising a personal statement, description of work experience and the teacher's reference (Turner and Nicholson, 2011). The selectors were all trained and comprised a mixture of clinical and non-clinical members,

administrative school staff and lay people. The most common reason for acceptance or rejection was medically related work experience, with the level of reflection on the part of the applicant, demonstrating an understanding of medical career, being judged as more important than type or duration of work. Teacher references were viewed as influential, especially for rejection but were hard to interpret as they involved guess work and reading "between the lines". Personal statements were deemed useful but considered highly subjective.

Perceptions of letters of reference were also examined in two other studies from USA, where they were viewed as most helpful when they were factual, descriptive and cited examples of specific behaviours (Johnson and Elam, 2001). In the case of postgraduate selection, they were considered more valuable when they were written by a clinician known to the assessor (Kaffenberger et al., 2014). Shortcomings of letters of reference included difficulty in ascertaining the strength of recommendation and reluctance on the part of the writer to give honest account of candidates' weaknesses (Kaffenberger et al., 2014).

Views on interview as a selection tool have been frequently surveyed, in particular, the views of MMI interviewers. Numerous studies clarified that interviewers comprised Faculty, students, non-clinical medical school staff, community members and laypersons- however no study was identified that reported the views of these stakeholder groups separately. As with applicants, interviewers generally perceived the MMI to be fair. There were several reports of this from good quality studies. (Mean fairness ratings 6 out of 7 (Campagna-Vaillancourt et al., 2013) 4.4 out of 6 (Humphrey et al., 2008) 3.93 out of 5 (Hoftmeister et al., 2008), 5.2 out 6 (Razack et al., 2009), 6.2 out 7 (Uijtdehaage et al 2011), 90% agree or strongly agree that MMI is fair (Dowell et al., 2012)).

 Importantly, MMI interviewers felt that they were able to get an accurate portrayal of applicants and that the scoring mechanisms allowed them to adequately differentiate between them. The evidence includes mean ratings for ability to differentiate between applicants of 4 out of 5 (Brownwell et al., 2007), 4.7 out of 7 (Campagna-Vaillancourt et al., 2013), 3.6 out of 5 (Hoftmeister et al., 2008), 5.7 out of 7 (Uijtdehaage et al., 2011), a mean rating 5.7 out of 7 for an accurate portrayal of candidates (Eva et al., 2004a) and 90% agree reasonable

portrayal of candidate (Dore et al., 2010). Kelly at al (2014a) found 75% (n=18) of interviewers felt that MMI was relevant, reasonably tested candidates' ability (79% n=19) and 62.5% (n=15) agreed that the content was sufficiently important to warrant inclusion in a selection test. Razack et al (2009) report that interviewers found MMIs appropriate for use with home and international applicants. There is also evidence that interviewers favour MMI over traditional interview (Campagna-Vaillancourt et al., 2013, Humphrey et al., 2008).

Also notable, in the case of a newly introduced selection tool, interviewers generally reported that they were confident and happy with the pre-interview information, clarity of instructions and time allocated (Brownell et al., 2007, Campagna-Vaillancourt et al., 2014, Eva et al., 2004a, Hofmeister et al., 2008, Uijtdehaage et al., 2011). In qualitative research Kumar et al (2009) established that interviewers felt much less anxious about their decision making due to the multiple assessment opportunities and appreciated the multidimensional view.

As part of a larger study, Eva and Macala (2014) sought to discern if variations in the test characteristics of MMI altered views of participants. Four MMI scenarios were presented three different ways: situational judgement type stations, behavioural interview type stations, and free form stations, which were unstructured in that the examiner was only given a brief explanation of the station purpose. Interviewers found that they were more able to generate an accurate portrayal of applicant by the SJT type station (p<0.05) but there was no significant differences between the station types in terms of difficulty, clarity of instructions, and available time.

Interviewers' concerns regarding MMI, include a fear that it might be primarily measuring communication skills (Kumar et al 2009), that candidate differences such as cultural background, personality or language may affect their performance (Kelly et al., 2014a (Chapter7), Razack et al., 2009), a lack of opportunity for MMI assessors to bench their marking against peers (Kumar et al., 2009), more time needed for interviewer to calibrate themselves (Hofmeister et al., 2008), the process was somewhat impersonal (Lubarsky and Young 2013) and that interviewers needed more specific training (Dowell et al., 2012).

Two studies reported assessors' views of SCs and found that they were favourable (Gale et al., 2010, Mitchson et al., 2009). Assessors rated it positively for relevance (3.8-4.7 out of 5), fairness (3.9-4.4), opportunity to demonstrate ability (3.9-4.7) and appropriateness to selection (4.9-4.7), with the simulated station being rated significantly higher (Gale et al., 2010). In addition, Mitchson et al (2009) found that 69% believed it was superior to traditional interview. Negative findings were few but included complaints about the inflexibility of the structured approach.

In summary, there is good evidence that this stakeholder group finds interview and in particular MMI fair, relevant and appropriate for selection, and emerging evidence for positive reactions to SCs. There is much less support for aptitude tests and academic record, which are sometimes viewed as lacking validity and acting as barriers to certain groups of applicants. The usefulness of letters of reference seems mostly to be for ruling applicants out rather than in.

The Views of Medical Students

Twelve studies were identified that explored the views of medical students, distinct from those where students were directly involved in the selection process as in the group above. These are synthesised under a) views of selection tools and b) under-represented and minority students.

Medical Students' Views of Selection Tools

A good quality mixed methods study (MERSQI rating 10.8) of first year medical students in five Scottish medical schools revealed that overall, the UK Clinical Aptitude Test (UKCAT) was poorly viewed (Cleland et al., 2011). Focus group interviews showed that students felt it lacked face validity, had poor predictive validity, was coachable, was discriminatory against less affluent applicants and there was lack of certainty about how the test was applied by medical schools. In the written survey (n=883, response rate 88%) only 20% of students agreed UKCAT was useful while 12% found it difficult and 42% found it moderately stressful.

Similarly, in a survey of two medical schools in New Zealand (n=1325, response rate 65%) the majority of students were unconvinced of importance of Undergraduate Medical and Health Professions Admission Test (UMAT) (Dhar et

al., 2012). Fifty-six percent thought UMAT was "not" or "not really" important, 67% believed it was not fair, 54% felt it assessed non cognitive attributes either "not really" or "not at all" and 81% found it either stressful or very stressful – this being significantly higher in the school that gave higher weighting to the UMAT. This contrasts with findings on a similar selection tool, HPAT-Ireland which has mixed student reaction. In one study of 175 First Year Medical students, 76% felt it fair, 37% felt it was easier for males, 70% felt questions well designed and relevant, except for the Non-verbal reasoning section which 32% felt was poorly designed and irrelevant (Stevens et al., 2014). While in a separate study, Kelly et al (2014a) found that only 38% found HPAT-Ireland relevant (see Chapter 7). Aptitude testing in general was found to be one of the least suitable selection tools (with only letters of reference and work experience being thought of as suitable by less students).

Two studies reported medical students' views on the value of coaching for aptitude tests. Wilkinson and Wilkinson (2013) found that the three commonest forms of preparation for UMAT were ACER practice materials (used by almost all students), a commercial coaching course (MED Entry) and student led tutorials. While the authors found no significant difference in overall performance attributable to coaching, they found that students who had taken a commercial course had significantly higher confidence (mean difference in Likert score rating for confidence 0.6) and that there was a moderately strong positive correlation between the amount of money spent and confidence (r=0.3, p<0.001). A finding that resonates with work by Stevens et al (2014) who reported the vast majority (79%) of those who had accessed commercial coaching (for HAPT-Ireland), felt it improved their performance and 49% would recommend it to a friend.

There is evidence that students prefer interview to aptitude testing (Kelly et al., 2014a, Stevens et al., 2014). International students are even more likely to support interview (p<0.01) (Stevens et al., 2014). In a study to establish the feasibility of MMI, Kelly et al (2014a) found that the majority of students thought the MMI was relevant and suitable for use in selection (See Chapter 7). Favourable free text comments included that MMI provided a "*…more wholesome picture*".   A criticism was inadequate time, and a concern that interviewers might be subjective in their marking.

Separately, Kumwenda et al (2013) surveyed medical students about their perceptions of the veracity of Universities and Colleges Admissions Service (UCAS) form. This form includes biographical information and a personal statement and is a required part of the admission process in the UK. Sixty-six percent of students suspect peers stretch the truth on their UCAS form, 16% believe deceptive practice is common (older students, those with lower UCAS scores more likely to agree $p<0.05$) 84% felt lying on UCAS is unacceptable even if other people are doing it (mean 1.8). Ninety-four percent believe exaggerating on UCAS is dishonest but 14% believe it is part of the admission game (males more likely to agree $p<0.05$) and 13% believe it is necessary / understandable.

In the qualitative strand of a mixed methods PhD thesis Wright (2012) interviewed thirteen medical students about their experience of applying to medicine. Family and the type of second level school attended were perceived as highly influential, both on the decision to apply and preparing the application. For example, being from a medical family was seen as a significant advantage in the gaining access to relevant work experience. Likewise, fee paying or independent schools were considered to be more proactive and supportive than state schools-offering practical guidance for interview preparation.

In summary, there were a small number of studies exploring students' views of selection tools. Based on what is available to date it appears that students prefer interview based selection to aptitude tests. They seem to view the latter as lacking in relevance, prone to bias and susceptible to coaching. They are also unconvinced about the transparency of written applications, including biographical essays and personal statements where they believe exaggeration is both common practice and a necessary part of the selection game. Students consider that access to supports in the application process is not equal for all applicants.

The views of Under-Represented and Minority Students

A UK based qualitative study explored mature students' experience of applying to medical school (Mathers and Parry, 2010). Mature candidates made strategic applications to medical schools that they felt were open and diverse. They highlighted the importance of open days and interviews to their sense of

identity and fit with prospective medical schools. University location was a key determining factor, as attending a local medical schools ensured ongoing access to crucial family and social support. As a group, they viewed the written application system (Universities and Colleges Admissions Service- UCAS) inflexible for mature applicants and there was a lack of transparency about what would constitute a good mature application. They highlighted the financial strain involved and linked this to the perception of applying to medicine as a '*risky business*' – often requiring the applicant to make commitments such as giving up existing paid employment in order to attend preparatory courses or build a CV without the certainty of a place at medical school ultimately.

A second study examined the views of students admitted through a widening access route, on the role of traditional interview for selection through this route (Rich, 2011). Interestingly students in their early clinical years supported traditional interview while students in the senior years felt that a multiple mini interview would be more appropriate.

The Views of Other Stakeholders

Three studies were identified that explored the views of other stakeholder groups. All three were conducted in Ireland and relate to the recent introduction of HPAT-Ireland.

In a national survey, O'Flynn et al (2013a) set out to establish views of career guidance counsellors on HPAT-Ireland and reforms to selection to medicine in Ireland. They report that 52% (n=97) were in favour of the reforms to selection (See Chapter 3-The Irish Context), 49% (n=92) felt that the new system fair. However, those who opposed it were worried about the negative impact of HPAT-Ireland on socially disadvantaged students. Overall they felt Section 2 (Interpersonal understanding) was most relevant and Section 3 (Non-verbal reasoning) least relevant to future job performance.

Elsewhere, Dennehy et al (2013) surveyed general practitioners who were not directly involved in selection, for their views of selection to medicine in Ireland and knowledge of HPAT-Ireland. While the majority (70%) supported the use of aptitude tests and academic record (97%) in principle, 32% disagreed or strongly disagreed with the changes to selection in Ireland. The majority strongly

supported academic record as a selection tool (118 respondents, 97%). Over 70% felt socioeconomic status would affect academic performance and 66% thought it would affect applicant performance on HPAT-Ireland.

Kelly et al (2014b) explored the views of fifteen doctors, from a variety of clinical backgrounds on HPAT-Ireland (See Chapter 6). On the whole this test was considered to have a moderately good degree of job-relatedness. However a non-verbal reasoning section was criticised, by all participants, for lacking clinical relevance. Doctors did not perceive the test to have good predictive validity. However half still felt it was acceptable as a selection tool in conjunction with academic record. Some doctors were undecided. Those who found it unacceptable were influenced by its perceived narrow focus, limited job relatedness, potential for socioeconomic bias, impact on gender and potential for negative influence on student diversity.

In summary, these three studies demonstrate that there has been mixed reaction to the introduction of an aptitude test to selection in Ireland, with both guidance counsellors and doctors supporting the use of aptitude tests in principle but with varying levels of approval of the actual reforms. Concerns re socioeconomic bias are uppermost.

2.4.4 Discussion

This review and synthesis of the evidence identifies a growing appreciation of the importance of understanding the views of stakeholders. Certainly, with respect to newer selection tools, such as multiple mini interview, the evidence indicates that those responsible for its introduction value and are keen to ascertain participants' perceptions. The emerging evidence demonstrates that there is good concordance of views between stakeholder groups. Generally, the research focuses on three main stakeholder groups: a) Applicants; b) Interviewers, Faculty and Admissions Committee Members (including community members and laypersons) - who are involved in some aspect of the selection process and c) Medical Students.

Stakeholders are clearly supportive of interview as a key part of the selection process to medical school and residency/ higher professional training programmes. In particular, MMI is widely accepted. The fact that applicants and

interviewers view MMI as providing adequate opportunity for candidates to demonstrate their ability and allows for differentiation between candidates is likely to be a key factor in acceptability. Secondly, the perception that MMI is free from gender or cultural bias is important and in keeping with the organisational justice theory of distributive justice.

Selection centres and SJTs are also highly acceptable to stakeholders. They represent high to medium fidelity assessments and the job relatedness of these tools is perceived as high by applicants and assessors. This matches with evidence from elsewhere that candidates have a preference for job related selection tools (Patterson et al., 2012b).

Aptitude tests, including MCAT, UMAT, HPAT-Ireland and UKCAT, enjoy mixed stakeholder acceptability. Underrepresented and minority applicants view them as barriers, while applicants and medical students question their fairness, face and predictive validity. Of concern to students and Faculty, is the perception that they are susceptible to coaching, and the associated fear that this may lead to economic bias. While the research shows coaching is of limited to no value in terms of performance, this perception of injustice is an example of negative consequential evidence and contravenes the principle that selection tools should do more good than harm (Downing, 2003). Selection tools that are *perceived* as unfair can deter potential medical students from applying which would be considered a profoundly negative consequential effect (Lambe et al., 2012, Patterson et al., 2012a).

Other tools such as letters of recommendation and personal statements are viewed with some reservation about the veracity of content. Their predictive validity is also known to be poorer than alternative tools, and coupled with poor stakeholder acceptability, questions their role in the selection process (Cleland et al., 2012).

This review also emphasises gaps and shortcomings in the research evidence of stakeholder views of selection to medicine.   These related to methodology and scope of stakeholder research.

Chapter 2: Review of Literature

With respect to methodology, this review revealed an over-reliance on quantitative research methods. Qualitative research, on the other hand, is ideally suited to understanding the meaning of selection for the respective stakeholder groups and can greatly add to our understanding of the needs and attitudes of stakeholders. The use of theoretical models, to conceptualise and interpret stakeholder views, was rare in this review, but again can help us to better appreciate and compare the nuances of stakeholder acceptability.

Within the quantitative paradigm, the need for the development and use of standardised tools for collating stakeholder views became clear. The use of standardised tools would better facilitate higher standards of reporting of the content and internal validity of the evaluation instrument, and would accommodate comparisons between different stakeholder settings. Equally there is a need for better prioritisation of stakeholder views as a legitimate aim for selection research. For many studies, that included valuable stakeholder opinions, this was listed only as a lesser aim and in some cases not even mentioned in the abstract. Future research should consider this, given the centrality of stakeholder views to the political validity of selection tools.

There was a tendency to rely on exit poll type surveys, particularly with respect to applicants. While this is understandable from a logistics viewpoint, this methodology has three shortcomings: a) these have to be short and not taxing for applicants to complete following the assessment, which limits the depth of information that they can provide b) applicants at this stage are unaware of the outcome and their opinion may differ depending on whether they were successful or not c) they do not allow for a period of reflection. Future research should aim to follow up the views of unsuccessful applicants.

This review also highlighted that there is a need to broaden the scope of this research. For example, no studies were identified during the time frame of this review, exploring participants' views of personality assessments and only scant exploration of academic record which confirms that there are gaps in our understanding of stakeholder views.

This review also highlights a need to seek the views of a wider pool of stakeholders. For example, no studies exploring the views of parents were found,

yet they are substantially invested in the application process. Also, there was limited research on the views of patients, or general public or members of the medical profession outside of those directly involved in the admission process or clinical teaching. Similarly, there was only one study of career guidance officers identified yet this group has been noted to be potentially very influential on applicants' preparation for admission (Lambe et al., 2012, Wright, 2012). Additionally, while there have been many studies of stakeholders' views, for the most part, each group is treated as if it is homogenous. For example, there were very few studies that compared and contrasted the differences in opinions between interviewers from different backgrounds. Within the applicant research, we saw that the experience of some subgroups differed depending on socioeconomic and cultural backgrounds. Future research should be mindful of these issues and seek to sensitively explore views in a manner that accommodates both differences and similarities within stakeholder groups.

Directions for Future Research

Future research should carefully consider the choice of methodology and efforts should be made to design good quality mixed methods or qualitative research to facilitate deeper insights into stakeholder views of a wider range of selection tools.

Limitations and Strengths

A limitation of this review is the restricted time frame. It is possible, that important studies may have been missed.

A second limitation is the relatively low MERSQI score for included papers. There were a number of reasons for this. A key reason was that, by virtue of all the studies included in this review being based on participant views/ opinions, this meant that the maximum score any paper could score for outcome was 1 out of 3.

Secondly, with respect to mixed methods studies, the score refers to the quantitative strand only. For many studies, collating the stakeholder view was a subsidiary aim or incidentally reported and sufficient details of the methodology to adequately complete the checklist were not included. Similarly, the restricted word count of conference abstracts resulted in very low scores. However recognising the time-lag from research completion to peer reviewed publication,

the inclusion of conference abstracts was acceptable to ensure that recent evidence was included.

A third limitation is that, at the study selection stage, the second reviewer assessed a random sample of 10% of *excluded* citations. This would have been better if a random sample of 10% of *all* citations had been used, as there is a risk of bias by knowledge that these titles were already deemed not eligible. However, both authors discussed the application of the inclusion and exclusion criteria after this random sample was checked and then the principal researcher (MK) rechecked all titles.

Lastly, in order to minimise bias, it would have been ideal practice to enlist an impartial reviewer to apply the MERSQI to papers on which MK and AWM were co-authors. This step will be taken prior to submitting this chapter to a peer reviewed journal.

Conversely, this review had a number of strengths. The search strategy (which included many sources of grey literature and not applying language restrictions during the searching phase) helped to ensure that valuable studies were not missed and limited the likelihood that publication bias would exclude studies with negative outcomes. The level of agreement between the two reviewers, who independently assessed the studies for inclusion, demonstrated that the inclusion and exclusion criteria were unambiguous. The use of a previously validated quality assessment tool was helpful in extracting the strengths and weaknesses of different studies and allowed for a comparison of the strength of evidence for the different stakeholder perspectives.

2.4.5 Conclusions

Stakeholders are a diverse and influential group. It is critical to the operation of fair and defensible selection processes that we understand and appreciate the range and depth of views that they hold. This review demonstrates that there is important work being done in this field, especially in respect to applicants. However, it highlights the need for better standards and more appropriate methodologies; for broadening the scope of the stakeholder groups included and for recognising that stakeholders even from the same group are not necessarily homogenous.

Table 7 Summary of the Research Evidence of Stakeholder Views of Selection to Medicine

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Adams 2009 * Research paper	USA Medical Schools	Quantitative	Academic record	Medical school administrators	Postal questionnaire survey.	Offer of interview	University academic record preferred over both community college course and online course.
Agrawal et al 2005*Research paper	USA Medical Schools	Quantitative	MCAT/ GPA / Other factors	Deans/ Medical Admission	Postal questionnaire	Percentage responses	MCAT, GPA , lack of financial aid, lack of role model barriers to under-represented/ and minority applicants
Brown & Griffin 2012 * Abstract	Australia 1 medical school	Quantitative	Academic record/ UMAT or GAMSAT and interview	Mixed stakeholder group-comprising applicants medical students patients and doctors	Questionnaire	Perceived validity, familiarity, overall confidence in selection	Confidence in selection tools low for all groups, even lower in medical professionals. Cognitive ability tests least valid/ interview most valid.
Brownell et al 2007 Research paper	Canada 1 Medical School	Quantitative	MMI	Applicants and Interviewers	Questionnaire	Mean ratings and standard deviations	MMI very acceptable to both interviewers & applicants. Free from gender, cultural bias, adequate time, orientation, differentiation. Fair.
Campagna-Vaillancourtet al 2013 Research paper	Canada Postgraduate residency programme – Ears, Nose Throat, Head & Neck	Embedded mixed methods	MMI	Applicants and Interviewers	Questionnaire	Mean ratings and standard deviations of	MMI very acceptable to both interviewers and applicants. Free from gender, cultural, age bias. Preferred over traditional interview
Christakis et al 2010 Research paper	Canada Postgraduate residency – ophthalmology	Quantitative	Interview day	Applicants	Questionnaire	Frequency of responses.	Interview very valuable, opportunity to express ideas/opinions, increased their likelihood of accepting a place.

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Cleland et al 2011* Research paper	Scotland 5 Undergraduate Medical Schools	Mixed Methods	UKCAT	1st Yr Medical students	Focus groups & Questionnaire	Framework analysis Percentage responses	UKCAT poorly viewed, lacking face validity, poor predictive validity, coachable, discriminatory against less affluent applicants. Only 20% s agreed that UKCAT was useful
Daram et al 2014 Research paper	USA Medical fellowship – Gastroenterology	Quantitative	Web based video interview	Applicants	Questionnaire	Percentage responses	Majority felt web based interview met expectations, and should be an option in fellowship interviews. ¼ equivalent/better traditional interview
Dennehy et al 2013* Research paper	Ireland Medical School Entry	Quantitative	HPAT-Ireland/ Academic record	General practitioners	27 item Questionnaire survey	Percentage responses	Good support for use of aptitude tests in principle but 30% unhappy with reforms to selection. Concerns re socioeconomic bias of academic record and HPAT-Ireland
Dhar et al 2012* Research paper	New Zealand 2 Undergraduate Medical Schools	Quantitative	UMAT	Medical students	Questionnaire survey	Percentage responses	56% UMAT was not important, 67% was not fair, 81% stressful, 54% assessed non cognitive poorly attributes either not really or not at all.
Dore et al 2010 Research paper	Canada Post graduate residency – OBGYN & Paediatrics	Quantitative	MMI	Applicants- MMI assessors	Questionnaire survey	Percentage responses	Almost 90% of applicants and assessors felt candidates had good opportunity to portray themselves, adequate time. ¾ assessors MMI better than traditional interview
Dowell et al 2012 Research paper	Scotland 1 Undergraduate Medical School	Quantitative	MMI	Applicants Interviewers –	Online Questionnaire	Percentage responses	94 % of applicants MMI fair, 33% MMI more stressful than standard interview 90% interviewers MMI fair, 23% more specific training
El Says et al 2013 Research paper	Saudi Arabia 1 Medical School	Quantitative	MMI	Applicants and interviewers	UTD	UTD	MMI was acceptable to both students and faculty.
Eva et al 2004a Research paper	Canada 1 medical School	Quantitative	MMI	Applicants & Interviewers	Questionnaire	Mean ratings and standard deviations	Applicants and assessors felt that candidates could accurately portray themselves, clear instructions, adequate time

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Eva & Macala 2014 Research paper	Canada 1 medical School	Quantitative	MMI- three variations of station type SJT, behavioural, Free form	Applicants & Interviewers.	Questionnaire	Mean ratings, standard deviations	Applicants found free form stations more anxiety provoking (p<0.05), and difficult (p<0.01). Interviewers more able to generate an accurate portrayal of applicant by the SJT type station (p<0.05)
Gale et al 2010 Research paper	UK Postgraduate training – anaesthetics	Quantitative	Selection Centre	Applicants & Assessors	Questionnaire	Mean and standard deviations	All four tools were positively rated by applicants & assessors for relevance (3.6 – 4.7), fairness (3.9-4.4), demonstrate ability (3.6-4.2), Assessors: highly appropriate
Goulston & Oates 2009 Report	Australia 1 Medical School	Mixed methods	MMI, GPA, GAMSAT	Applicants, medical students, Faculty, alumni, Health Services, clinical training, professional bodies & more	Focus groups, submissions, email surveys, face to face interviews-	Medical School report - 27 recommendations	Commitment to widening diversity, MMI as interview tool, including community interviewers, GAMSAT to ranking for call to interview
Greenhalgh et al 2004 * Research paper	UK 6 London secondary Schools	Qualitative	Not specified	Academically able school students aged 14-16, from mixed social/ ethnic groups	Focus groups	Pupils perceptions of medical school and motivation to apply	Less affluent students felt medicine for "posh" people, less informed about application, underestimated their chances of getting in. Finance a huge barrier.
Griffin et al 2008 Research paper	Australia 1 medical school Medical School	Quantitative	MMI and UMAT	Applicants	Questionnaire survey	Perceptions of the usefulness of coaching , previous interview experience and practice run on MMI	Those who had attended UMAT coaching rated it more helpful than those who had not (P= 0.001). A MMI practice run considered most effective way to prepare and coaching least helpful for MMI.
Gula et al 2014 Abstract	Canada 1 medical school	Quantitative	Standardised interview	Applicants and interviewers	Survey	Views on atmosphere of interview, confidence in interview,	Standardised interview positively received by applicants and interviewers
Harris & Owen 2007* Research paper	Australia – 1 medical school	Mixed methods	Non-cognitive characteristics MMI	Medical students, early graduates, health academics, clinical health workers and administrative staff. Also surveyed MMI applicants	Using q methodology	Ranking of statements	Characteristics Love of medicine and learning, groundedness, self-confidence, balanced approach, mature social skills and realism. MMI based on these was well received by applicants.

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Henry 2006 * Research paper	USA 1 premedical preparatory programme	Quantitative	MCAT/ GPA	97 premedical students African Americans Hispanic	Perceived Educational and Career Barriers Inventory	Mean responses to inventory items.	Barriers- Not having a high enough GPA and MCAT were seen as the most significant barriers
Hofmeister et al 2008* Research paper	Canada Residency Programme- Family Medicine at 2 Medical Schools	Embedded mixed methods	MMI	Applicants and Interviewers	Survey with quantitative and qualitative components	Analysis of Likert scale responses and content analysis	Applicants: Preferred MMI over other interviews, free from culture/gender bias. Interviewers: well prepared/ fairness (3.93/5)/ ability to differentiate (3.61/5). MMI helpful assessing professionalism. Needed more time to calibrate.
Hopson et al 2014* Research paper	USA 3 Emergency medicine training sites	Quantitative	MMI	Emergency Medicine (EM) interns	Pre and post experience surveys	Mean Likert responses	MMI would negatively influence decision to accept offer of interview 2.7 (out of 5) pre and 2.8 post MMI. More in favour of mixed MMI and traditional interview. Higher MMI scores had preference for MMI (not significant).
Humphrey et al 2008* Research paper	UK 1 post graduate deanery Paediatric training programme	Quantitative	MMI	Applicants & Interviewers	Questionnaires- Reliability Cronbach alpha 0.88 &.62	Means, standard deviations and Confidence intervals	Applicants: MMI fairer than traditional interview 4, preferable to traditional interview 3.7, (IMGs preferred MMI significantly more). Interviewers- MMI better than traditional interview (4.8).
Husbands et al 2014 (a)* Abstract	UK Undergraduate Medical school	Quantitative	SJT	200 Medical School applicants response rate = 36.2%	UTD	Applicants perceptions of relevance and validity of SJT	Most applicants strongly agreed that SJT appeared relevant and valid
Jauhar et al 2008* Rsearch paper	Scotland National survey	Quantitative and open ended comments	Shortlisting / interview and feedback	123 Doctors on Psychiatry training programme	Questionnaire	Percentage responses to Likert scale questions	76% lack of confidence in shortlisting process. 45% interview invalid (unsuccessful candidates significantly more so) 92% felt references should be available at interview with 63% supporting structured references

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Jayasuriya et al 2012 * **Abstract**	UK 1 medical school	Qualitative	Not specified	Medical Students	Focus groups	Students perceptions	Students were aware of the components of selection but unsure how they were used. Inconsistency in student advice. Preferred non-academic interviews that used personal statements and communication scenarios
Johnson & Elam 2001*Short research report	USA 1 Medical School	Quantitative	Letters of recommend-ation	Admission committee members & Premedical advisors	10 item checklists set.	Perceptions of usefulness	Both thought most helpful when they factual, descriptive and cited examples of specific behaviours.
Kaffenberger et al 2014* Letter – original research	USA National Survey	Quantitative	Letters of recommend-ation (LOR)	Professors of Dermatology -	Survey	Percentage responses	LOR from Dermatology Professors and "Physicians I know" more reliable than other sources. Frequently have difficulty in ascertaining the strength of recommendation and reluctance to give honest account of weaknesses
Kelly et al 2014 a* **Research paper**	Ireland 1 Undergraduate medical School	Quantitative	MMI / HPAT-Ireland,	First Year Medical Students MMI Interviewer	Electronic questionnaire	Percentage responses	Students: Relevant: 90% MMI, 60% traditional interview, 38% HPAT-Ireland. Suitable selection tool: 73% MMI, 79% academic record. Interviewers: 75% MMI was relevant, 71% MMI would be a useful addition to selection
Kelly et al 2014 b* **Research paper**	Ireland 1 Undergraduate medical School	Qualitative	HPAT-Ireland	Qualified doctors from various disciplines (n=15)	Interviews – analysed using principles of grounded theory	Perceptions of job relevance, acceptability of HPAT-Ireland	Sections 1 and 2 perceived to have good job-elatedness, but Section 3 criticised. Split views on acceptability, with those opposed being principally concerned re possible negative impact on diversity.
Kleshinski et al 2008* Research paper	USA 1 medical school	Quantitative	Interview	Interviewers & Applicants	Survey	Percentage responses to questionnaire items	Applicants: 74% asking about professionalism positively impacted & enhanced view of medical school. Interviewers: 50% felt applicant response impacted on assessment of applicant. Applicants more positive re importance of including professionalism p= 0.0001

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Koczwara et al 2012 Research paper	UK Post graduate GP training in one geographical area	Quantitative	Cognitive ability tests clinical problem-solving test (CPST), situational judgement test (SJT)	Applicants	Validated candidate evaluation questionnaire	Percentage and frequency of responses	Cognitive ability tests: 30% not fair-fair, 35% content not appropriate, 54% not relevant. By contrast figures from the overall 2009 GP applicant pool (n=2947) showed that the CPST and SJT were regarded as relevant by 89%, 63%, appropriate 85%, 68% and fair 85% and 53%.
Kumar et al 2009 * Research paper	Australia & Canada 2 Graduate Entry Medical Schools	Qualitative	MMI	MMI Interviewers & Applicants	6 Focus groups and open-ended survey	Framework analysis	Very positively viewed. Candidates valued the independence of the interviewer & multiple opportunities. Interviewers less anxious about decision making, but concerns re measuring communication skills and lack of opportunity to bench their marking
Kumwenda et al 2013* Research paper	UK 6 Medical schools and 1 dental school	Quantitative	UCAS	First year entrants to medical and dental school	Online questionnaire	Average and percentage responses Cronbach alpha 0.77.	66% suspect peers stretch the truth, 16% deceptive practice is common, 84% lying unacceptable, 949% exaggerating on UCAS is dishonest but 14% think part of the admission game (males agree more p<0.05)
Lambe et al 2012 Research paper	UK 1Medical School	Quantitative	UKCAT	Applicants	Online questionnaire	Percentage responses	86% thought that you can prepare for the UKCAT, 44% felt that advice on the UKCAT was confusing, 55% felt test was fair and 44% agreed it was relevant
Lievens 2013 Research paper	Belgium National survey Medical and dental undergraduate	Quantitative-longitudinal multiple cohort study (1999-2002)	SJT & Cognitive tests	Applicants	Validated questionnaire	Mean, Standard deviation of responses	Face validity of the SJT (3.19 SD 0.88) was significantly higher than cognitive test (2.76 SD 0.68) p<0.01. SJT as significantly less difficult than the cognitive tests.
Lievens & Sackett 2006 Research paper	Belgium National survey Medical and dental undergraduate selection	Quantitative	Two formats of SJT video based versus written formats & Cognitive tests	Applicants from two cohorts	Validated questionnaire Cronbach alpha (0.66, 0.76)	Mean, Standard deviation of responses	No significant difference between the mean face validity perceptions of the video SJT (3.41) and written SJT (3.44). Both significantly higher than face validity of the cognitive test (2000= 2.75, 2003 =2.79).

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Lubarsky &Young 2013 Abstract	Canada 1 hospital Neurology residency program	Quantitative	MMI	Applicants & Interviewers	UTD	UTD	Both applicants and interviewers felt MMI allowed applicants to showcase their unique attributes and skills, but that the process felt somewhat 'impersonal'
Marrin et al 2004* **Research paper**	Canada 1 Medical School	Quantitative	Qualities of selection process	Admission stakeholders- Faculty, students and community mean	Paired comparison approach-	Z scores for probability of each characteristic being chosen from the pairing	No significant difference between stakeholders. Fairness (mean z score 0.92), Validity (mean z score 0.87), comprehensiveness (0.44), accessibility (0.1), defensible (-0.3), leads to diversity (-0.31), affordable (-0.8), public statement (-0.9).
Mathers & Parry 2010* Research paper	UK 3 Medical Schools	Qualitative	Not specified	Older mature students	Unstructured one to one interviews	Framework analysis	Consideration & weighing up feasibility. University location/ access to family support/ identity and fit were key. Inflexibility and uncertainty of process/ UCAS inflexible/ Risks involved in making the application. Financial cost
Milne et al 2001* Research short report	USA 1 Medical Residency Programme	Quantitative	Interview	Medical Interns	Questionnaire survey	Percentage responses	Interview goals: to learn about programme (84%), sell myself (80%), 86% interview necessary/ 93% no interview was unacceptable.
Mitchson 2009* **Research paper**	UK 1 Post graduate deanery	Quantitative	Selection centre –	Assessors-	Feedback questionnaire	Frequency and percentage responses	Overall 69% felt the new selection centre was an improvement on traditional interview while 4% felt it was worse
Monroe et al 2013* **Research paper**	USA / Canada Large scale survey of 142 medical school	Quantitative	MCAT/ GPA	Admission Deans from all US and Canadian medical schools using MCAT	Online survey	Means Standard Deviations and frequencies of responses	MCAT and GPA viewed most important for shortlisting to interview but less important in the decision of who to admit where interview (4.5 out of 5) and letter of recommendation (3.7) more valued
Niyomdecha et al 2012* Abstract	Thailand 1 medical school	Quantitative	MMI	Medical students and instructor	UTD	UTD	88% of instructors thought MMI process was good and 100% of students thought MMI was fair and pleasant

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
O'Brien et al 2011 Research paper	UK 1 Undergraduate medical School	Quantitative	MMI and standardised interview (SI)	Applicants & Interviewers-	Questionnaire	Likert scale scoring	Interviewers' ranking of SI and MMI re fairness, accuracy and ability to pick best candidate equal. School leavers: MMI more accurate, less difficult than SI (p=0.03 and 0.01). Graduate entrants: MMI more difficult than SI (p=0.005).
O'Flynn et al 2013 (a)* Research paper	Ireland National survey	Quantitative	HPAT –Ireland and academic record.	Guidance counsellors	Questionnaire	Percentage responses	52% in favour of the introduction of HPAT-Ireland, 49% felt system fair. Those opposed concerned re socially disadvantaged. Majority- non-verbal reasoning least relevant.
Patel et al 2011* Abstract	USA Post graduate medical residency programme	Quantitative	Group Interview	77 Applicants	Online anonymous	Percentage responses to survey questions	75 % liked group interviews, 89% felt group interview was effective, IMGs felt they had a significantly harder time impressing interviewers than local candidates (p=0.004)
Patterson et al 2009* Research paper	USA 1 medical school with mission to recruit from and serve underserved populations	Quantitative	MCAT/ GPA	American Indians and Alaskan natives applicants	Questionnaire	Frequency and percentage responses	Access to supports was limited for both accepted and rejected applicants – but latter had less (p<0.05). MCAT very difficult obstacle by 64.7% and financing the application process by 42.4%
Patterson et al 2011 * Research paper	UK Postgraduate training in GP- national survey	Quantitative	SJT / Clinical problem solving test (CPST) and Selection Centre (SC)	Applicants to GP training- 3 cohorts 2007-09.	Online and paper questionnaire survey.	Mean, Mode Standard deviation of responses Cronbach alpha survey 0.7-.94.	Shortlisting: SJT viewed very job relevant but CPST as more so p<0.001. Perceptions of fairness (formal test characteristic and interpersonal treatment) were good. Selection centre- All three tasks positively viewed in terms of job relevance, SP more positively p< 0.001. Perceptions of fairness high
Patterson et al 2013 Abstract	UK Large scale survey	Quantitative	SJT	Candidates to Foundation Year Training Programme	UTD	Candidate reaction	Feedback from candidates indicate SJT relevant and fair

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Randall et al 2006 Research paper	UK 1postgraduate Paediatric Deanery	Quantitative	Selection centre (SC)-	Applicants	Questionnaire survey	Frequency of responses	89% agreed that the content of the SC was appropriate, 89% good opportunity to demonstrate skills, 89% more relevant than other selection tools
Razack et al 2009* Research paper	Canada 1 Undergraduate medical school	Embedded mixed methods	MMI	Applicants and Interviewers	Questionnaires with quantitative and qualitative components	Mean ratings and standard deviations Content Analysis of free comments.	Applicants: MMI more fair than standard interview (p=0.001) & more effective at evaluating non-academic aptitudes (p=0.001) more stressful (p=0.016). Interviewers: Fair, effective, appropriate for use with home and international applicants, however misses some of benefits of traditional interview.
Rich 2011 * **Abstract**	UK 1 medical school- with a widening access agenda	Quantitative	Traditional Interview	Medical student s and Interviewers	Questionnaire	Satisfaction with traditional interview for widening access selection	49% students in early clinical years felt traditional interview should be retained. Only 25% of in clinical years and 20% of interviewers agreed favouring instead a change to multi station interviewing
Rodgerson et al 2013 * **Abstract**	UK Medical / Dental School	Quantitative with free text	MMI	Medical and dental applicants	Online survey post	Percentage responses and mean ratings	94% agreed MMI suitable means of assessing potential/ 45% agreed traditional interview suitable. MMI was more favourable to both groups in terms of enjoyment, stressfulness and fairness.
Samarasekera et al 2014 **Abstract**	Singapore 1 medical school	Quantitative	Focused skills assessment and SJT.	Applicants and assessors	UTD	UTD	92 .4% of candidates happy with format. 82% assessors positive perceptions of the process

*Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim*

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Stagg & Rosenthal 2012* Research paper	Australia 1 Medical School	Qualitative	Not specified	Community Members and Members of the rural based Community Liaison Committee-	Semi structured individual interviews	Thematic analysis	Overwhelmingly saw involvement in selection of students as positive. Opportunity for professional &personal growth; responsibility to represent the broader community; protecting the student and public interest and self-interest in shaping the future workforce...
Stevens et al 2014* Research paper	Ireland 3 medical schools	Mixed methods – embedded	HPAT –Ireland and other selection tools	First Year medical students –	Questionnaire survey	Percentage responses and simple content analysis of free text	Almost all support academic record as suitable tool, 78% interview, 74% personality tests, 68% adjunct admission tests. International students more likely to support interview, knowledge about course, references and personal statements (all p<0.01), Of those who had sat HPAT – 76% felt it fair, 36.7% felt it was easier for males, 32% felt non-verbal reasoning section irrelevant
Tiller et al 2013 Research paper	Australia Graduate entry medical & dental school	Quantitative	Internet based MMI	International applicants and interviewers	Online survey	Percentage responses and mean ratings	Mean satisfaction ratings with use of skype technology 4.25, overall interview process 4.2, being interviewed online as part of overall selection process 4.10, video quality 4.09, audio quality 4.08. 68% would prefer an in-person MMI and 32% a skype interview. 78% of Interviewers satisfied with the iMMI
Turner & Nicholson 2011* Research paper	UK 1 Undergraduate Medical School	Qualitative	UCAS/ Personal Statement/ LOR	Medical school selectors - clinical / non-clinical and lay members	Focus groups and document review	Thematic framework analysis and triangulation with recorded reasons	Most common reason for rejection of candidate for interview was medically related work experience. Teacher reference viewed as influential esp. for rejection but hard to interpret- reading "between the lines" Personal statement – useful but considered highly subjective. Ideal candidate extremely difficult to judge

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Uijtdehaage et al 2011 **Research paper**	USA 1 medical School Specialised to develop leadership & serve disadvantage	Quantitative	MM	2 cohorts of Applicants & 1 cohort of Interviewers	Questionnaire surveys	Mean ratings and standard deviations	Applicants: able to present abilities (5.7), adequate instructions (6.4), sufficient time (3.9), free from gender cultural bias (6.5-6.7) Interviewers – accurate portrayal (5.6), prepared (6), clear instructions (6.2), adequate time (4.7), allow differentiation (5.7), Overall fair (6.2)
UKCAT Consortium 2009/2010 Report	National Survey UK Medical Schools using UKCAT	Quantitative	UKCAT / Aptitude	Applicants 6821 (27% response rate) (of these 33% mature)	Questionnaire	Percentage responses and majority opinions	Considered a difficult test. Unconvinced it tests right attributes. 40% felt their college or school were not well informed about UKCAT/ ¾ had used online practice materials/ books, and found them useful. 90% happy with testing environment.
UKCAT Consortium 2011 Report	National Survey UK Medical Schools using UKCAT	Quantitative	UKCAT / Aptitude	Applicants (19.5% response rate)	Questionnaire	Percentage responses and majority opinions	44% found out it from websites /prospectuses. 33% found out from their schools. 36% of candidates from independent schools rated their advice as good or very good, only 18% from comprehensive schools agreed with this. Majority very supportive of practice tests (93%) and books (90%). Timing in the test is crucial.
Vermeulen et al 2012 Abstract	Post Graduate GP training. Netherlands	Quantitative	Behaviour specific interview, knowledge test, SJT and simulated consultation versus standard panel interviews	47 candidates	UTD	UTD	Both panel interviews and behaviour specific interview were considered job relevant and fair/ but latter offered better opportunity to show competencies. Both SJT and knowledge based test were considered job relevant. SJT considered highly fair (95.7%), simulated consultation (78%) knowledge based test (64%).

111

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Waheed et al 2011* Research paper	Pakistan 1 medical school	Quantitative	Interview	2nd Yr Medical Students & Faculty members	Lecture and survey	Frequency of responses	77% of students highly positively influenced by the lecture compared to 10% faculty, 85% students / 76% Faculty agreed it influenced their impression of Medical school values Faculty more likely to feel important to include such scenarios in admission interviews (p=0.01)
Westwood et al 2007* Research paper	UK 1 Post graduate Deanery Cardiology	Quantitative	Structured Interview	Applicants	Questionnaire	Median and interquartile range 1=strongly agree	Satisfaction rating high, (2), objective (2), appropriate duration (2) offered sufficient scope to express individuality (2) and was relevant to the job (2).
White et al 2011* Research paper	Canada 1 Medical School	Qualitative	Essay	Applicants	Review of essays and interviews with Applicants- Qualitative analysis using modified grounded theory	How applicants approach writing the essay	Applicants expressed the idea that they had approached the essays as a way to ''show themselves'' and ''tell their own story'' in a subjective way which they felt was missing from other parts of the admission process.
Wilkinson & Wilkinson 2013 Research paper	New Zealand 1 Medical School	Quantitative	UMAT	Year 1 University students who sat UMAT twice	Online survey	Percentage response rates, comparisons of scores mean and standard deviation	Commonest forms of preparation were ACER practice materials, MED Entry course, and student led tutorials. Students who took a MED Entry course had significantly higher confidence (mean diff Likert score 0.6), No significate differences for those taking student led tutorials. Moderately strong positive correlation between amount of money spent and confidence r=0.3, p<0.001.

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

112

Table 7 Continued

Author /Year/ Type of article	Location / Setting	Study Design	Selection tool(s)	Stakeholder	Data Collection method	Outcome variable	Overall findings
Wright 2012 **PhD Thesis**	UK 1 Medical School	Mixed Methods thesis, with Qualitative strand used to explore stakeholder views		Medical Students (n=13)	Interviews - Qualitative analysis using Framework analysis	Students' views of influences on decision to apply to medical school and preparedness	Family & School were highly influential on decision, support for application activities such as work experience, preparing personal statements and interview practice. Students from medical /professional backgrounds and fee paying schools were better supported/prepared.
Ziv et al 2008 **Research paper**	Israel 1 Medical School	Quantitative	MOR – Selection centre	Applicants (two cohorts) MOR Assessors	Questionnaire	Frequency of responses	76% of applicants rated MOR as fair 76% felt they had opportunity to express their capabilities 85% (n=299) of rater found it fair. 92% rated MOR assessment items as appropriate

Key: NR - not relevant because qualitative study, UTD –unable to determine as insufficient information provided, * Studies where establishing stakeholder view was principal aim

Chapter 3: The Irish Context

3.1 Selection to Medicine in Ireland

Ireland has six medical schools located in the National University of Ireland Galway, University College Dublin, University College Cork, University of Limerick, Trinity College Dublin and the Royal College of Surgeons in Ireland. With the exception of the University of Limerick which is a graduate entry medical school, medicine is largely a 5-6 year undergraduate degree programme- with applicants coming directly from second level school. This chapter focuses on selection to undergraduate medicine as that is the subject of this thesis.

Ireland mirrors international trends as the demand to pursue a medical career far outweighs the number of places available in medical school. In 2010, for example of the 13,198 EU school leavers who applied to study medicine, 1,885 met the eligibility criteria and (excluding a small number of special access places) 434 were awarded a medical school place (Central Applications Office, 2012a). Based on these figures, the selection ratio in Ireland is 0.23, making admission to this course highly competitive. As indicated in Chapter 1, the selection pathway and admission criteria depend on the country of origin of the applicant. These are discussed in more detail below.

3.2 Selection and Entry Criteria for EU Applicants

Between 1977 and 2009, selection to medicine for EU applicants was based solely on a competitive process, dependent on the score applicants attained in their final second level school exit exam (O'Flynn et al., 2013b). This examination is known as the Leaving Certificate Examination (LCE) and is administered and overseen by the State Examination Commission (2009). To this day selection to medicine is still largely dependent on performance in the LCE or equivalent European school exit exam such as A-levels or Scottish Highers.

In 2009, entry and selection for EU applicants to undergraduate medical school in Ireland underwent a major revision. In essence, selection to medicine in Ireland moved from a system which was based exclusively on a measure of attainment – namely performance in the LCE to one based on a mixture of attainment and ability as measured by a new aptitude test (McManus et al., 2013b). The most substantial change was the introduction of HPAT-Ireland (Australian Council for Educational Research, 2014a). Two other changes

introduced at the same time and have gone largely unnoticed; a moderation of the way in which applicants scores on the LCE were calculated and a stipulation that the LCE score had to be based on results awarded in *one sitting* of the examination- i.e. the examination had to be in the one academic year.

Since the changes were introduced, Irish and EU applicants must satisfy a minimum academic achievement requirement (LCE or equivalent) and sit HPAT-Ireland, with applicants being rank ordered based on a combination of academic and HPAT-Ireland scores, in an approximate 2:1 weighting.   Offers of places are made to the top performing applicants with the numbers of places available to EU candidates capped by government.

Applicants to medicine typically sit Higher Level LCE in English, Irish, Mathematics, another European language, a minimum of one science subject and two further subjects of their own choice. For entry to medicine, applicants' scores on six subjects are taken into account in a system that awards a certain number of *points* for each grade achieved. See Chapter 3 Appendix 1 (a) for details on how points are calculated. For EU applicants outside of Ireland, an equivalent academic record is accepted such as the A level and where is a system of converting to the appropriate points. Applicants are required to score a minimum threshold of points in the LCE to be considered eligible; however, to secure a place in medicine, applicants must score substantially higher.   There is no minimum HPAT-Ireland requirement. Hence, successful applicants can include those who have achieved maximum LCE points and relatively low HPAT-Ireland scores and vice versa.

The introduction in 2009 of *moderation* of the LCE score entails an adjustment being made to applicants' raw LCE score. See Chapter 3- Appendix 1 (b) for details on how LCE raw scores are moderated. In essence, moderation damps down the range of scores at the top end of the LCE points scale so that when moderated LCE points in excess of 550 are adjusted so that every 5 points earned on the raw count, are allocated just 1 point. These moderated scores are known as *Adjusted Leaving Certificate Scores*.

The weighting of scores between LCE and HPAT-Ireland was originally approximately two to one. There has been a minor adjustment to this since its

introduction but essentially the ratio is still close to a 2:1 ratio in favour of academic points. See Chapter 3-Appendix 1(c) for an explanation of the weighting.

The medical schools are entitled to demand their own science subject matriculation requirement in addition to the usual higher level English, Irish, Mathematics and European language. Since 2009, applicants are required to meet both basic matriculation and specific subject requirements in *one* sitting of the LCE.

It is important to note that the administration and processing of selection data, as well as matching and issuing offers of places to all EU applicants is predominately handled by agents external to the medical schools. Hence, by comparison with other countries, the selection of EU applicants requires minimal input from Irish medical schools and faculty.

3.3 Why Change Was Considered Necessary

The impetus to change the way medical students were selected in Ireland came from a number of sources, however, the motivation was not primarily driven by any sense of dissatisfaction with traditional selection processes on the part of the medical profession generally.

When selection was based solely on academic record, the points necessary to secure entry to medicine steadily increased so that by the late 1990s successful medical school applicants were required to secure points in the range of 92-97% of maximum achievable points (Quinn et al., 2010). The relentless increase in points year on year was dubbed "*the points race"* and was the source of much public dissatisfaction*.* The points race was seen as both educationally detrimental and unsustainable. It was argued that it promoted rote learning, gave unfair advantage to those who could afford access to private grind schools and favoured females over males (Byrne, 2000, Clarkson, 2007, Dempsey, 2003, Dunne, 2007). Another negative educational consequence of the points race was that it encouraged students to repeat their LCE- sometimes strategically spacing subjects over two years, so that a situation existed where in 2008 41% of successful

entrants to medicine were repeat Leaving Certificate students (O'Flynn et al., 2012).

At the same time as public discontent with the points race grew a major review of undergraduate education and training in medicine was underway. In 2006 the report from the Working Group on Undergraduate Education and Training in Medicine- Medical Education in Ireland a New Direction was published (Fottrell, 2006). This highly influential report highlighted a number of key changes which should be made to medical education in Ireland. These included expanding the numbers of medical school places available for EU students, which were under strict government control, improving the quality of the educational programmes and curricula, stricter accreditation of clinical training sites and importantly a recommendation that the "*Leaving Certificate results should no longer be the sole selection mechanism for undergraduate students*" (Fottrell, 2006). A second landmark publication was the report from the Medical Education and Training Group which addressed a wide number of issues related to medical workforce planning, recruitment and retention of medical graduates and the quality assurance of post graduate training places (Buttimer, 2006).

Based on the recommendations from both these reports, and in response to the points race, the Departments of Health and Education, through then Minister of Health, Mary Harney and Minister of Education, Mary Hanafin, announced a €200 million package of reforms to undergraduate medical education and postgraduate professional training (Department of Education and Skills, 2006).

As part of the implementation of this package of reforms three key changes were introduced to the selection process. The moderation of LCE points was intended to reduce the emphasis on attaining exceptionally high points while still rewarding excellence in academic achievement. The single sitting rule was intended to reduce the advantage given to repeat LCE applicants. HPAT-Ireland was intended to test skills in domains that are deemed important to the practice of medicine that are not tested or insufficiently tested by the LCE.

3.4 Reaction to Change and Media Coverage

The changes to entry and selection to medicine in Ireland, and in particular HPAT-Ireland, have been an issue of significant public interest as evidenced by the widespread coverage in the national newspapers, television, radio and social media sites as well as parliamentary questions (Houses of the Oireachtas, 2010, Houses of the Oireachtas, 2012, Houses of the Oireachtas, 2014). It could be argued that media coverage both reflected, and was instrumental in shaping public opinion and confidence in the test and continues today. Hence, it is important and relevant to appreciate the extent and nature of media coverage afforded to HPAT-Ireland, particularly with respect to comprehending the public mood at the outset of this thesis.

On February 2nd 2006, the day after the Department of Health launched the investment package in medical education reforms, the paper of record in Ireland, the Irish Times, ran with a headline *"Aptitude Test Likely for Aspiring Medical Students"* (Donnellan, 2006). This was followed over the next number of weeks and months by a series of articles exploring the possible impact of the new reforms (Donnellan, 2006, Kerr, 2006, The Irish Times, 2006). In general, the proposed reforms to medical student selection were initially welcomed. Then Minister for Education, Mary Hanafin was quoted as saying that "*This new entry mechanism will have the effect of recognising a broader range of educational performance and reducing the negative educational impact of intense pressure on students for an exceptional Leaving Certificate performance*" (Flynn, 2007). This view was echoed in other news publications and media sites (Collins, 2009, Department of Education and Skills, 2009, Murray, 2009).

Thus, the public expectation was created that, against the backdrop of an expansion of the number of undergraduate places available, HPAT-Ireland would help lessen the academic scores required for entry to medicine and therefore, would make it more accessible to a larger number of potential applicants. This widely welcomed belief was also in keeping with the National Plan for Equity of Access to Higher Education and fitted with a public mandate to widen access and increase diversity in institutes of higher education nationally (Higher Education Authority, 2008). Shortly after Trinity College Dublin Communication Office officially announced the launch of the new selection criteria for medicine, a

leading Irish Academic went on record stating that one of the main benefits of HPAT-Ireland was that "*the test measured ability rather than prepared learning. This would help avoid a situation whereby scores could be improved through repeated testing and grind courses*" (Flynn, 2008).   The understanding that HPAT-Ireland was not a test that one could study for became widespread in the public domain along with the perception that this was a core aspect of the fairness of the test.

However, a report in the Irish Times by Dr Muiris Houston on the changes to the gender balance in the September 2009 medical school intake sparked a media furore and marked a distinct turning point in public reaction to HPAT-Ireland (Houston, 2009). In his article Dr Houston pointed out that following the first year of operation of the reforms to selection 52% of offers of medical school places went to females and 48% of offers went to males, reversing a trend in the previous years for a 60:40 ratio in favour of female candidates. Interviewing two prominent medical school academics, one of whom stated that changing the gender balance had been one of the considerations in the introduction of HPAT-Ireland in the first instance, the article ran with the headline "*Welcome for more men doing medicine- Medical Schools pleased by return of gender balance*".

Over the next month, the Irish Times published a series of heated letters to the editor, both from within and without the medical profession, on the issues of gender and selection to medicine. Readers of the paper were "*horrified by the notion that the Health Profession Admission Test (HPAT) is considered a good idea because it restores gender balance to the medical profession*" (Cotter, 2009). Some were struck by irony that "…*With regard to our allied health professions namely nursing, pharmacy, physiotherapy, clinical nutrition and occupational therapy I cannot help but note that there has been no effort to redress the imbalance*" (Keily and Byrne, 2009). Others expressed concern until the ratio of females to males in the "*applicant*" pool was made public knowledge, then informed discussion about the gender issue would not be possible (Henry, 2009). Some saw the debate as an opportunity to consider the need for a more family friendly medical career structure (Hussey, 2009).

The letters debate resulted in the publication of a response from the Deans of the Medical Schools clarifying that "there *was no intention at any level that this test, introduced at the behest of the Department of Education and Science, would introduce any form of gender bias and indeed this was one of the exclusion criteria in the tender process*" (Kelleher et al., 2009).

However, this response did little to reassure the public and the media coverage continued, some with more sensational headlines including "*Barred from Medical School – why Grade A rejects feel sick*" which took up the story of one female applicant who failed to secure a place in medicine despite scoring maximum Leaving Certificate Examination scores (Bielenberg, 2009). Interestingly, virtually no media attention was paid to the important effect of moderating the Leaving Certificate Examination scores, a change that was introduced at the same time as HPAT-Ireland and in subsequent reports proved to be as influential as HPAT-Ireland in altering gender patterns in the intake (O'Flynn et al., 2013b, O'Flynn et al., 2013c).   In particular, moderating the Leaving Certificate scores was likely to affect female applicants more, as they were more likely to achieve the higher Leaving Certificate points.

Over the next number of years, media coverage of HPAT-Ireland waxed and waned in response to the timing of relevant published academic articles and impending significant dates in the academic calendar such as the closing date for applications to university (Houston, 2010, Siggins, 2010, The Sunday Times, 2013). Many publications including the Irish Times, the Irish Medical Times and Medical News ran supplements for candidates on how to prepare for taking HPAT-Ireland. Twitter threads such as #hpatresults predominately echoed candidates' anticipation of and reactions to HPAT-Ireland results while others advertised commercial coaching and some medical schools and journalists cited links to relevant HPAT-Ireland publications and reports (Brant Toni [@ToniBrant], 2014, Murchan Emilie [@EmilieMurchan], 2012, O'Sullivan Ryan[@Jandyman95], 2014). Boards.ie (http://www.boards.ie/), on of Ireland's largest online discussion forums hosted a number of different discussions tagged with HPAT. In excess of 5,000 entries range from preparation for HPAT-Ireland, to repeating it and reactions to candidate results.

Following the publication of the interim report on HPAT-Ireland in late 2012, media coverage again intensified (O'Flynn et al., 2012). HPAT-Ireland was criticised for having failed to deliver on expectations on a number of fronts (Mooney, 2012, RTÉ News, 2012b). Media reports suggested that commercial coaching did in fact lead to improved performance. This flew in the face of widening access, as only those who could afford commercial coaching stood to benefit, and as time progressed HPAT-Ireland was widely criticised for failing to "*level the playing field*"(Mooney, 2012). There were repeated calls for the abolition of HPAT-Ireland from Minister of Health Dr James O Reilly (O'Regan, 2012). However, it was recommended by the interim report group that HPAT-Ireland would stay on as part of the entry requirement to medicine for a further four years to allow for a fuller analysis based on evidence.

In summary, from the very outset HPAT-Ireland has and continues to receive significant and widespread media attention. This is not surprising given its novelty, the large number of applicants to medicine each year and the extent to which securing a place in medicine requires commitment and perseverance on behalf of applicants and their families. Initially welcomed as a method of broadening access to medicine and reducing the pressure on school leavers to attain a perfect LCE, HPAT-Ireland has also been widely criticised for perceived unfairness on the grounds of gender and susceptibility to coaching.

3.5 Selection and Entry Criteria for Non-EU Applicants

According to the Organisation for Economic Co-operation and Development (OECD) international students are one of the fastest growing parts of the global education industry (Ranchin and Rebolledo-Gómez, 2013). International students refer to those who complete higher education outside of their country of citizenship (Hallock et al., 2007). In an Irish medical education context, an international student generally refers to those whose country of citizenship is outside of the EU.

There are a number of entry routes to undergraduate medicine in Ireland for Non-EU applicants (School of Medicine, 2012) (School of Medicine, NUI Galway 2012). These routes generally operate through agents. Examples include Atlantic Bridge (http://www.atlanticbridge.com/) for applicants from North America and

the Irish University Medical Schools Consortium (IUMC) (iumc@po.jaring.my.) for applicants from Malaysia.

Individual medical schools are allowed to set their own requirements for the selection of Non-EU students. Typically, the selection process includes academic record, evidence of English language proficiency, reference plus interview. Non-EU applicants to graduate entry programmes are required to sit either the GAMSAT or MCAT, depending on the medical school involved. Non-EU applicants are not required to sit HPAT-Ireland.

By contrast with EU applicants, the selection of Non-EU applicants involves significant input from Medical Schools. Typically this entails engagement with the relevant agent at both the shortlisting and selection stage but the degree of input can vary between the individual medical schools, and the specific agent involved. In the case of NUI Galway, senior members of the Medical Faculty contribute to the screening of academic records, biographical essays and letters of reference. Additionally, members of Faculty conduct structured panel interviews for South East Asian applicants biannually in Malaysia and Singapore. Table 8 summarises the main differences between EU and Non-EU selection criteria.

Table 8 Criteria for Medical Student Selection in Ireland

EU Applicants Selection Criteria	Non-EU applicants Selection Criteria
Academic Record: Leaving Certificate Examination or equivalent	Academic Record : Grade Point Average
English Language Proficiency- if required	English Language Proficiency: International English Language Testing System (IELTS) or equivalent
Health Professions Admission test –Ireland (HPAT-Ireland)	+/- Traditional Interview
	+/- Others including MCAT, Personal statement, reference

Chapter 3: The Irish Context

All the medical schools require Non-EU students who do not have English as a first language, to provide evidence of English proficiency. The most commonly used test of is the International English Language Testing System (IELTS) (International English Language Testing System 2014). First used in the 1980's, it is taken by over 1 million candidates per year and recognised in 120 countries (Wilson, 2010). It comprises four modules; Listening, Reading, Writing and Speaking. Results are represented as a band score from a nine-band scale, with each band corresponding to a specified competence in English. A score of 1 corresponds to a few isolated words and 9 represents an expert user of the language. The total overall score as well as Listening and Reading scores are reported in whole and half bands. The Writing and Speaking band scores are reported in whole bands only. Guidelines recommend an IELTS score of 7 as probably acceptable for "linguistically demanding" academic courses with an IELTS score of 6 recognised as the minimal acceptable requirements for a university course (Feast, 2002). At present in Ireland, the minimum acceptable overall IELTS score is 6.5 with not less than 6 in each of the 4 domains.

Data on the reliability of IELTS are publicised by the test developers. The most recent reports indicate that in 2013 the internal reliability of the Listening and Reading tests ranged from a Cronbach's alpha of 0.88 to 0.93 across the various versions of the test used. Using generalisability studies, the reliability of the Speaking and Writing sections ranged from 0.81–0.89 (International English Language Testing System 2015). On the whole, there is conflicting evidence regarding the predictive validity of IELTS. Study outcomes have ranged from no significant findings to strong correlations. However, the majority of studies report significant, small to medium correlations between IELTS and future academic performance (Feast, 2002, Hyatt, 2012). A study of the acceptability of IELTS to higher education academic and administrative staff, indicated that the majority felt that it was a very useful indicator of English proficiency required for the demands of the academic course (Hyatt, 2012). They concluded that it was a "fit for purpose test". However, qualitative interview data highlighted concerns regarding the tension between setting acceptable cut-off standards and the pressure to recruit and identified a need for post entry language support for successful applicants (Hyatt, 2012).

3.6 Funding of Irish Medical Schools

The issue of funding of Irish Medical Schools and the selection of Non-EU students are closely intertwined. Undergraduate Irish Medical Schools are funded by two principal sources: state funding and fees from Non-EU students. Under the free fees initiative tuition fees for EU students studying medicine are paid by the state (Citizens Information Board, 2015). For over 30 years, the number of places available for EU medical students is tightly regulated by government and state funding to individual medical schools is based on the allowed intake of EU students per school.

By contrast, medical schools have relative autonomy to decide how many Non-EU students they recruit annually, subject to periodic review by the Medical Council. State funding for the EU intake approximates to €12,000 per student per year and is inadequate to meet the costs of modern undergraduate medical education and training.

As a consequence of the cap on EU places and the shortfall in state funding for the costs of training these EU students, medical schools have increasingly relied on the selection of Non-EU students for financial viability. Non-EU students pay fees of circa €30,000 per academic year, and as such are a vital revenue stream for Irish Medical Schools. Fottrell (2006) calculated that Non-EU students contribute over 50% of the funding for medical education nationally. Non-EU students account for a very significant proportion of medical school places in Ireland, in some cohorts accounting for up to 60% of the medical school intake (Fottrell, 2006).

3.7 Summary of Irish Context

In summary, selection and entry to Irish undergraduate medicine has come under scrutiny since changes to the system were introduced following the publication of both the Fotrell and Buttimer reports (Fotrell, 2006, Buttimer, 2006).

In 2009, three changes were made to the selection criteria for EU applicants: the introduction of HPAT-Ireland, moderation of the LCE scores and a stipulation that LCE subject and matriculation requirements have to be met in one sitting of

the LCE. Selection of EU applicants is via a central application system, which is tightly regulated and capped by government. As selection of EU medical students is largely managed centrally by the CAO and ACER, the resource implications for medical schools in administering this entry route is minimal. Media attention and debate has focused on HPAT-Ireland. While initially welcomed the public response to HPAT-Ireland has been critical.

Non-EU applicants are selected by a different process which does not include HPAT-Ireland, but involves other tools including IELTS and in most cases, an interview. Non-EU students are fee paying and hence provide a vital source of revenue to medical schools. The number of places available for Non-EU students is not subject to the same tight control as their EU counterparts. Individual medical schools have more direct input in to the administration of this selection pathway. The public debate regarding selection to medicine has not included the Non-EU pathway.

Chapter 4: Methodological Considerations

4.1 Research Methodology

According to Patton "*Purpose is the controlling force in research*" and from purpose stems all decisions regarding the choice of methodology, design and methods (Patton, 2002). The purpose of this research thesis is to establish the predictive validity and stakeholder acceptability of selection tools for medicine. To this end the chosen research methodology is mixed methods.

Greene identifies that the overall purpose for mixing methods in a programme of research is "*to develop a better understanding of the phenomena being studied*" (Greene, 2007). The development of mixed methods as a distinct research methodology began in earnest in the 1980s. Since then, researchers have sought to explore, define, challenge and defend the merits of mixed methods as a distinct research methodology (Creswell and Clark, 2011). Mixed methods research is defined as "*research in which the investigator collects and analyses data, integrates the findings, and draws inferences using both qualitative and quantitative approaches, or methods, in a single study or a program of inquiry*" (Tashakkori and Teddlie, 2010). Implicit in this definition is the fact that mixed methods allow the researcher to attempt to answer research aims and objectives where one source of data, be that qualitative or quantitative, may be insufficient to adequately address them. The objective of mixed methods research is to "*draw from the strengths and minimise the weaknesses*" of both quantitative and qualitative research methods (Johnson and Onwuegbuzie, 2004).

Mixed methods have been identified as an appropriate for studying *complex* medical education topics (Schifferdecker and Reed, 2009). In a critical narrative review, Maudsley (2011) concludes that they are "*increasingly relevant to medical education".* Likewise a recent AMEE guide concludes that there is growing evidence that a "*combination of quantitative and qualitative methods are important*" in medical education research (Tavakol and Sandars, 2014). Medical student selection can be fittingly described as a *complex* medical education topic. In terms of metrics of good assessment, public accountability and transparency, there are defined and objectively measurable outcomes such as feasibility, reliability, predictive validity and economic cost of selection tools. However, in terms of utility and ensuring generality of use it is also necessary to consider the multiple perspectives of stakeholders and various stakeholder groups, their

interpretations of and reactions to selection tools, their level of engagement with changes to selection processes and to identify any unintended negative consequence that may be overlooked by the more objective measures . Understanding how these different facets fit together is indeed complex.

Greene identifies five main purposes for mixing methods in social inquiry (Greene, 2007). The main purpose of using mixed methods in this thesis was *complementarity* which occurs when researchers strive for broader and more comprehensive understanding by utilising methods that purposefully set out to investigate *different elements of the same complex phenomena* (Greene, 2007). The results from the different methods are then brought together to enhance, deepen and extend the overall interpretations and inferences from the research. For example, quantitative methods were employed to establish the predictive validity of HPAT-Ireland and MMI with respect to medical student assessment. Whereas qualitative methods were necessary, to meaningfully explore both doctors' and students' views of these selection tools.   The rich and diverse insights gleaned from utilising this mixed methods approach provided valuable information that neither method could have done in isolation.

A second purpose for using mixed methods was *development* which involves sequentially using data from one method to inform the development of the other method with the intention of *improving understanding by capitalising on each method's inherent strengths* (Greene, 2007). The author goes on to state that this definition of development is to be broadly construed to include aspects of implementation, as well as sampling and instrument construction. For example, results from the quantitative feedback survey of MMI candidates and assessors used in the MMI Feasibility Study (Study 3) were used to inform the topic guide for the focus group interviews used in the qualitative strand of Study 4.

Although using mixed methods for the purposes of *initiation* and *triangulation* were not amongst the original motivations for choosing a mixed methods approach for this thesis, there were some instances where these benefits were also realised by adopting this methodology. The purpose of initiation is to use different methods to examine the same complex phenomenon *with the intention of provoking paradox, contradiction or divergence* (Greene, 2007). Using mixed

methods to establish the predictive validity of HPAT-Ireland and to explore doctors' views of the ability of this selection tool to predict future performance, highlighted the divergence between what doctors' believed to be true of HPAT-Ireland's capacity to predict performance and what was established by measurement. The purpose of triangulation differs from initiation in that it *seeks convergence, corroboration and correspondence of results from different methods* (Greene, 2007). Using quantitative methods in the MMI studies established that Non-EU students' mean MMI score was significantly lower than their EU counterparts. This tallied with stakeholder perceptions that language and culture were significant barriers to Non-EU student performance on MMI.

The final purpose for mixing methods is *expansion-* whereby the researcher chooses mixed methods *to expand beyond the scope and range of the study to assess different phenomena* (Greene, 2007). This was not a consideration for adopting mixed methods in this thesis.

Mixed methods research is not without its difficulties. Its greatest challenges are that it demands of the researcher that they become versed in the skills of both the qualitative and quantitative paradigms; it requires more in-depth planning and preparation of the various phases of the research, data analysis is more complex and the integration and synthesis of conclusions more demanding. Additionally there is the challenge of reducing the inevitably large amount of generated data to focused publishable papers with concise messages. All of these challenges were encountered and ultimately overcome throughout this thesis.

4.2 Paradigm/ World View

One of the main philosophical questions faced by mixed methods researchers is how to accommodate the divergent worldviews of the quantitative and qualitative traditions, within a mixed methods programme of research. An important aspect of mixed methods is that it is practical and based in the real world. It does not demand that the researcher rigidly adhere to one world view or paradigm but rather asks the researcher to consider which world view is most appropriate to understanding and answering the research aims and objectives. Creswell and Clarke (2011) identify four major worldviews used in research namely post positivist, constructivist, participatory and pragmatist.

Pragmatism focuses on the consequences of actions, is problem centred, pluralist and real-world orientated (Creswell, 2007). It is very commonly adopted in mixed methods studies (Tashakkori and Teddlie, 2010). Pragmatism was chosen as the best fit world view for this thesis for four main reasons:

 1) In terms of ontology pragmatism recognises both singular realities (such as measureable values of predictive validity, economic cost of selection tools) and multiple realities (such as perceptions, interpretations and stakeholder acceptability). Having this openness to the nature of reality was fundamental to addressing the research objectives of this thesis. It was also helpful to start out from this pluralist standpoint when conducting four separate studies throughout the course of this thesis, involving fourteen collaborators each with their own opinions on, and experiences of, the nature of reality.

2) With respect to epistemology pragmatism is driven by a "what works" approach to acquiring knowledge, which is essential to conduct research in the complex field of medical student selection, and encompasses quantitative and qualitative methods.

3) Regarding methodology pragmatism accommodates both deductive and inductive approaches to the process of research; both of these approaches were required to adequately address the objectives of this thesis.

4) From a personal perspective, pragmatism aligns well with my philosophical outlook and complements my background of general practice specialist training, which emphasises accepting multiple ways of seeing the world and finding workable solutions to real world problems.

4.3 Research Design

Mixed methods research designs are varyingly classified in the literature and there is no one agreed inclusive design typology. Tashakkori and Teddlie (2010) summarise the different approaches to classifying mixed methods research designs into six categories. The basic design classifications (Types I-III) differentiate mixed methods designs along relatively simple lines relating to the number of research strands, the manner in which the qualitative and quantitative methods are incorporated and the stage of the research process at which the

mixing occurs. While the more complex design typologies classify mixed methods designs according to a combination of the basic differences above and a more integrated consideration of the research purpose, worldview and decisions about quality related to both the quantitative and qualitative aspects of the study (Type IV); the degree of iteration between phases of the study (Type V) and the amount of synergism resulting from mixing of quantitative and qualitative components (Type VI).

Creswell and Clarke's (2011) classification of mixed methods designs is an example of a complex integrated design typology (Type IV). They differentiate six major mixed methods research designs: convergent parallel, explanatory sequential, exploratory sequential, embedded, transformative and multiphase. According to this classification, a multiphase design was selected for this thesis. A multiphase study "*combines the concurrent and /or sequential collection of quantitative and qualitative data sets over multiple phases of a programme of study*" with the purpose of addressing a "*set of incremental research questions that all advance one programmatic research objective*" (Creswell and Clark, 2011).

The main features of this research design that were pertinent to the decision to adopt it for this thesis include:

- The researchers or team of researchers examine a subject through a series of connected quantitative or qualitative studies
- It encompasses both sequential and concurrent strands over a period of time
- It is compatible with a pragmatic world view
- Equal emphasis is given to both quantitative and qualitative strands
- Mixing occurs within the programme objectives
- It allows for each study to address its own specific research questions that "*evolve to address a larger programme objective*"
- It is a sufficiently flexible design to incorporate emerging questions as they arise throughout the course of the research
- It facilitates researchers to publish papers from the individual studies – while still contributing to an overall research programme.

Figure 2, Chapter 1 depicts the multiphase research design diagram for this thesis, highlighting the different phases of data generation, collection and interpretation, and paper publications.

4.4 Theoretical Perspective

4.4.1 Background

It is recommended that researchers consider a theoretical perspective on their work (Creswell and Clark, 2011, Crotty 1998). The use of theory can sensitise the researcher to issues and interpretations that they may not have considered themselves (Mac Farlane and O Reilly-de Brún, 2012). In the case of a multiphase research design, a theoretical perspective assists the researcher to integrate data, draw relevant interpretations and synthesise meaningful conclusions (Creswell and Clark, 2011).

Organisational justice theories are particularly relevant to medical student selection. These justice theories were outlined in Chapter 1 and expanded on below. As described elsewhere, the data from the qualitative strands of this thesis were initially analysed inductively, categorised and reported as themes independent of these theories (Mac Farlane and O Reilly-de Brún, 2012). Subsequently the findings from the thesis overall were mapped against the constructs of organisational justice theories (See Chapter 9).

Although, to date, organisational justice theories have been predominately used to explore first hand reactions, there are a number of other examples where these theories have been used to understand third party views (De Cremer et al., 2005, Van den Bos and Lind, 2001). Furthermore, Gilliland a leading authority on organisational justice recommends that researchers should consider the impact of justice in selection processes on a broad group of stakeholders (Gilliland and Hale, 2005). Hence, their application to this thesis is justifiable.

In the original article extending organisational justice theories to selection, Gilliland (1993) emphasised the "*social side of selection*". He argued that "*perceptions of test fairness*" were equally as important as psychometric

measures, such as validity and reliability in considering the utility, ethicality and legality of selection tools.

Although some scholars differ, the generally accepted model of organisational justice theories utilises a tripartite typology: procedural justice describes the fairness of the decision making procedure, distributive justice describes the fairness of the allocation of outcomes and interactional justice the fairness of the interpersonal treatment during the implementation of procedures (Leung, 2005).

4.4.2 Procedural Justice

Procedural justice can be thought of in terms of the stakeholders' view of the satisfaction or violation of a number of rules. It is one of the most influential determinants of perceived fairness of selection tools. Originally Gilliland described ten procedural justice rules (1993). These have been slightly modified in the intervening years, as interactional justice became recognised as a domain in its own right. However, it is still useful to consider the rules relating to formal characteristics of the selection tool. These include perceptions of the job relatedness of the selection tool, opportunity to perform, reconsideration of opportunity and consistency of administration.

The dimension that exerts the greatest procedural influence on perceptions of fairness is the extent to which a selection tool is viewed as job related (Gilliland, 1993). Selection tools that approximate actual aspects of the job, such as work samples or assessment centres with relatively high fidelity selection processes, are generally considered to be more job related and to be found fairer by applicants (Gilliand and Hale, 2005). A study of selection to General Practice training in the UK, found that applicants rated a clinical problem-solving test and a simulated patient consultation, as highly job related and fair at the short-listing and selection stage respectively (Patterson et al., 2011). Perceptions of the degree to which the content of the selection tool is considered to be relevant to the job have been shown to be distinct from ideas regarding its ability to predict future performance (Bauer et al., 2001).

Opportunity to perform refers to the fact that applicants views selection tools as fairer when they believe they have had adequate chance to demonstrate their skills and abilities (Gilliland and Hale, 2005). It relates to the concept of "voice"

which means that "*procedures are perceived to be more fair if recipients of the decision outcome have the opportunity to express themselves prior to the decision*" (Colquitt, 2001, Gilliland, 1993). Voice helps to explain why candidates generally prefer interviews over aptitude tests, as they offer an opportunity to demonstrate one's abilities directly to the selectors.

Reconsideration of opportunity means that the stakeholder sees there were ample opportunities to have their scores re-checked and that they were satisfied with the process for reviewing their test results (Bauer et al., 2001). This has been shown to be important in the case of applicant views of multiple choice tests (Truxillo et al., 2001).

Lastly, consistency of administration relates to perceptions of the degree to which the test was administered equally to all applicants (Bauer et al., 2001). In particular, giving the same selection procedure to all applicants has been shown to impact on perceptions of fairness (Gilliland and Hale, 2005). Linked with this is the degree to which the test is in common usage, widespread use impacts positively on perceptions of fairness (Gilliland and Hale, 2005).

4.4.3 Distributive Justice

Distributive justice is governed by the degree to which an organisation is seen to divide resources and rewards fairly, according to a particular allocation rule (Ambrose and Arnaud, 2005). One of the commonest allocation rules is equity. This means that rewards are distributed in accordance with contribution and people will judge a process to be fair by comparing their own input and outcome ratio to that of some other referent person (Gilliland, 1993). In general, selection is considered fairer when the rule of equity is applied (Gilliland, 1993). Perceptions of distributive equity may be influenced both by expectations of receiving a place and offers of places (Gilliland and Hale, 2005). When expectations of receiving an offer are high, a disappointed candidate may view the process more unfair than a situation where the expectation of receiving an offer was low (Gilliland, 1994).

Two further allocation rules are commonly recognised. Equality, meaning everyone gets the same irrespective of input, and need which means that more is given to the recipient with the greatest need (Ambrose and Arnaud, 2005). One

way to apply the equality rule to selection is to ensure an un-biased test. The expectation is that everyone has the same opportunity of being selected (Gilliland and Hale, 2005). An example of how the need rule applies to selection is the practice of banding. Some selection processes operate by offering places to all applicants within a certain band of scores, the band being determined by the estimated measurement error. This system can be used in an effort to increase hiring rates of minority applicants (Gilliland and Hale, 2005).

Cultural variations can impact on people's perceptions of the particular allocation rule likely to be acceptable (Leung, 2005). For example, in countries with individualistic cultures the equity rule is generally preferred, whereas in collectivistic cultures such as Japan, the equality rule is more acceptable.

4.4.4 Interactional Justice

Interactional justice focuses on the interpersonal side of organisational structure (Cohen-Charash and Spector, 2001). It has two components; informational and interpersonal (Cropanzana et al., 2007). Informational is concerned with providing knowledge about the procedures. Rules regarding informational justice include that the information is timely, accurate and includes explanations for decisions (Gilliland and Hale, 2005). Explanation includes feedback to stakeholders, honesty and provision of selection information.

Interpersonal justice describes the respectful, dignified and considerate treatment of individuals (Cohen-Charash and Spector,2001, Gilliland and Hale, 2005). It has been found that the experience of injustice in dealing with an organisation (such as not getting the job, or not being selected to medical school) can be somewhat mitigated by positive experiences of interactional justice (Cropanzana et al., 2007). For example, Kanerva et al (2010) found when they explored nurses' views of selection interviews in which they were unsuccessful, one of the most influential factors on overall impression of fairness was interactional justice. For the nurses in this study, respect and opportunity for two way exchange were considered markers of just interaction.

Co-workers value the interactional justice of being kept informed of decisions regarding selection processes in which they feel they have a stake (Gilliland and Hale, 2005).

In summary, organisational justice theories can be categorised as procedural, distributive and interactional. Stakeholder's reactions to selection tools can be understood by the extent to which they view the tool as complying with a set of rules associated with each category. Procedural justice in particular is highly influential on views of selection. The extent to which the selection tool is considered job related is a key factor in perceptions of procedural justice.

4.5 Ethical Approval

Ethical approval for the thesis was granted by the NUI Galway Research Ethics Committee. As the thesis was constructed over phases, the applications to ethics mirrored these.

The main ethical consideration in Study 1 was that the possibility existed that students could be concerned that their exam results were being reviewed by research personnel who were independent of their actual exam process. This was overcome through the use of a linked anonymised database, with only the data enterer having access to the students' identity (Chapter 4 Appendix 1).

In Study 2, there were two principal concerns. Firstly, there was a possibility that the doctors may have experienced exam stress while sitting HPAT-Ireland. This was minimised primarily by prior warning to participants that they may experience some exam stress and that the invigilator would be sensitive and the option to withdraw from the study at any time was reiterated. No issues arose with respect to this concern during the study. Secondly, the Research Ethics Committee queried how the anonymity of study participants would be protected if HPAT-Ireland results were linked to their interviews and included in the published paper. This caused us to carefully consider the pros and cons of reporting practising doctor's performance on HPAT-Ireland. We decided that reporting how qualified doctors performed on the test was of limited value at best because they were not the target audience of the test so comparisons with school leavers were largely unfounded. Hence, we were able to reassure participants that their scores on HPAT-Ireland would be returned to them personally for their private information and would not form part of any analysis or publication. This was particularly important because, as the study was qualitative small numbers of participants were involved, it was based in one institution, hence it would be

potentially possible for readers to identify study participants and their individual scores (Chapter 4 Appendix 2).

Studies 3 and 4 were covered by the same ethical application. The study was granted provisional approval first with instructions to ensure that students' MMI scores would not be made available to others within the medical school and that they would in no way influence student's progress. Students were to be given ample advance information to adequately prepare and the use of senior medical students to act as role players for MMI stations was discouraged. Having undertaken to comply with these conditions full ethical approval was awarded (Chapter 4 Appendix 3). As the studies were underway, a number of unanticipated issues gave rise to the need to reapply for ethical extensions. Firstly, through expanded reading about selection, came the realisation that the socioeconomic background of candidates was an important consideration in determining the fairness of a selection tool, especially with respect to widening diversity. Therefore further ethical approval was sought to allow asking students for their parents' occupation. This was granted and subsequently students were categorised according to the Central Statistics Office socioeconomic grouping. Secondly, it became apparent very early in the recruitment phase for the MMI, that the comparison between EU and Non-EU students would be an important research component of this study. Therefore, further ethics approval was sought to allow the conduct of EU and Non-EU student focus groups. This extension was approved.

4.6 Thesis by publication

This is an article based thesis. Selection to medicine is a controversial and quickly evolving field of research. Thesis by publication enabled me to add to the body of published evidence regarding the use of aptitude tests and multiple mini interview, in a timely fashion. Thesis by publication is congruent with the principles of the Multiphase Mixed Methods design adopted for this research (Creswell and Clark, 2011).   This format ensured that, as well as becoming skilled in carrying out research, I also developed and honed the skills required to successfully publish in the peer reviewed medical education press, which presents its own challenges.   One of the most important aspects of my publication strategy was to secure international peer reviewed publications.

4.7 Contribution to Research

This thesis comprises four research studies with their respective peer reviewed published papers. Collectively there were fourteen collaborators involved in this programme of research. I was the principal investigator for all four studies and took responsibility for the original research concept, establishing and maintaining collaborations with the researchers, drafting the research protocols for each study and editing with input from relevant collaborators, securing ethical approval and funding for the studies, recruitment of participants to each study, quantitative data collection and data base set up, data entry and maintenance(except for Study 1 – where a collaborator (DR) inputted the student selection and assessment data and Study 3 where a collaborator (AJ) inputted the stakeholder feedback), conducted the qualitative data interviews for Study 2 and led the analysis of the transcripts, planned the administered and oversaw the running of the MMI, co-analysed the MMI focus groups interviews, carried out the economic costing for Study 3, statistically analysed the quantitative data for Study 4 (with a collaborator (JN) providing statistical guidance) and co-analysed the quantitative data for Study 1 (with a collaborator (DR) leading this statistical analysis, and another collaborator (JN) providing statistical guidance). I wrote the first draft of the four papers and ensured that co-authors were kept involved in the writing-up process. I was corresponding author for all papers and took responsibility for involving the co-authors in responding to reviewers' comments. I oversaw the day to day administration of each study and was responsible for dissemination of study findings.

In addition I was responsible for securing funding, through competitive processes for each of the studies and maintaining the research accounts- (See Appendix A for summary of research funding awarded).

4.8 Reflexive Account

Reflexivity is a core consideration of qualitative research (Pope and Mays, 2006). It describes the awareness a researcher has of the "*biases, values and experiences that he or she brings to a qualitative research study*" (Creswell, 2013) (page 216). It is widely recommended that qualitative researchers make their position explicit (Corbin and Strauss, 2008, Patton, 2002). In particular the

research should describe his or her "*experiences of the phenomenon being explored*" and "*how these past experiences shape the researcher's interpretation of the phenomenon*" under study (Creswell, 2013 page 216).   Creswell (2013) goes on to suggest that this reflexive account can be included either at the outset, the end or throughout the research study- with the researcher free to choose the most suitable placement. Reflexivity is an important aspect of the quality and rigour of the qualitative strands of this mixed methods thesis. It contributed substantially to the choice of research objectives, the design of the study, the generation and subsequent interpretations of the qualitative data and hence the most logical place to include this reflexive account is here at the outset.

Past Experience

 I came to this thesis with no background experience of selection to medicine research or the literature in this field. My experience in selection was limited to my own selection to medicine in Ireland in the 1980s. At that time selection was based solely on academic record. I secured a place to study Medicine in the then University College Galway (now NUI Galway).   My interest in the topic was sparked when my own nephew decided to apply for medicine and was amongst the first applicants to sit HPAT-Ireland. Like most Medical Faculty in Ireland, even though I was working for many years in a Medical School, my only knowledge of selection processes to medicine consisted of anecdotal stories of HPAT-Ireland. This is because, prior to the changes, selection to medicine of EU applicants was administered centrally by the Central Applications Office (CAO), with virtually no input by Medical School Faculty. With respect to the selection of Non-EU students, I was aware that they were selected via a separate pathway but sketchy on the detail of this.

Biases

    At the outset of this thesis, I had no real strong feelings with respect to HPAT-Ireland, and came to the thesis with a relatively open mind.   I was however, aware of the fact that some within the medical profession and indeed the public were sceptical of the ability of any selection tool to predict future practice. I was also aware that there was controversy regarding the possible role of gender and socioeconomic group on candidates' performance. In general my stance in life is to seek out my own evidence before committing to one viewpoint or another, so

even though I was aware of these controversies, I would not necessarily have been swayed by them. However, I was acutely aware that the changes to selection were brought in with little consultation and without any pilot testing. In my opinion this is contrary to the principle of stakeholder involvement and demonstrates a lack of respect. Hence, it is possible that on a subconscious level I would have been predisposed to having a negative impression of the changed selection processes, on these grounds alone. It is also possible that as a woman in medicine I would have been subconsciously negatively influenced by the media coverage of HPAT-Ireland and gender.

Values

Throughout my training as an undergraduate medical student, non-consultant hospital doctor, General Practitioner and medical educationalist I have come to value the following: inclusivity, partnership, consultation, fairness, transparency, feasibility, respect, reliability and evidence. These values were very influential in deciding on the research aims of this thesis, and shaped the dual foci of predictive validity and stakeholder acceptability. Coming from a non-medical family background and being the first generation to go to university, I particularly value widening diversity to medicine and respect moves to ensure that our medical workforce reflects the diversity of the population that we serve.

Possible Impact of the Above on my Interpretation of the Data

Throughout this study, I have been conscious of how my own experiences, biases and values could impact on data generation, interpretation and representation of findings. To this end, I have taken a number of steps to 1) heighten my own awareness of this – through keeping a reflective dairy at various stages throughout the research – (See Chapter 4 Appendix Excerpts from Reflective Research Diary) 2) actively pursue research collaborations with others both *with* established experience in medical student selection internationally balanced by other researchers outside of the field of medicine and *with no* special interest in selection but expertise in other research dimensions and to *involve* these collaborators in the generation and interpretation of data; c) reduce the possibility that my own stance could interfere with study participants' freely expressing their views- for example in Study 4 Strand B, I employed an experienced focus group moderator to conduct the MMI focus groups in case

students or MMI assessors would feel compromised in being honest with their views of MMI in front of me, as they were aware that I was heavily invested in the running of the MMI; d) member check with the study participants to help ensure that my interpretation of their transcript did in fact reflect what they understood that they said (See Chapter 4 Appendix 7); e) become very informed about medical student selection through reading the peer reviewed literature, attending international conferences and engaging in debate and discussion on selection with other in the field. This has been invaluable in helping to educate, shape, challenge and inform my underlying ideas regarding selection. Throughout the course of this thesis, as I became more informed about the complexities of selection, including important aspects of social justice such as equity of access I have become more open to the merits of adjunct selection tools and the importance of considering academic and non-academic attributes in the selection of medical students. Hence, throughout this thesis my attitude towards selection would have become more positive about these aspects.

4.9 Summary

Figure 4 is based on Creswell and Clarke's (2011) adaptation of Crotty's research levels and summarises the overall framework for this thesis in terms of worldview, theoretical background, research methodology, study design and methods of data collection (Crotty,1998).

Worldview: Pragmatism

Theoretical Perspective: Organisational Justice Theories

Research Methodology: Mixed Methods

Research Design: Multiphase

Methods: Semi- structured one-to-one interviews, focus groups, questionnaires, hand-searching of admission data, review of selection records, collation of undergraduate examination results, collation of scores on selection tools, piloting Multiple Mini Interview, economic costing.

Figure 4 Research Framework (Adapted from Crotty 1998/Creswell and Clarke 2011)

Chapter 5 Predictive Validity of HPAT-Ireland (Study 1)

Maureen E Kelly, Daniel Regan, Fidelma Dunne, Patrick Henn, John Newell, Siun O'Flynn. To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills? – An observational cohort study. BMC Medical Education 2013, 13, 68 (10th May) Reproduced with permission - Chapter 5 Appendix 1)

5.1 Abstract

Background

Internationally, tests of general mental ability are used in the selection of medical students. Examples include the Medical College Admission Test, Undergraduate Medicine and Health Sciences Admission Test and the UK Clinical Aptitude Test. The most widely used measure of their efficacy is predictive validity.

A new tool, the Health Professions Admission Test- Ireland (HPAT-Ireland), was introduced in 2009. Traditionally, selection to Irish undergraduate medical schools relied on academic achievement. Since 2009, Irish and EU applicants are selected on a combination of their secondary school academic record (measured predominately by the Leaving Certificate Examination) and HPAT-Ireland score. This is the first study to report on the predictive validity of the HPAT-Ireland for early undergraduate assessments of communication and clinical skills.

Method

Students enrolled at two Irish medical schools in 2009 were followed up for two years. Data collected were gender, HPAT-Ireland total and subsection scores; Leaving Certificate Examination plus HPAT-Ireland combined score, Year 1 Objective Structured Clinical Examination (OSCE) scores (Total score, communication and clinical subtest scores), Year 1 Multiple Choice Questions and Year 2 OSCE and subset scores. We report descriptive statistics, Pearson correlation coefficients and Multiple linear regression models.

Results

Data were available for 312 students. In Year 1 none of the selection criteria were significantly related to student OSCE performance. The Leaving Certificate Examination and Leaving Certificate plus HPAT-Ireland combined scores correlated with MCQ marks.

In Year 2 a series of significant correlations emerged between the HPAT-Ireland and subsections thereof with OSCE Communication Z-scores; OSCE Clinical Z-scores; and Total OSCE Z-scores. However on multiple regression only the

relationship between Total OSCE Score and the Total HPAT-Ireland score remained significant; albeit the predictive power was modest.

Conclusion

We found that none of our selection criteria strongly predict clinical and communication skills. The HPAT- Ireland appears to measures ability in domains different to those assessed by the Leaving Certificate Examination. While some significant associations did emerge in Year 2 between HPAT Ireland and total OSCE scores further evaluation is required to establish if this pattern continues during the senior years of the medical course.

5.2 Background

The use of tests of general mental ability, including aptitude tests, is widespread in the selection of medical students internationally (Patterson and Ferguson, 2010). Examples include the Medical College Admission Test (MCAT), the BioMedical Admissions Test (BMAT), the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and the UK Clinical Aptitude Test (UKCAT) (Association of American Medical Colleges, 2012, Australian Council for Educational Research, 2012b, Cambridge Assessment Group, 2008, UKCAT Consortium, 2012b). The hypothesis that establishing medical school applicants' aptitude at the outset enables one to rank applicants in order of their likelihood to succeed in medicine and become good doctors appears sound on the surface. However the evidence for the effectiveness of such tests, as a selection tool, is mixed and their use is controversial (Prideaux et al., 2011).

The most widely used measure of their effectiveness is predictive validity; the ability of the selection tool to predict medical students' performance in undergraduate assessments. There is consistent (albeit not perfect) evidence for the predictive validity of the MCAT (Donnon et al., 2007, Julian, 2005). In relation to the UKCAT findings are conflicting. Two studies report no significant correlation between UKCAT scores and medical student performance (Lynch et al., 2009, Yates and James, 2010). In a recent follow up study, the authors reported that the UKCAT did not independently predict student performance in clinical course work, whereas prior academic attainment was highly predictive (Yates and James, 2013). Conversely a study from Newcastle University found that the UKCAT significantly

predicted exam performance in all but one major exam over two years (Wright and Bradley, 2010). Two recently published papers found evidence of little or no predictive validity with respect to the UMAT (Poole et al., 2012b, Wilkinson et al., 2011). The modest predictive validity of the BMAT appears to be most related to applicants' performance in the scientific knowledge section (Emery and Bell, 2009, McManus et al., 2011a).

Possible reasons for the variability in reports of predictive validity may stem from comparing research that is limited to single institutions with that from multi-centred studies. Findings reported from single institutions may reflect specific associations with particular curricular or assessment techniques and may not be generalisable to medical schools at large. The reliability and validity of individual medical school assessments, and indeed selection tools may also impact on predictive validity studies. Other potential reasons for variability is the number of students followed up and the duration of follow-up-with larger scale studies, having longer follow up times being more likely to yield valid results.

Although the tests described above all purport to measure aspects of general mental ability there are subtle but important differences between them (See Table 9). One of the most important differences between these tools lies in the domains they assess (Patterson and Ferguson, 2010). For example the MCAT tests both knowledge of physical and biological sciences (termed crystallized intelligence) *and* candidates' logical reasoning and processing skills (known as fluid intelligence). The BMAT also has a section that tests candidates' knowledge of science and mathematics. On the other hand the UKCAT and UMAT focus largely on testing candidates' fluid intelligence in terms of mental processing, reasoning and decision making without testing underlying background knowledge. Whether or not this is fundamental to the differences in predictive abilities has not been fully explored. Knowledge based performance is associated with subsequent success in medical school. However in a large meta-analysis Ferguson et al (2002) have established that only approximately 23% of variance in medical school performance can be explained by previous academic performance. Admission tests and aptitude tests therefore are supported because they may measure domains not measured in school exit exams. However their added value to the selection process must be carefully evaluated.

Table 9 Features of a Variety of General Mental Ability/Aptitude Tests Compared

Assessment tool	HPAT	UMAT	UKCAT	MCAT	BMAT
Target candidates	Undergraduate medical school applicants – predominately school leavers	Undergraduate medical school applicants – predominately school leavers	Undergraduate medical school applicants – predominately school leavers	Medical School applicants – predominately college students	Undergraduate medical school applicants – predominately school leavers
Type of test	MCQ	MCQ	MCQ	MCQ plus written essay*	MCQ, written answers and written essay
Duration	2 hrs 30 mins	2 hrs 45 mins	2 hrs	4.5 to 5 hours *	2 hr
How administered	Paper based	Paper based	Computer based	Computer based	Paper based
Standard Cost to applicant	€95	€161	€78	€181	€50
No of participating schools	5	14	26	Required by almost all medical schools in North America	6
Year it was first used	2009	First used in 1991 at Newcastle University, Australia with expansion to other institutions in 1997/98	2006	Earliest versions commenced in 1946 and have evolved over time. Current format exists since 1992 with some minor adjustments since	2003
Subsections of test	1. Logical Reasoning & Problem Solving	1. Logical Reasoning & Problem Solving	1. Verbal reasoning	1. Verbal Reasoning Skills	1: Aptitude and Skills
	2. Interpersonal Understanding	2. Understanding People	2. Quantitative reasoning	2. Physical Sciences – chemistry, physics and data interpretation	2: Scientific Knowledge and Application
	3. Non-Verbal Reasoning	3. Non-verbal Reasoning	3. Abstract reasoning	3. Biological Sciences – biology and organic chemistry	3: Writing Task
			4. Decision analysis		

Table Footnote: *HPAT*-Health Professions Admission Test Ireland. *UMAT*-Undergraduate Medical Admissions Test. *UKCAT*- UK Clinical Aptitude Test. *MCAT*- Medical College Admission Test. BMAT-BioMedical Admissions Test. * The writing sample section will be removed in 2013 thus shortening the test

Chapter 5 Predictive Validity of HPAT-Ireland (Study 1)

The Ottawa Consensus Statement on assessment for the selection of health care professions and specialty training strongly recommends that further research and evidence, coupled with an examination of supporting theoretical philosophies is conducted to fully inform the international debate on selection (Prideaux et al., 2011).

A new tool, the Health Professions Admission Test- Ireland (HPAT-Ireland), was introduced in 2009 (Australian Council for Educational Research, 2012c). The main impetus for its introduction was the publication of a Government initiated report which recommended that medical student selection, in Ireland, should no longer be based on academic grades alone. The report acknowledged the increasing use of specialized admission tests which recognize the importance of factors other than academic achievement in the development of a doctor (Fottrell, 2006). A key motivator for this recommendation was a sense of social responsibility for widening access to medicine. Candidates from socioeconomically disadvantaged backgrounds are under- represented in Irish medical schools; accounting for less than 4% of all applicants (O'Flynn et al., 2012).

The HPAT-Ireland is designed and independently delivered by the Australian Council for Educational Research-ACER (Australian Council for Educational Research, 2012a). ACER, a not-for-profit organization specialising in educational decision making, also designs the UMAT exam used by over a dozen institutions in Australia and New Zealand. Information on the development of HPAT-Ireland test items and domains, in particular how these domains are blueprinted against the domains of professional competencies, is not readily available.

The HPAT-Ireland is a multiple choice test. In terms of intelligences tested it largely focuses on fluid intelligence. There are three sections. According to the test designers they measure the following abilities: Section 1: Logical reasoning and problem solving consists of 44 multiple choice questions based on a passage of text or a diagram presenting certain information. Applicants are required to analyse and logically reason through the information presented. Section 2: Interpersonal Understanding consists of 36 multiple choice questions based on a scenario representing specific interpersonal situations. Applicants have to

identify, understand, and, where necessary, infer the thoughts, feelings, behaviour and/or intentions of the people represented in the situations. Section 3: Non-Verbal Reasoning consists of 30 multiple choice questions based on recognition of patterns and sequences of shapes. The questions test the applicant's ability to reason in the abstract and solve problems in non-verbal contexts.

Since the introduction of the HPAT-Ireland, undergraduate medical school places are now offered to Irish and EU school leavers based on a combination of second level school academic achievement (predominately measured by the state run school exit exam the "Leaving Certificate Examination"- LCE) and the applicant's performance on the HPAT-Ireland (see Chapter 5- Appendix 2 for full explanation of selection criteria). Applicants from outside of the EU undergo separate selection processes, outside of the scope of this study.

The National Research Group Evaluating Revised Entry Mechanisms to Medicine is a consortium of medical educators, researchers and statisticians who meet under the auspices of the Council of Deans of the Medical Faculties of Ireland. This group is currently examining the relationship between medical students' selection scores and their performance on undergraduate cognitive tests. A preliminary report is available but final reports from this work will be available when the initial cohort has completed the five year undergraduate cycle and will be essential to the validation of these selection tools (O'Flynn et al., 2012).

The focus of this study however is the relationship between student scores in the selection tools and subsequent performance on tests of communication and clinical skills. It is intended that this study will compliment findings from the National Research Group Evaluating Revised Entry Mechanisms to Medicine group and lead to a fuller picture of the utility of these selection tools. Communication and clinical skills are at the heart of sound medical practice. They are cited as two of the eight key domains of good professional practice by the Irish Medical Council (Medical Council, 2010a). According to the CanMEDS framework communication skills are an essential ability that physicians need for optimal patient care (Frank, 2005). The corollary is also evident. A breakdown of complaints to the Irish

Medical Council reveals that communication problems rank in the top three categories of all complaints received from the public (Medical Council, 2010b). A similar pattern exists internationally; a survey of three separate American State Medical Boards reported that unprofessional behaviour accounted for 92% of all violations (Papadakis et al., 2005). Sui and Reiter (2009) contend that the tradition of demanding high levels of academic excellence for selection to medicine has resulted in limiting the number of complaints in terms of cognitive issues. The new challenge is to identify and include selection tools that screen for other important non-cognitive attributes such as communication skills and professionalism (Siu and Reiter, 2009). In modern day curricula, communication and clinical skills are introduced early and built up in a spiral fashion throughout the medical course. A selection tool that could predict strengths in these areas would make a valid contribution to the selection process.

Therefore the aim of this research was specifically to establish whether a relationship existed between student scores on the HPAT-Ireland (including subsections thereof) and the Leaving Certificate Examination and subsequent performance on tests of communication and clinical skills in the early undergraduate years.

5.3 Methods

This study was conducted across two medical schools; National University of Ireland Galway (NUI Galway) and University College Cork (UCC). The competencies of communication and clinical skills are taught at comparable levels throughout the undergraduate courses. At both institutions Objective Structured Clinical Examinations (OSCE) were conducted at the end of Year 1 and Year 2 to assess clinical and communication skills.

Sample

The sample comprised all students who were enrolled, in their first year of study, at the medical schools of NUI Galway and UCC in the academic year 2009. At NUI Galway, students are either enrolled in Foundation Year (GFY) or First Year Medicine (GMed1) depending on their science subjects grades in the LCE. All students were followed up for two years. Undergraduate examination results for the year of intake, Year 1 (academic year 2009–2010) and the following, Year 2

(academic year 2010–2011) were examined and their association with the selection criteria of LCE and Health Professions Admission Test (HPAT-Ireland) determined.

Data

ACER and the Central Applications Office provided the HPAT-Ireland and LCE data. The respective medical schools provided the undergraduate examination results. Written consent to use HPAT-Ireland data was given by all the applicants at the time of sitting the HPAT-Ireland. Ethical approval was granted by the Research Ethics Committee, NUI Galway and tabled in UCC. A linked anonymised data base was used for the study. Only the data enterer and a senior academic administrator had access to the link.

The following data were collated: gender, HPAT-Ireland total and subsection scores (Section 1, 2 and 3); LCE, LCE/HPAT-Ireland combined score, Year 1 structured clinical examination (OSCE) scores (Total and subtests (i.e., communication and clinical components), Year 1 Clinical MCQ (total scores only); and Year 2 OSCE (Total and subtest scores).

The LCE adjusted and LCE/HPAT-Ireland combined scores are based on agreed national selection criteria (Central Applications Office, 2012b). (See Chapter 5 Appendix 2). The minimum entry points for medicine (comprising LCE adjusted plus HPAT-Ireland score) in the two medical schools for 2009 were: UCC 715, NUI Galway 712.

The OSCE is designed to test communication and clinical skill performance and competence (Harden et al., 1975). The stations in this study assessed a range of skills including diagnosis, history taking, medical procedures and interpretation of results. The score sheets at each medical school allowed for the communication and clinical scores to be extracted from each OSCE station total score. Three outcome OSCE variables were computed (Communication, Clinical and Total) for the samples Galway Year 1 (GY1), Galway Year 2 (GY2), Cork Year 1 (CY1) and Cork Year 2 (CY2). Similarly Multiple Choice Examination (MCQ) outcome scores from clinical modules were extracted to reflect communication and clinical attributes of students. (See Chapter 5 Appendix 3 for further details).

While extraction of examination scores was conducted identically in both universities, the OSCE stations were designed and marked differently and so were re-coded as Z-scores (describing each score in terms of its relationship to the class mean score). For GY1, single scores were re-coded for: OSCE Communication (GMed1 and GFY), OSCE Clinical (GMed1 only), Total OSCE (GMed1 only) and finally Multiple Choice Questionnaire (MCQ) which included Communication and Clinical components (GMed1 only). For CY1, single scores were re-coded for: OSCE Communication, OSCE Clinical, Total OSCE, and MCQ. For Y2 at both Galway (minus GFY) and Cork, single scores were re-coded for: OSCE Communication, OSCE Clinical, and Total OSCE. The OSCE stations had both communication and clinical skill components.

Data were analysed using SPSS 17.0 for Windows (SPSS, Inc., Chicago, IL, USA). Descriptive statistics; mean, standard deviation (SD) and median were used to describe continuous variables, and frequencies and percentages to describe categorical variables.

There was no evidence against normality for the continuous explanatory (i.e. HPAT-Ireland and LCE scores) and response variables (i.e. OSCE results) and all were compared between groups (e.g., gender, Foundation Year vs. Med1), using two sample t-tests. The Pearson correlation coefficient was deemed adequate to describe the degree of linear relationship between continuous explanatory and response variables. As outlined in a previous, similar study, limits for correlation coefficients of ≥ 0.20 or ≤ – 0.20 were set as a priori criteria for practical significance (Lynch et al., 2009). Multiple linear regression models were used to identify significant predictors of the OSCE response variables. Variable selection techniques and the magnitude of the variance inflation factor were used to adjust for multicollinearity due to the correlation between the HPAT-Ireland predictors. A significance level of p <0.05 was required for a variable to be included in a model. Given that the percent of missing data varied for each explanatory variable, multiple imputation, using chained equations, was used to impute missing data in order to check the sensitivity of missing data to the identification of significant predictors.

5.4 Results

Demographics

The total sample was 324 (National University of Ireland, Galway, *n* = 193 [1st Med., *n* = 133; Foundation Year (FY), *n* = 60]; University College Cork, *n* = 131). Of this sample, 46% were male (*n* = 150), and 54% were female (*n* = 174). There was no appreciable difference in gender between the two universities (i.e., % Male: Female, 47: 53 and 45: 55, NUI Galway and UCC respectively). The majority of the sample comprised Irish nationals (83%, n = 269). Age was not ascertained; however, given the typical profile of first year medical students at NUI Galway and UCC, it is anticipated that most participants were between the ages of 18 and 21. A total of 131 students (42%) sat neither the HPAT-Ireland nor the LCE in 2009 (largely comprising non-EU entrants who are selected via a separate process, but also those re-sitting exams or who had deferred entry). Twelve participants were selected via a number of special access routes to study medicine and were excluded from further analysis; leaving a final sample of 312. Descriptive statistics for Years 1 and Year 2 outcome variables are outlined in Table 9.

Table 10 Descriptive Statistics for the Variables of Interest and Outcome Measures

Variable	N	M (SD)	Median
Leaving Certificate Examination (LCE)	177	567.57 (21.17)	565.0
Combined LCE/HPAT- Ireland	177	728.36 (14.47)	724.0
HPAT- Ireland Section 1: Logical Reasoning and problem solving (Max = 100)	181	58.82 (7.67)	58.0
HPAT- Ireland Section 2: Interpersonal understanding (Max = 100)	181	56.91 (7.27)	58.0
HPAT- Ireland Section 3: Non-verbal reasoning (Max = 100)	181	60.61 (9.61)	60.0
Total HPAT- Ireland (Max = 300)	181	176.20 (14.38)	174.0
OSCE Year 1 Communication *Z*-scores	277	.002 (1.00)	.62
OSCE Year 1 Clinical *Z*-scores	217	-.01 (.99)	.97
OSCE Year 1 – Total OSCE Score	216	- .02 (1.82)	- .05
OSCE Year 2 Communication *Z*-scores	215	.01 (1.00)	.008
OSCE Year 2 Clinical *Z*-scores	210	-.003 (1.01)	.01
OSCE Year 2 – Total OSCE Score	208	.03 (1.85)	-.002
Multiple choice examination	197	.002 (.99)	-.05

Footnote: SD = Standard deviation; HPAT- Ireland = Health Professions Admissions Test Ireland; OSCE = Objective Structured Clinical Examination.

Year 1 Group comparisons

A series of two sample t-tests, using a Bonferroni adjustment for multiple testing, were carried out to examine potential differences amongst the students in terms of gender and year of entry to programme. NUI Galway students who entered directly from secondary level schooling into 1st year medicine (n = 39), were compared with those entering Foundation Year (n = 53) on the variables of interest and the outcome measures (i.e., selection criteria, and medical school examinations). No significant differences were observed on any of these measures, with the exception of isolated differences in HPAT-Ireland Section 3 performances. Therefore all Galway medical students were treated as a single sample.

Further Bonferroni adjusted two sample t-tests revealed that the average score for males was significantly higher, than the average score for females on HPAT- Ireland Sections 1, t (179) = 3.51, p < .001, d = .52, and HPAT-Ireland Section 3, t (179) = 3.40, p < .001, d = .50, but not on the HPAT- Ireland total score. Due to the small numbers in the gender groups (males who undertook HPAT-Ireland and completed Year 1 examinations n = 46, females n = 61) and lack of gender difference on Total HPAT- Ireland performance, analyses were undertaken for the entire sample [i, ii] (See Chapter 5 Appendix 4 for results section notes)

Year 2 Group comparisons

A similar series of Bonferroni adjusted two sample comparisons were conducted for Year 2 (i.e. gender and year of entry to programme). There were no significant differences between the groups. Students were therefore treated as a unified sample across all further analyses.

Correlations

Table 10 shows the correlations between the Communication and Clinical OSCE marks, for Years 1 and 2 respectively. Table 11 shows the correlations between the selection criteria and student performance on the OSCE and the MCQ represented by Z-scores [iii].(See Chapter 5 Appendix 4). Results for Year 1 are presented below the diagonal and for Year 2 above the diagonal.

Table 11 Correlation Between Communication and Clinical Elements of OSCE (Galway and Cork – Years 1 and 2)

Year 1 Variables	Galway Yr 1 Clinical	Cork Yr 1 Clinical	Year 2 Variables	Galway Yr 2- Clinical	Cork Yr 2- Clinical
Galway Yr 1 Communication: Body Mass Index	.32**		Galway Yr 2- Communication- Chest Pain	.20*	
Galway Yr 1 Communication: Vitals	.42**		Galway Yr 2- Communication - Eye exam	.39**	
Galway Yr 1 Communication: Blood Pressure	.55**		Galway Yr 2 Communication – Gastrointestinal	.55**	
Galway Yr 1 Communication: Urinalysis	.45**		Galway Yr 2 Communication - Joint exam	.55**	
Cork Yr 1 Communication: First Aid		.36**	Cork Yr 2 Communication - Neurology History		.73**
Cork Yr 1 Communication: Clinical Anatomy		.45**	Cork Yr 2 Communication – Cardiology		.65**
Cork Yr 1Communication: Respiratory		.78**	Cork Yr 2 Communication - Neurological Lower Limb Examination		.43**
Cork Yr 1 Communication: Abdominal Examination		.68**	Cork Yr 2 Communication - Neurological Cranial Nerve Examination		.49**
			Cork Yr 2 - Communication - Cardiovascular examination		.71**
			Cork Yr 2 - Communication - Respiratory examination		.65**

Footnote: * P<.01, ** P<.05

Table 12 Correlations Between Selection Criteria and Outcome Measures (Year 1 Below Diagonal; Year 2 Above Diagonal)

Variable	HPAT-Ireland 1	HPAT-Ireland 2	HPAT-Ireland 3	Total HPAT-Ireland	LCE	LCE + HPAT-Ireland	OSCE Comm	OSCE Clin	OSCE Total	1MCQ Total
HPAT- Ireland 1	-						.03	.04	.05	n/a
HPAT-Ireland 2	-.05	-					.27**	.15	.23*	n/a
HPAT-Ireland 3	.29**	-.23**	-				.19	.16	.20	n/a
Total HPAT-Ireland	.70**	.32**	.70**	-			.29**	.21*	.28**	n/a
LCE	-.12	-.16*	-.06	-.18*	-		-.13	.12	.02	n/a
LCE + HPAT-Ireland	.59**	.28**	.61**	.85**	.31**	-	.17	.24*	.24*	n/a
OSCE Comm	-.03	.14	.02	.07	-.03	.04	-			n/a
OSCE Clin	.06	-.002	.13	.13	.04	.14	.66**	-		n/a
OSCE Total	.10	.07	.07	.18	.02	.18	.91**	.91**	-	n/a
MCQ Total	.002	.04	.09	.09	.32**	.27**	.24**	.30**	.30**	-

Footnote: *HPAT-Ireland* = Health professions admissions test; *LR and PS* = Logical Reasoning and problem solving; *N-VR* = Non-verbal reasoning; *LCE* = Leaving certificate examination; *OSCE* = Objective Structured Clinical Examination; *MCQ* = Multiple choice questions; Total = Total scores (i.e., Communication and Clinical elements combined); n/a = not applicable i.e. there is no result for this test in the corresponding year.: * $P<.01$, ** $P<.05$

Chapter 5 Predictive Validity of HPAT-Ireland (Study 1)

Correlations between the selection criteria and outcome measures were undertaken for the entire sample.

In Year 1 none of the selection criteria were significantly related to Total OSCE scores. Neither were they related to either OSCE Communication or OSCE Clinical scores. The LCE and LCE/HPAT- Ireland scores were however, positively associated, with MCQ marks (r = .32 & .27 respectively, p values all < .01).

In Year 2 moderate, significant associations emerged between HPAT- Ireland 2 and Total HPAT- Ireland, and OSCE Communication Z-scores (r = .27, .29 respectively; all p values < .01). Total HPAT- Ireland and LCE/HPAT- Ireland were significantly correlated (r = .21 & .24 respectively; all p values < .05) with OSCE Clinical Z-scores. Finally HPAT- Ireland 2, Total HPAT- Ireland, and LCE/HPAT- Ireland were all significantly correlated with Total OSCE Z-scores (r = .23, .28, .24 respectively; p < .05, .01 & .05 respectively).

Multiple regression analysis

For the outcome measure Year 1 MCQ score the LCE explanatory variable was identified as the single significant predictor (b = 0.02, p = 0.001, 95% CI 0.007 to 0.024) with an adjusted R^2 of 0.09 suggesting a positive, predictive association between LCE scores and Year 1 MCQ. See Figure 5. No significant predictors were identified for the OSCE Communication and OSCE Clinical variables at Year 1. For the Total OSCE response, no explanatory variables were deemed useful for inclusion. However the HPAT- Ireland and LCE combined explanatory variables achieved borderline significance (p = 0.06). These results suggest that, based on the sample provided, none of the selection criteria currently used in the Irish system, are predictive of Total OSCE scores in Year 1.

No significant predictors were identified for the separate OSCE Communication and Clinical response variables at Year 2 response. However, when considering the Total OSCE Year 2 response, Total HPAT- Ireland (b = 0.04, p =0.008, 95% CI 0.01 to 0.07), was identified as a significant predictor with a model R^2 adjusted of 0.07. This suggests that, based on this sample, higher scores on Total HPAT- Ireland scores are related to higher marks on the Year 2 OSCE score however the predictive power is moderate[v]. See Figure 6. See Chapter 5 Appendix 4)
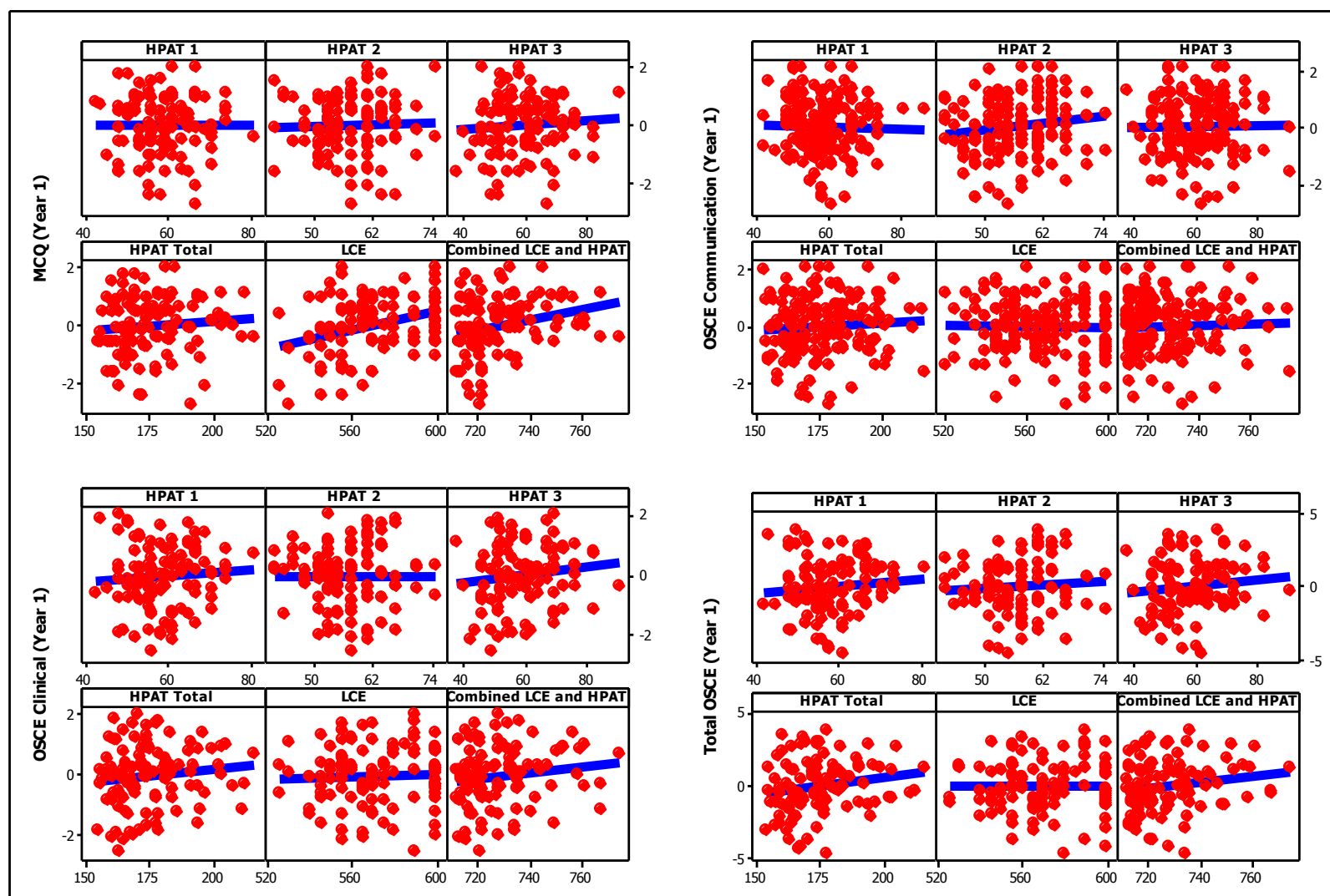
Figure 5 Scatter Plot of Year 1 Results Versus Selection Criteria
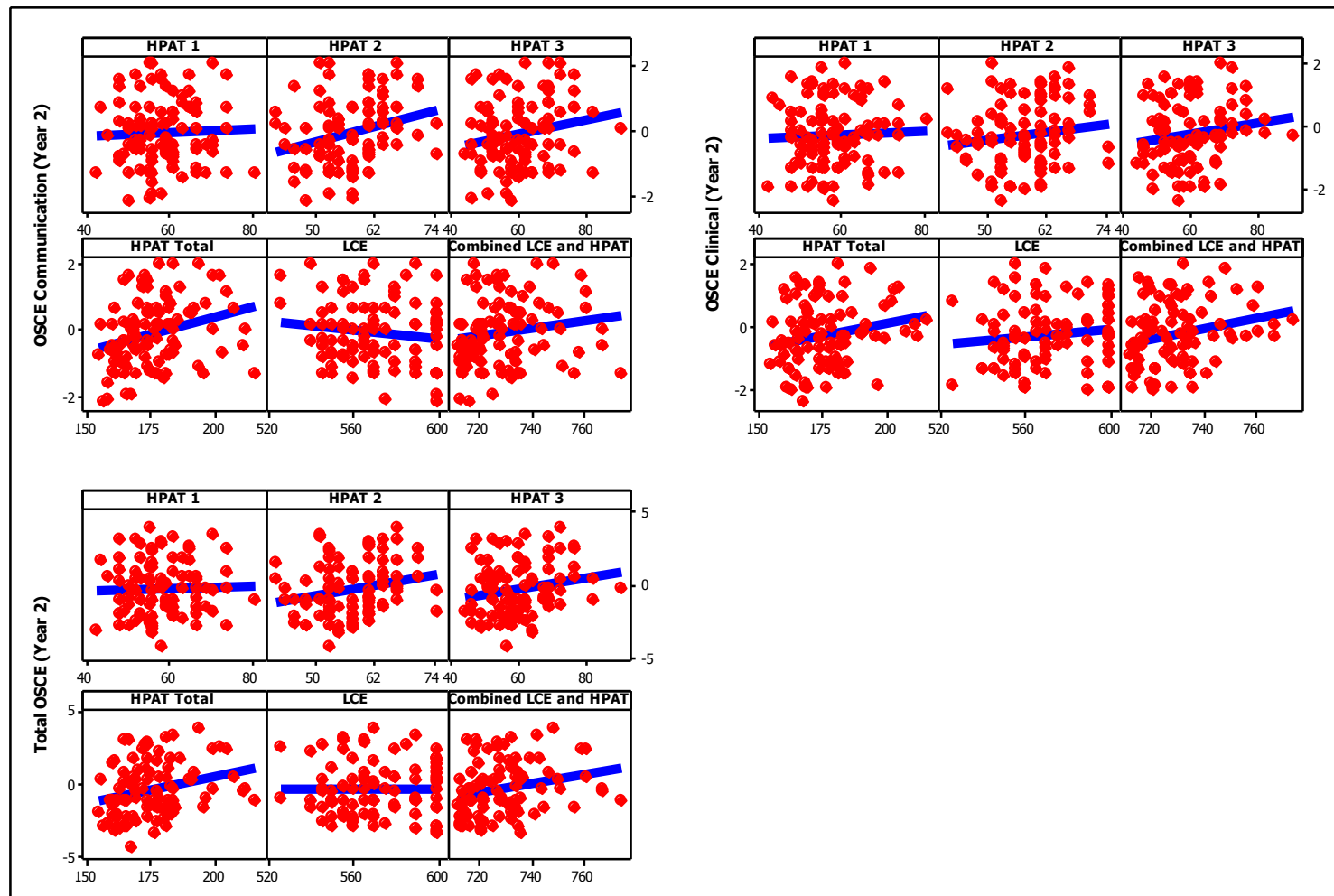
Figure 6 Scatter Plot of Year 2 Results Versus Selection Criteria

5.5 Discussion

This is the first paper to report on a prospective study establishing the predictive validity of the HPAT- Ireland. We conducted a two year follow up of the first cohort of students, selected to two different medical schools, by the LCE and HPAT- Ireland combined. We examined the relationship between applicant performance on the selection tools, and subsections thereof, and subsequent performance on undergraduate tests of communication and clinical skills.

According to Patterson and Ferguson (2010) in criterion related validity studies, such as this one, it is unusual to obtain validity coefficients greater than r =0.5. Values in the range of r = 0.2 to r = 0.29 bracket can be described as low from a practical viewpoint albeit they may reach statistical significance (Yates and James, 2010). In a large BEME systematic review of the predictive values of measures obtained in medical school and later performance in medical practice correlations up to and including r = 0.37 were reported as low (Hamdy et al., 2006). Whereas Julian in an analysis of the predictive validity of the MCAT deems values above r = 0.4 or higher as indicative of a fairly strong relationship (Julian, 2005). When reporting predictive validity studies therefore, it is desirable that correlation coefficients reach at least 0.30 to be considered meaningful (Coates, 2008).

Our first year correlation findings are unremarkable apart from the finding that the LCE and the LCE/HPAT- Ireland correlated with performance in a clinical MCQ. This relationship is to be expected given that both the LCE and the MCQ test in the knowledge domain. This relationship was moderate (*r* = .32 & .27 respectively) and on regression testing only the LCE remained predictive. This is consistent with observations that although prior academic achievement is one of the best predictors of undergraduate medical student performance the majority of the variance in medical student performance lies outside of the influence of this domain (Ferguson et al., 2002).

In Year 2 a number of correlations emerge between: OSCE Communication Z Score and HPAT- Ireland Section 2 and Total HPAT- Ireland (r = .27&.29 respectively); OSCE Clinical Z Score and Total HPAT- Ireland and LCE/HPAT- Ireland (r = .21&.24 respectively); and finally between Total OSCE Z Score and HPAT-

Ireland 2, Total HPAT- Ireland and LCE/HPAT- Ireland (r = .23, .28 &.24 respectively). However although these correlations reach significance, they are at best moderate. Further analysis, using multiple regression, did not robustly support these correlations, with only Total HPAT- Ireland being somewhat predictive of the Total OSCE Year 2 Z Score.

Specific attention was focused on correlations between HPAT- Ireland 2 and OSCE Communication Skills sub-scores as this section of HPAT- Ireland purports to assess interpersonal skills. While we did find a correlation, it only emerged in Year 2 and the strength of this relationship was somewhat disappointing. In terms of Clinical Skills sub-scores, our data does not demonstrate a firm relationship with HPAT- Ireland either. Whilst recognising that performance in summative assessment is influenced by a host of variables, (Ferguson et al., 2002) meaningful correlations between entry criteria and subsequent clinical performance in test conditions would be expected. Indeed for many this is the only added value and justification in the use of adjunct admission tests (McManus and Powis, 2007, McManus et al., 2003). It is possible however that stronger correlations may emerge as the course progresses and the complexity of clinical assessments increases.

In terms of any evidence of incremental validity (the increase in predictive power by the addition of another selection tool) (Patterson and Ferguson, 2010), the data in Table 12 suggest that there may be a possible gain in validity resulting from the addition of the HPAT to the existing selection process. However serial cohort data needs to be analysed to demonstrate this conclusively and multiple regression, at least in Year 1, undermines this observation.

Three types of error are common in validation studies: sampling error due to small sample sizes, poor measurement precision in either the selection tool or the undergraduate assessment tool, and restricted range of scores (Patterson and Ferguson, 2010). Our sample is small by international norms. We attempted to offset this by following up the cohort for two years. Assessment practices at both schools were not identical; and every attempt has been made in the analysis to account for this variance. There is a scarcity of published data on the development and reliability of the HPAT- Ireland. Although it is our understanding

that Medical Schools are provided with confidential annual reports on the performance of HPAT- Ireland, these are not readily available in the public domain. We have not adjusted the data to correct for range restriction in HPAT-Ireland. There is not uniform agreement about whether to routinely correct or not. Any one of these limitations could have reduced the size of the correlation between the selection criteria and undergraduate results observed in our study. It is also possible that the HPAT- Ireland and/or the LCE predict performance outside of the domains we examined.

Two previous publications reported on the HPAT-Ireland (Halpenny et al., 2010, Quinn et al., 2010). However, drawing generalised conclusions from these studies is limited by the fact that in both cases a scaled down, modified version of the HPAT- Ireland was used.

We found that on average males scored significantly higher than females in HPAT-Ireland Section 1 (logical reasoning & problem solving) and HPAT- Ireland Section 3 (non-verbal reasoning). We found no gender difference in our sample in relation to Leaving Certificate or HPAT- Ireland Section 2 scores. This is surprising, as it is well established with respect to the Leaving Certificate that females perform better overall (O'Flynn et al., 2013c). It may be that our sample size was too small to detect true difference between the genders. We report no gender difference on total HPAT- Ireland score. However further research is required before confident statements can be made about the role of gender in HPAT-Ireland performance. Similar concerns have been raised with respect to the UKCAT (James et al., 2010).

Correlations between the LCE and the total HPAT- Ireland showed a very weak negative relationship (r = −.18). This may reflect that the LCE and the HPAT-Ireland are examining different applicant attributes. A recent study compared the predictive validity of the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and Grade Point Average (GPA) (Poole et al., 2012b). GPA was found to be a better overall predictor of medical school exam performance than the UMAT, but the UMAT and GPA together were marginally better again. For senior students the UMAT offered no predictive advantage over the GPA, with respect to communication and clinical skills. These findings are of particular relevance as the

HPAT-Ireland and the UMAT are both designed by ACER and have comparable subsection domains.

The HPAT-Ireland is one of the latest tests of general mental ability to appear on the selection scene. Its design and item content closely resembles that of the UMAT. The inclusion of this test was controversial with many suggesting that reforms in Ireland represented a missed opportunity to introduce a test which demonstrably added value to the selection process (Murray, 2011, Walshe, 2009). For example the incorporation of situational judgment tests looks promising and has the potential to improve the utility of tests of general mental ability as a selection tool (Koczwara et al., 2012). The real benefit of this class of tests is their ability to be taken by large numbers of candidates with minimal cost in terms of finance and medical school faculty time. However the challenge for test designers is to continually improve the design of such tests so that the domains that they assess help us to rank medical school applicants in a meaningful way.

5.6 Conclusions

At present it appears that none of the entry and selection criteria used in the Irish system strongly predict clinical and communication skills performance in the early stages of the course. Some correlations emerge between total HPAT – Ireland scores, HPAT section 2 (measuring interpersonal understanding) and subsequent OSCE performance but correlations are weak to moderate. Further analysis is necessary and is ongoing. Any additional selection test must add value to the selection process in general and it is desirable that such tests enhance of the ability of schools to select candidates with an aptitude for clinical and communication skills. While the HPAT- Ireland appears to measures ability in domains different to those assessed by the LCE it remains to be conclusively established whether this correlates robustly with subsequent medical school performance. This cohort will be followed up for their remaining years in medical school and further evaluations will be conducted to establish if this pattern continues into the senior years of the course.

Chapter 6 Stakeholder Acceptability of HPAT-Ireland (Study 2)

6.1 Abstract

Background

Selection tools for medicine must achieve political validity and enjoy stakeholder acceptability. This qualitative study aimed to establish the perspectives of doctors, from various clinical specialities, on HPAT- Ireland, a new selection tool for undergraduate medical students.

Methods

Fifteen doctors participated over three iterative cycles of recruitment, interviewing and analysis. Prior to interview, participants sat a practice HPAT-Ireland test. HPAT-Ireland has three sections: 1: Logical reasoning/problem solving; 2: Interpersonal understanding; 3: Non-verbal reasoning.

Summary of results

Three themes emerged: Job relatedness; Utility of HPAT-Ireland and Diversity. Sections 1 and 2 were considered very job related however Section 3 was widely criticised for lacking clinical relevance. Doctors did not think that the test would reliably predict future performance. However one third felt it was acceptable as a selection tool in conjunction with academic record. Those who found it unacceptable were influenced by its perceived narrow focus, limited job relatedness, potential for socioeconomic bias, impact on gender and potential for negative influence on student diversity.

Conclusions

A selection tool that does not enjoy the confidence of the medical profession is unlikely to achieve political validity and may ultimately fail, regardless of other objective measures of its effectiveness such as predictive validity.

6.2 Introduction

The British Medical Journal in 1946 highlighted the challenges of medical student selection, considering them "*formidable"* (Smyth, 1946b). Medical schools internationally still struggle to identify selection tools which meet the demands of credibility, fairness, validity and reliability (Prideaux et al., 2011). More recently the requirements of good selection tools have expanded to include political agendas such as widening diversity and future workforce planning (Cleland et al., 2012).

Stakeholder acceptability receives less academic attention but generates much public interest (Beckett, 2008, Cresswell, 2011, Henry, 2010, Mc Donagh, 2010, Molloy, 2010, Neligan, 2009).   Acceptability is highly influential and can determine if a selection tool achieves widespread use or otherwise (Murphy et al., 2008). For example traditional interview and personal statements remain popular despite poor reliability and validity records (O'Flynn, 2010). Stakeholder acceptance and positive student reactions are amongst the recognised standards for judging the quality of a selection tool (Patterson and Ferguson, 2010).

Job-relatedness refers to the extent to which a selection tool measures job content or is perceived to be a valid predictor of job performance (Tippins and Adler, 2011). In order to be legally defensible selection tools for employment need to tie directly to the requirements of the target job (Aamodt, 2012). Regarding medical student selection information on how selection test items blueprint against the professional competencies of medical graduates are often omitted from the literature hence the relevance of certain tests may not be apparent to stakeholders (Patterson et al., 2012b).

Aptitude tests, or tests of general mental ability, remain amongst the most popular selection tools in use. Confidence in these tools has been undermined by conflicting reports of predictive validity and concerns over stakeholder acceptability (Cassidy, 2008, Cleland et al., 2011, James et al., 2010, Julian, 2005, Poole et al., 2012b, Wright and Bradley, 2010).

The Health Professions Admission Test Ireland (HPAT-Ireland) (Australian Council for Educational Research (ACER), 2013) was introduced in 2009 with the intention of widening access to medical school, removing the sole reliance on

academic achievement and bringing Irish medical schools' recruitment policy in line with international norms (Fottrell, 2006). It is similar to the Undergraduate Medical Admissions Test (UMAT). A full description of HPAT-Ireland, comparisons with other aptitude tests and reports of its predictive validity are reported elsewhere (Kelly et al., 2013).

The introduction of HPAT-Ireland met with strongly voiced criticism in the national media. Concerns were raised about the potential for cost to act as a barrier to socioeconomic disadvantaged applicants and possible gender bias, with the popular view holding that males would out-perform females (Donnelly and Heffernan, 2011, Murray, 2011, RTÉ News, 2012a).

One study of stakeholder acceptability: a survey of doctors' knowledge and opinion of HPAT-Ireland has been published (Dennehy et al., 2013). In this study 75% of respondents reported they had little or no knowledge of HPAT-Ireland content, while 70% supported the use of aptitude tests in selection.

Study Aim

This study aimed to establish the perspectives of doctors from a range of specialities on HPAT-Ireland, explore the degree to which the skills measured were considered to be job related, and establish opinions on its acceptability as a selection tool.

6.3 Methods

The study design was qualitative, drawing on the broad precepts and techniques of the grounded theory tradition; a popular methodology used in medical education research (Glaser and Strauss, 2009, Harris, 2002). Medical student selection has been identified as an area that would benefit from more widespread use of "grounded theory approaches" (Harris, 2003). We employed purposeful sampling, an iterative approach to data generation and analysis and the constant comparison technique. Data were coded and categorised using the three step process of open, axial and selective coding supported by detailed memoing (Corbin and Strauss, 2008).

Following ethical approval and a pilot study the main study was carried out in the School of Medicine, NUI Galway and the Western Research and Education

Network (WestREN) (www.western.nuigalway.ie). To ensure that study participants were fully informed of the content of HPAT-Ireland they first sat a full sample paper under examination conditions. As qualified doctors are not the target audience for HPAT-Ireland, comparisons of scores with school leavers are largely unfounded hence individual participant scores are not reported.

Sampling and recruitment

Demographics factored into a purposeful sampling framework comprised: gender, age, specialty and experience in medical education (Coyne, 1997). Sample size was guided by the data saturation principle (Lincoln and Guba, 1985). Sixty-one doctors were invited to participate; 23 agreed however 8 of these opted out due to clinical commitments, 1 declined and no response was received from the remainder.

Study Participants

Fifteen doctors took part (nine male: six female). Participants were assigned an alphabetical code. See Table 12 for outline of sample demographics.

Table 13 Demographics of Study Participants

	Total	Males	Females	Age in years	Speciality	Special interest in medical education
Pilot Study	n=3	n=2	n=1	40-49: n=1 50-59: n=1 age>60: n=1	3 General Practitioners	n=2
Cycle 1	n=4	n=4	n=0	40-49: n=2 50-59: n=2	2 General Practitioners 1 Paediatrician 1 Psychiatrist	n=1
Cycle 2	n=6	n=5	n=1	20-29: n=1 30-39: n=3 40-49: n=1 age>60: n=1	1 Paediatrician 1 Surgeon 1 Clinical Pathologist 1 Nephrology Registrar 2 Clinical lecturers in Medicine	n=4
Cycle 3	n=5	n=0	n=5	30-39: n=1 40-49: n=4	1 General Practitioner 1 Psychiatrist 1 Paediatrician 1 Radiologist 1 Clinical lecturer in Paediatrics	n=2

Data generation

In Phase I participants sat the practice HPAT-Ireland test. See Table 14 for details of HPAT-Ireland. Each participant received a corrected copy of his/her paper. In Phase II semi-structured interviews were conducted by MK (May, 2011). The topic guide was developed and informed by the study objectives (Rubin and Rubin, 2004), relevant literature (Roberto et al., 2005) and ongoing analysis (Corbin and Strauss, 2008). (See Chapter 6 Appendix 2). Interviews took place in NUI Galway or the participant's workplace, lasted between 40-60 minutes, were audio-taped and transcribed verbatim. Participants were given an opportunity to amend the views expressed in their interviews following knowledge of their HPAT-Ireland score; however no participant did so.

Data Analysis

Open coding was conducted independently by MK and AWM. Descriptions of codes and emerging themes were discussed and agreed upon with the other authors and the remaining interviews were conducted and coded in an iterative fashion. N-Vivo10 software was used (QSR International, 2012).

Table 14 Description of HPAT-Ireland Subsections According to ACER

Sections	Duration in Minutes	No. of Questions	Description- (All MCQ type questions)
Section 1 Logical reasoning and problem solving	65	44	Questions based on a passage of text or a diagram presenting certain information. Applicants are required to analyse and logically reason through the information presented.
Section 2 Interpersonal understanding	45	36	Questions based on a scenario representing specific interpersonal situations. Applicants have to identify, understand, and, where necessary, infer the thoughts, feelings, behaviour and/or intentions of the people represented in the situations.
Section 3 Non-verbal reasoning	40	30	Questions based on recognition of patterns and sequences of shapes. The questions test the applicant's ability to reason in the abstract and solve problems in non-verbal contexts.
Total test time	2 ½ hours		

Quality and Rigour

Open questions, summarising and clarification were employed during interviews to encourage the doctors to fully express their views (Patton, 2002). Member checking was carried out with a random sample of participants to ensure that participants' views were accurately represented (Carlson, 2010). (See Chapter 6 Appendix 3). To facilitate reflexivity a coding diary was kept (Corbin and Strauss, 2008). To ensure fair dealing care was taken not to over emphasise the views of any one group of participants and specific attention was paid to both seeking and understanding "deviant" views (Pope and Mays, 2006).

6.4 Results

Three main themes emerged.

Theme 1 Job relatedness

1.1 Section 1: Logical reasoning and problem solving

Four fifths of participants agreed that Section 1 resonated to a "*moderate*" (Dr C) degree with clinical practice:

"..I think it's [Section 1] *assessing something that's an important quality, that physicians and surgeons, or psychiatrists… require, and that's interpreting data"* (Dr P).

Section 1 was considered "*time-pressured*" (Dr V) and "*very difficult*" (Dr E) - features which were thought to indirectly test prioritisation, coping with stress and time-management.

Three doctors, from different clinical specialities, disagreed and felt that Section 1 "*was more to do with scientific and literature review rather than clinical reasoning*" (Dr J).

Suggestions for improvement included reducing the number of questions in Section 1 and constructing the questions so that they could "*..give you the same information much more concisely*" (Dr V).

1.2 Section 2: Interpersonal Understanding

There was consensus that Section 2 tested "*emotional reasoning*" (Dr M) and "*..insight into the way we think and just understanding people*" (Dr R) and almost unanimous agreement across all specialities that it resonated the most with clinical practice.

This was considered distinct from testing communication skills: "*..communication is a huge skill that doctors need…but you can't really test that in a paper format*" (Dr W).

Many suggested "*expand*[ing] *the middle section*" (Dr X), to include analysis of a "*video or even an auditory clip*" (Dr P) ..which would be "*a more realistic way of assessing interpretative empathic skill(s)*" (Dr J).

One participant was especially critical of Section 2: " [it was] *a missed opportunity…..it could have tested …personality factors that may be of relevance to picking future doctors…especially the attitudes of the candidate*" (Dr D) .

1.3 Section 3: Non Verbal Reasoning

Section 3 baffled participants irrespective of gender, speciality or age. They struggled to identify which skill it actually tested and although many suggested it was concerned with "*pattern recognition*" (Dr X) a couple were left " *not even sure what it's testing..*"(Dr S)

There was unanimous agreement that Section 3 was the least job related section

"*Section 3 really eluded me as to where that was in relevance in terms of medicine at all*" (Dr H)

The most common suggestion for improvement was to "*get rid of the third section.*"(Dr P) and in its place "*expand the middle section* [Section 2]" (Dr X).

There were a number of key skills and attributes that participants felt were important to succeeding in medical school and becoming a doctor that were not tested including "*stamina*" (Dr C), "*focus, attention to detail, dedication,*

persistence"(Dr X) and *" really hard work"* (Dr Y). These relate to the theme on utility below.

Theme 2 Diversity

2.1 Gender, age and maturity

There was broad awareness of media coverage of potential gender bias across the specialities, gender and age groups of participants. Participants considered it "..*a complex assessment for either male or female candidates*" (Dr V). A third of participants thought that "*boys may do better on Section 3"* (Dr H) and *"women would do better"* (Dr C) on Section 2. Yet there was no perception that the overall test was likely to be gender biased. There was consensus that HPAT-Ireland was "very *challenging for a seventeen/eighteen year old …. it seemed to be looking for a fair degree of maturity"* (Dr J).

2.2 Socio economic group

Participants were concerned that socioeconomic background could impact on applicant performance on HPAT-Ireland through access to commercial coaching:

 "You have to play fair…. you do better if you repeat it or if you get loads of grinds." (Dr Y).

Participants appreciated that one of the reasons for introducing HPAT-Ireland was to *".. try and diversify intake"* (Dr J). However they did not think that HPAT-Ireland would *"… have much effect on* [widening]*diversity, because no matter what system you put in place, the articulate middle classes will jump that particular hurdle"* (Dr J).

2.3 Student skills

Participants felt that HPAT-Ireland may favour applicants with strong logical, analytical and reasoning skills, potentially reducing student diversity at a potential cost to the profession: "*So if you look at advances in medicine, its often the most creative people that come up with the most ingenious solutions"* (Dr K).

173

Theme 3 Utility of HPAT-Ireland

3.1 Interpreting HPAT-Ireland results

Almost two thirds of participants likened HPAT-Ireland to a standard "*aptitude* [test]" (Dr Y) and interpreted the result predominately as a measure of one's "*logical reasoning*" (Dr Y) abilities. The majority were struck by the extent to which it tested a "*different set of skills*" (Dr E) to the Leaving Certificate, the Irish state run secondary school exit exam.

Participants wondered   "*.. if you were good at the HPAT does that mean that you'll be a good doctor, if you were bad at the HPAT does that mean that you will be.. a bad doctor?* (Dr W).

In response the majority were "*reluctant to put too much weight..*"(Dr C) on the result of HPAT-Ireland for a number of reasons. It was not considered "*specifically geared at medical* [selection]" (Dr Y); there was doubt that performance on paper would match performance in clinical situations and it was considered a very "*tall order*" (Dr J) for a selection tool to reliably predict future performance "*..when you think of the diversity of doctors and the varied skills they actually need, so such a test in my opinion doesn't exist, never has and presumably never will."* (Dr H).

Factors such as the quality "*..of under graduate training*" (Dr V), the "*enthusiasm and motivation*" (Dr K) of the student and the omitted skills referred to in Theme 1 all had an important role to play.

3.2 Acceptability of HPAT-Ireland as a selection tool for medicine

Acceptability of HPAT-Ireland as a selection tool varied. Participant's clinical speciality, age or gender did not appear to consistently influence opinions.

Seven participants found it broadly acceptable however many of these remained sceptical of its ability to select good future doctors. These participants considered that HPAT-Ireland "*tests a different set of skills*" (Dr E) to the Leaving Certificate, was in keeping with "*international*" (Dr S) norms and "*levelled the playing field*" (Dr H) in terms of combining it with applicants' academic record .

"..it's just one more tool to allow you to select from a very good cohort of students". (Dr S)

Four participants were opposed to HPAT-Ireland because they perceived it to have limited predictive validity, potential for socioeconomic bias and that it was an "*extra hurdle to jump over*" (Dr V). One doctor was particularly opposed to it "*.I think it falls far short* [as a selection tool]*... as a matter of fact I think it's both random and dangerous*" (Dr D).

Four participants remained "*undecided*" (Dr Q) about its role and acceptability. This group reserved judgement until *"more research and evidence"* (Dr Q) emerges to inform the debate one way or the other: "*So, I think it was a noble attempt to try and check different things, and time will tell whether it has achieved that or not"* (Dr J).

6.5 Discussion

The degree to which a selection tool measures skills considered to be job related impacts greatly on stakeholder acceptability. A strength of this study is that participants sat HPAT-Ireland, ensuring that opinions were informed by knowledge of the test. Perceptions of its job relatedness were based on the skills participants used to complete HPAT-Ireland (response process) and the degree to which those skills resonated with their clinical practice (content evidence), two of the five sources of evidence for construct validity (Downing, 2003). Adequate construct validity is a requirement of good selection tools and is defined as the extent to which a tool measures the construct that it is intended to measure (Cleland et al., 2012). Powis (1994) describes the construct being tested as suitability to *"become good medical students....and ultimately good doctors"*.

Section 2 was thought to test skills that relate well to those used by doctors. Hence this section was deemed most "job related" by participants across all specialities, gender and ages reflecting the central role of interpersonal communication in medicine. Participants recommended expanding the section to provide a more comprehensive test of interpersonal skills. Lievens and Sackett (2012) established that Situational Judgement Tests of interpersonal skills show acceptable predictive validity for future job performance and these could be

readily incorporated into HPAT-Ireland. One participant criticised Section 2 for missing the opportunity to assess personality factors. In their comprehensive model for the selection of medical students Bore et al (2009) argue for the inclusion of personality factors on the grounds that in the workplace they have incremental predictive validity over cognitive abilities alone. Adam et al (2012) demonstrated a range of weak to moderate correlations between written tests of personality factors and student examination results and outcomes of tutor assessments. A recent review concludes that personality assessment may provide a moderate level of predictive validity however more research is required particularly with regard to stakeholder acceptability (Cleland et al., 2012).

Section 1 resonated to a moderate degree with practice although the intricacies of clinical reasoning were deemed different to those of general logical reasoning. Interestingly the indirect concerns of time management, coping under pressure and prioritisation were considered congruent with the demands of clinical practice. However their relevance might not be obvious to medical school applicants. Thorough explanation of the rationale supporting test items has been shown to positively impact on test acceptance (Patterson et al., 2011).

Section 3 did not find favour with the doctors in this study. According to ACER the main reason for assessing non-verbal reasoning is to gain a measure of cognitive ability independent of language ability and specific cultural knowledge (Australian Council for Educational Research (ACER), 2007). However participants in this study struggled to find any resonance with clinical practice and hence did not see the relevance of this section. Participants were also concerned that coaching could improve performance presenting an additional barrier to applicants from lower socio-economic groups. There is evidence that coaching improves performance on Section 3 of both HPAT-Ireland (O'Flynn et al., 2012) and UMAT (Griffin et al., 2012, Griffin et al., 2008). Hence there is a strong argument for removing this Section entirely from the paper.

"*Consequential*" evidence requires that selection tools do more good than harm and is a measure of their effect on applicants, faculty, patients and society (Downing, 2003). A motivation for introducing HPAT-Ireland was widening diversity. Doctors in this study did not consider that HPAT-Ireland would achieve

this aim, a view supported by enrolment data which has not demonstrated any change in the socioeconomic background of candidates since its introduction (O'Flynn et al., 2012). Lack of diversity however may reflect the *applicant pool* (O'Neill et al., 2013). A study of UK teenagers, found that academically able students from lower socio-economic backgrounds, viewed medicine and university as alien to them and restricted to "*posh*" people (Greenhalgh et al., 2004). Further consideration needs to be given to school outreach workshops, mentoring programmes, fostering links with disadvantaged schools and expansion of special access routes.

Participants' concerns about the potential impact of gender on performance were restricted to applicant performance on test subsections, which tended to balance out between Sections 2 and 3. This is in keeping with reports which have not demonstrated a significant gender discrepancy in *overall* HPAT- Ireland results but do indicate that males slightly outperform females on Section 1 and 3 (O'Flynn et al., 2013c). Similar patterns exist with the UMAT (Griffin et al., 2012). Selection tools that are *perceived* as unfair can deter potential medical students from applying which would be considered a profoundly negative consequential effect (Patterson et al., 2012a).

Test validity is a measure of the weight of evidence that supports the *interpretation* of results for a *given purpose* at a *certain time* (Downing, 2003). It was the predominant view of doctors in this study that scores on HPAT-Ireland should not be interpreted as a measure of applicants' likelihood to be a good doctor. Academic record is no longer considered sufficient grounds for selection (Powis, 2010) due in part to potential for socioeconomic bias and diminished ability to differentiate between top performing applicants. It is in this respect that HPAT-Ireland received moderate acceptability amongst study participants as a tool to further enable rank ordering of applicants. This interpretation and use is not fully in keeping with testing the construct according to Powis (1994) and may be at odds with the perception of the general public. Consideration of alternative adjunct selection tools including Multiple Mini Interview or Situational Judgement Tests which may have better predictive profiles and hence be more likely to fulfil both aspects of the construct is worthy of more open discussion and debate.

Political validity differs from construct validity in that it is a measure of the extent to which these stakeholders consider the tool to be appropriate for use in selection (Cleland et al., 2012). Acceptability, job relatedness *and* the interpretation of applicant performance on the selection tool are all important determinants of political validity.

Limitations of this study include that for pragmatic reasons most doctors were interviewed prior to receiving their marked HPAT-Ireland paper. Although no participant wished to amend their original interview after receiving their results it is possible that this may have altered their views. Secondly although this study utilised many of the broad precepts and key techniques of grounded theory analysis it is not a grounded theory study per se. Elsewhere it has been accepted that the grounded theory approach can be justified without the generation of a theory as long as it is acknowledged that this was not the express intention of the work (Coyne and Cowley, 2006, Gantley et al., 1999). In accordance with current recommendations we have clearly outlined which aspects of grounded theory were employed (Kennedy and Lingard, 2006).

6.6 Conclusions

Due to the high stakes nature of medical student selection there are many stakeholder groups of which the medical professions itself is a key one. Doctors were critical of the lack of clinical relevance of Section 3 and the potential for negative impact on diversity. Improvements to test design could impact positively on its job relatedness and subsequent acceptability. A selection tool that does not enjoy the confidence of the medical profession is unlikely to achieve political validity and may ultimately fail, regardless of other objective measures of its effectiveness such as predictive validity. Almost seventy years on the task of selection remains "*formidable"* (Smyth, 1946b).

Chapter 7 Feasibility of MMI in an Irish Setting (Study 3)

Chapter 7 Feasibility of MMI in an Irish Setting
(Study 3)

7.1 Abstract

Multiple Mini Interview (MMI) is a new selection tool for medical school applicants. Developed at McMaster University in 2004 it comprises a series of interview stations designed to measure performance across a range of competencies including communication skills, team work, and ethical reasoning.

In September 2012, 109 First Year Medical students underwent the MMI. It consisted of 10 stations. The median total score, out of 150, was 100 (min 63, max 129). Cronbach Alphas for the 10 individual stations range from 0.74 to 0.80. Overall Cronbach's Alpha of MMI items was 0.78.   Staff and student feedback was positive. The outline cost per student was estimated at €145.

This study demonstrates that it is feasible to hold a MMI with acceptable levels of reliability and stakeholder approval in an Irish setting. Further work is ongoing to establish the concurrent and predictive validity of MMI in this cohort of medical students.

7.2 Introduction

Medicine is a highly sought after career choice amongst Irish school leavers. In 2010, for example there were 3,292 applicants for 434 medical school places (O'Flynn et al., 2013b). An ideal selection tool would reliably rank applicants in accordance with valid criteria enabling predictions that they would make good doctors (Patterson and Ferguson, 2010). However there are many facets to being a good doctor (Albanese et al., 2003). Designing a selection tool that measures across these facets in a reliable and valid way is challenging. One newer tool that is gaining popularity is the Multiple Mini Interview (MMI) (Eva et al., 2004a).   First developed at Mc Master University in 2004 it comprises a series of interview stations, each designed to measure performance on a different non cognitive competency such as communication skills, team work, moral reasoning and ethical decision making. It takes place in a timed circuit, similar to an OSCE.

MMI is emerging as a promising selection tool with respect to its ability to predict student performance in undergraduate tests (Husbands and Dowell, 2013, Pau et al., 2013). A recent systematic review has indicated that MMI is growing in

popularity across Canada, UK, Australia and USA. It has been applied in both undergraduate and graduate Medical Schools as well as higher professional training. Its use has spread to dental, health sciences, pharmacy and veterinary programmes. The average number of stations is 10, each lasting 8 minutes and generally with one interviewer per station (Pau et al., 2013). The aim of this study was to establish the feasibility of running a MMI in an Irish setting.

7.3 Methods

All students enrolled, for the first time, in First Year Medicine, NUI Galway, September 2012 were eligible. Ethical approval was granted by NUI Galway Research Ethics Committee. Participation was voluntary. Volunteers were entered in to a draw for an iPad. Funding was granted by WestREN (http://westren.nuigalway.ie/). Interviewers and administrators were recruited from the School of Medicine, Nursing and Health Sciences and Western Training Programme in General Practice.   Role-players were selected from the Simulated Patients Group. A small number of senior cycle medical students assisted with role playing and administration. One of the authors (MK) and an acting coach trained the role players. MMI interviewers received written information in advance while interviewer briefing on station content and marking grids was conducted immediately before the MMI (by MK). Interviewers underwent online training to use OMIS software to electronically mark the MMI (Kropmans et al., 2011). This software enables live psychometric analysis of station and interviewer performance.

The MMI circuit consisted of ten, seven minute stations, with one interviewer per station. Material for the stations was provided by Dundee Medical School and blueprinted against the Irish Medical Council's eight domains of professional practice (Medical Council, 2010a). Minor station modifications were made to ensure authenticity in an Irish setting. Five stations involved an interviewer, a role-player and the candidate. The other five stations were interview based (one interviewer: one candidate). Each station was scored across three domains and one global rating scale.   Domain scores ranged from 0-5 (0= poor; 5 =excellent) with detailed written descriptors of excellent and poor performances. Global score were on a five point scale ranging from unacceptable to excellent

performance. The MMI circuit ran over two days. Post MMI students received a debriefing lecture. In addition students obtained individual written feedback on their performance. Post MMI student and interviewer evaluation was collected anonymously by electronic questionnaire, entered into SPSS and analysed.

7.4 Results

There were 241 eligible students. Of these**,** 109 students (45% of class) completed the MMI comprising 41 males, 68 females. There were 64 (58.7%) EU nationals and 45 (41%) were Non-EU which is reflective of class norms. There were 49 interviewers, nine role-players, nine senior-cycle medical students and three administrators. An MMI cycle consisted of two parallel circuits. The MMI cycle was repeated 6 times to accommodate up to 120 students. Each station was scored out of a total of 15. The median total score, out of 150, was 100 (min 63, max 129). Cronbach Alphas for the 10 individual stations range from 0.74 to 0.80. Overall Cronbach Alpha of MMI items was 0.78.

Feedback was returned by 71 students (65% response rate). Ninety per cent either agreed or strongly agreed that the content of the MMI was relevant to their understanding of the practice of medicine (See Figure 7). To put that in context of the students who had undergone a traditional selection interview (n=30) only 60% thought that the issues raised during the interview were relevant, correspondingly only 38% (n=47) of students who had taken an admission test (such as the HPAT) (n= 47) thought the issues covered in the test was relevant. There was no significant difference in these opinions based on student gender or nationality.

Students rated the suitability a number of selection tools on a five point Likert scale ranging from very unsuitable to very suitable (see Figure 8).   MMI was considered almost on a par with academic achievement as suitable grounds for selecting medical students MMI 73%; Academic achievement 79%; whereas the other tools were less favoured. Student feedback was collected on the best and worst aspects of the MMI. Representative favourable feedback is that MMI "…*allows for a more wholesome picture of the candidate*".   A criticism was that "*The time allocated for each station is too short. I didn't really have time to think*

of the issues asked". Another concern was that interviewers might be subjective rather than objective in marking applicants.
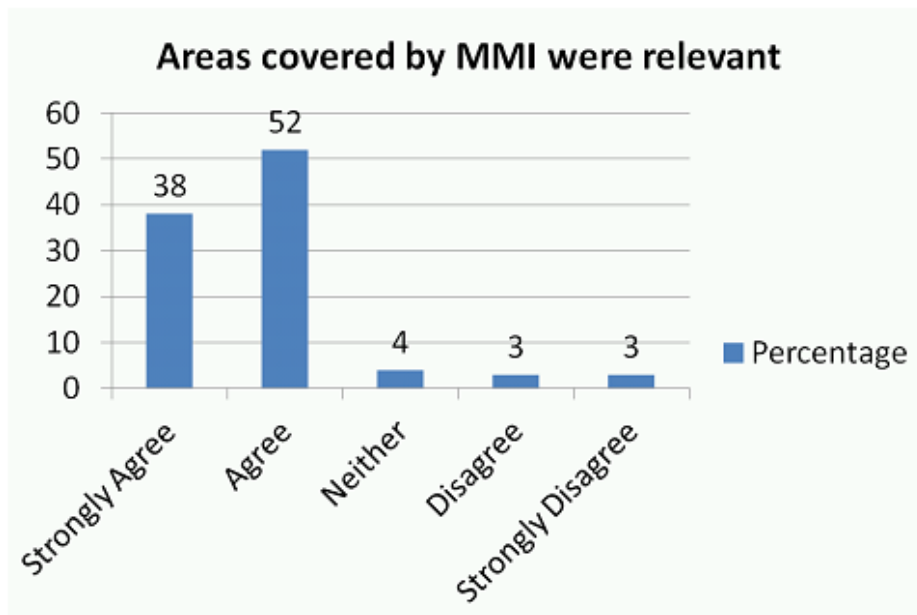
Figure 7 Student Feedback on the Relevance of the MMI
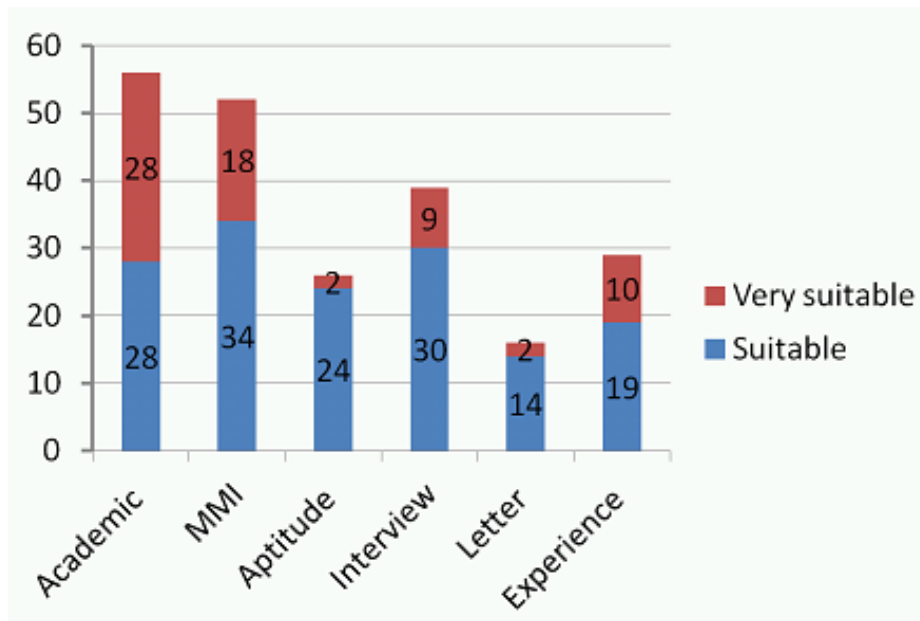
Figure 8 Student Feedback on the Suitability of a Variety of Selection Tools

There was a 49% (n=24) response rate to online interviewer feedback. Three quarters of interviewers felt that MMI was relevant to the practice of medicine and that the stations reasonably tested candidates' ability. Almost two thirds thought that the content was sufficiently important to the practice of medicine to warrant inclusion in a selection test. The majority of respondents (70.83%) thought that an MMI would be a useful addition to medical student selection in Ireland (see Table 15). Interviewers felt that the main advantage of MMI was its ability to "*assess candidates' actual performance objectively and consistently in tasks that are relevant to performing as a clinician*" (n=4). A second advantage was that it was a "*good assessment of non-cognitive and inter-subjective skills*" (n=4). As one interviewer put it "*The MMI seems to provide a 'best of all' option in terms of selection methods by striking a balance between objectivity, aptitude, and 'the human factor*." However MMI was considered "*Expensive in terms of personnel, time and resources*" (n=3); with the "*Potential for enhanced inequity in student selection due to potential for preparation at 'grind schools'*" (n=4) and the

potential exists for quieter or international students to underperform *"The MMI can struggle to allow for cultural and language differences"* (n=6).

An analysis of the cost involved in the running an MMI was conducted, based on an assumption that hosting the MMI would be external to core academic activity and hence would incur additional costs. As interviewers and administrators who took part in this study received no payment, we estimated costs based on typical OSCE rates for licencing exams (see Table 16). The total cost excludes the cost of investing in software support and station development. The cost per applicant, based on 120 applicants, is estimated at €145 per person.

Table 15 Interviewer Feedback on the MMI

Question Stem	Agree or Strongly agree	Neither agree nor disagree	Disagree or strongly disagree
The areas covered by the MMI were relevant to the practice of medicine.	75% (n=18)	8.33% (n=2)	**16.67% (n=4)**
The MMI stations were constructed in a manner that tested the candidates' ability to a reasonable degree	79.17% (n=19)	12.50% (n=3)	**8.33% (n=2)**
The issues raised by the MMI stations were sufficiently important to the practice of medicine to warrant inclusion in a selection test	62.5% (n=15)	25% (n=6)	**12.5% (n=3)**
MMI would be a useful addition to medical student selection in Ireland	**70.83% (n=17)**	**8.33% (n=2)**	**20.83% (n=5)**

Table 16 Estimated Costs of Running the MMI

Item	Unit cost	Estimate cost for MMI for 120 applicants
Interviewer expenses (20 interviewers plus 2 backup interviewers =22)	Based on OSCE examiner rates of €500 per full day	11000
Administrative support X4	Based on OSCE administration rates of €250 per full day	1,000
10 actor / role players	Actor training @€50 per actor	500
	MMI @ €150 euro per day	1,500
Acting Coach	Sourcing and training actors estimated @ €500	500
	Supervision of acting at MMI @ €150 per day	150
Venue rental	€2,000 per day	2000
Catering for 37 people (lunch/ teas/ coffee)	Estimated @€15 per person	555
Office supplies / laminating / paper/ photocopying / Various MMI station material props	€200	200
Estimated total cost		€17405
Estimated cost per applicant		**€145.04**

7.5 Discussion

Medical student selection is a complex and emotive issue. At its heart is a sense of social responsibility to select, from amongst hundreds of very able applicants, those best placed to become good doctors. This study demonstrated that it is feasible to hold a MMI in an Irish setting. Due to the level of expertise with OSCE examinations, the move to MMI proved both practicable and achievable. Student performance was comparable to that of Dundee applicants as was the station item Cronbach alphas which demonstrated a satisfactory level of reliability (Dowell et al., 2012). Station material was blueprinted against the eight domains of professional practice, thus ensuring both face and content validity. Feedback from students indicated that the test achieved an acceptable level of approval amongst this stakeholder group. Interviewers were overall

supportive of MMI as a selection tool. This is in keeping with reports from MMI
feasibility studies internationally which also note stakeholder approval (Brownell
et al., 2007, Razack et al., 2009).

Economics are an important aspect of feasibility as MMIs are labour intensive
and potentially costly. Our estimated costs assume that no cost is absorbed by the
respective medical schools which would not necessarily be the case.
Redistributing these costs to applicants risks giving rise to socioeconomic bias and
due consideration needs to be taken to avoid this. In Canada and UK faculty
involvement in selection is seen as core academic activity. Evidence from
international experience is that MMIs are a more economical use of faculty time
than traditional interviews. Costs can also be kept to a minimum by utilising
senior cycle medical students as trained role players and venue rental costs could
be avoided if MMIs become core institutional activity (Rosenfeld et al., 2008).

The study did have some important limitations. Participants were already
selected to medical school, therefore the range of scores achieved is unlikely to
represent the spectrum from an applicant pool. Secondly the students who
volunteered for the study may differ from their class mates in some important
respects. This paper reports on the feasibility and face validity of the process of
MMI as opposed to establishing its concurrent or predictive validity. Further work
is required and is currently underway to determine these in an Irish setting as well
as establishing the impact of age, gender and nationality on performance.

The real question is whether MMI would be implementable on a national
level. The main determinant of this is the numbers of places available in medical
school, coupled with the ratio of applicants called to MMI for places offered. For
example with approximately 450 undergraduate places a ratio of 3:1 would imply
1,350 applicants are called to MMI. Such numbers would be best accommodated
via a central process. It may be possible to shortlist applicants by rank ordering
them either on Leaving Certificate or Leaving Certificate/HPAT combined scores.
The timing of release of Leaving Certificate results would necessitate hosting the
MMI in late August. Scheduled MMI dates could be announced by the CAO at the
time of application to medicine, with advice for all applicants to keep these dates
available. Invites to MMI could be made via the CAO system, once Leaving

Certificate/ HPAT results were available. The use of OMIS software in the marking of MMI would facilitate a quick turnaround of final offers to medicine. MMIs require time, effort and commitment on the part of medical schools. One may ask is it worth it? Reforms to entry and selection to medical school in Ireland have provoked debate and are under review (Kelly et al., 2013, O'Flynn et al., 2012) Attrition in medical schools in Ireland is low and therefore those enrolled are highly likely to graduate (Maher et al., 2013). Therefore is it not a necessity to employ the best available tools to ensure we enrol, educate and graduate the most suitable candidates? We contend that the use of MMIs is worthy of further consideration in the Irish context.

Chapter 8 Predictive Validity and Stakeholder Acceptability of MMI (Study 4)

8.1 Abstract

Background

International medical students, those attending medical school outside of their country of citizenship, account for a growing proportion of medical undergraduates worldwide. This study aimed to establish the fairness, predictive validity and acceptability of Multiple Mini Interview (MMI) in an internationally diverse student population.

Methods

This was an explanatory sequential, mixed methods study. All students in First Year Medicine, National University of Ireland Galway 2012 were eligible to sit a previously validated 10 station MMI. Quantitative data comprised: demographics, selection tool scores and First Year Assessment scores. Qualitative data comprised separate focus groups with MMI Assessors, EU and Non-EU students.

Results

109 students participated (45% of class). Of this 41.3 % (n=45) were Non-EU and 35.8% (n=39) did not have English as first language. Age, gender and socioeconomic class did not impact on MMI scores. Non-EU students and those for whom English was not a first language achieved significantly lower scores on MMI than their EU and English speaking counterparts (difference in mean 11.9% and 12.2% respectively, P<0.001). MMI score was associated with English language proficiency (IELTS) (r=0.5, P<0.01). Correlations emerged between First Year results and IELTS (r=0.44; p=0.006; n=38) and EU school exit exam (r=0.52; p<0.001; n=56). MMI predicted EU student OSCE performance (r=0.27; p=0.03; n=64). In the analysis of focus group data two overarching themes emerged: Authenticity and Cultural Awareness. MMI was considered a highly authentic assessment that offered a deeper understanding of the applicant than traditional tools, with an immediate relevance to clinical practice. Cultural specificity of some stations and English language proficiency were seen to disadvantage international students. Recommendations included cultural awareness training for MMI assessors, designing and piloting culturally neutral stations, lengthening station duration and providing high quality advance information to candidates.

Conclusion

MMI is a welcome addition to assessment armamentarium for selection, particularly with regard to stakeholder acceptability. Understanding the mediating and moderating influences for differences in performance of international candidates is essential to ensure that MMI complies with the metrics of good assessment practice and principles of both distributive and procedural justice for all applicants, irrespective of nationality and cultural background.

8.2 Background

International medical students, those who attend medical school outside of their country of citizenship, account for a significant proportion of medical school undergraduates (Hallock et al., 2007). For example they make up 7.5 % of medical school undergraduates in the UK, (Medical Schools Council, 2014b) and account for over 15% of Australian medical graduates (Medical Deans Australia and New Zealand, 2014). International medical graduates (IMGs) account for approximately 25% of practicing physicians in the USA, with returning USA citizens accounting for a growing proportion of these (Boulet et al., 2009). According to the OECD the number of students migrating in pursuit of higher education is on the increase rising from 0.8 million worldwide in 1975 to 4.3 million in 2011 (OECD, 2014).

Interview remains a stalwart of medical student selection internationally (Cleland et al., 2012, Poole et al., 2012a). A survey of First Year medical students in Ireland (n=291) revealed that over 70% were in favour of selection interviews with international students significantly more likely to hold this view (p≤ 0.001) (Stevens et al., 2014). Multiple Mini Interview is the most robust structured interview in use today; avoiding many of the shortcomings of the traditional interview (Eva et al., 2004a, Razack et al., 2009).

A recent systematic review of MMI specifically called for research to determine its acceptability in different cultural contexts which has particular relevance for international medical school applicants (Pau et al., 2013).

There is limited exploration of the challenges of international medical selection in the literature, with existing research largely restricted to postgraduate

professional training. There is controversy with respect to the performance of IMGs in postgraduate examinations and progression on higher professional training programmes. Three recent BMJ articles highlight the complexities and sensitivities in this area particularly with regard to issues of fairness and equivalence, issues which are also relevant to selection (McManus and Wakeford, 2014, Peile, 2014, Tiffin et al., 2014b).

Stakeholder opinion is an important consideration in the *political validity* of selection tools (Cleland et al., 2012). In one study of paediatric higher professional training in the UK international graduates were more likely than UK graduates to prefer MMI over traditional interview as a selection tool (p=0.01) (Humphrey et al., 2008). A study of Canadian graduates and IMGs, applying to postgraduate training, found that MMI was considered reasonable by 88% of candidates and 90% of assessors although there was no differentiation of the views of Canadian and IMGs, nor were assessors' views reported (Dore et al., 2010). Patterson et al established the usefulness of *organisational justice theories* to conceptualise stakeholder reactions to selection tools in medicine (Gilliland, 1993, Patterson et al., 2011). Two of these theories are particularly relevant: *Distributive justice* relates to the fairness of the selection *outcome* - such as securing a medical school place, in terms of equal opportunity and equity. *Procedural justice* relates to the fairness of the selection *process* in terms of job relevance, characteristics of the test and interpersonal treatment (Gilliland, 1994). These values complement the criteria expected of selection tools as defined by the Ottawa consensus statement, which conceptualised selection as "*assessment for selection*" and recommended that the principles of good assessment be applied to selection tools (Prideaux et al., 2011).

The five undergraduate medical schools in Ireland host a significant number of international students annually.   MMI is not routinely used in the selection of medical students in Ireland. However a pilot study in 2012 established the feasibility of running a MMI in an Irish setting (Kelly et al., 2014a).

The aims of this study were to run an experimental MMI in an internationally diverse student population to establish its a) Fairness with respect to age, gender, socioeconomic group and candidate background b) Predictive validity in year one

assessment outcomes c) Stakeholder (MMI candidates and assessors) acceptability.

With respect to selection the term fairness implies that everyone has a fair opportunity and chance of being selected based on talent and merit alone (Milburn, 2009). Predictive validity determines how well scores on a selection measure predicts some future outcome, such as work performance or examination scores (Cleland et al., 2012). Stakeholder acceptability describes the views of those who are affected by or can affect selection processes, such as applicants, employers, parents and the regulator, on the suitability and appropriateness of the tool for use in selection (Cleland et al., 2012, Freeman, 2010). In this paper we discuss stakeholders' views of the acceptability of MMI through the lens of organisational justice theory.

This study was set in the School of Medicine, National University of Ireland, Galway an undergraduate medical school with over 1000 students, in collaboration with the Medical School Dundee (UK) and University College Cork, Ireland.

The selection criteria for medical students in Ireland are determined by whether their country of origin is within or out with the European Union (EU) (See Table 16). EU school leavers apply via a single national system administered by the Central Applications Office (CAO) and are selected on the basis of academic record and score in an adjunct admission test -HPAT-Ireland, with an additional requirement for an English language test if prior academic record has not included English as a subject (Australian Council for Educational Research, 2014a, Central Applications Office, 2014 , Central Applications Office (CAO), 2014). There is no national protocol for selection of Non-EU students however most schools require academic record, evidence of English language proficiency and interview. Non-EU applicants are not required to sit HPAT-Ireland. The most commonly used test of English proficiency is the International English Language Testing System (IELTS) (International English Language Testing System 2014). This test comprises four domains; Listening, Reading, Writing and Speaking. At present in Ireland the minimum acceptable IELTs score is 6.5.

Table 17 Criteria for Medical Student Selection in Ireland

EU Applicants Selection Criteria	Non-EU applicants Selection Criteria
Academic Record: Leaving Certificate Examination or equivalent	Academic Record : Grade Point Average
English Language Proficiency- if required	English Language Proficiency: International English Language Testing System (IELTS) or equivalent
Health Professions Admission test –Ireland (HPAT-Ireland)	+/- Traditional Interview
	+/- Others including MCAT, Personal statement , reference

8.3 Methods

The research paradigm was mixed methods, adopting a pragmatic worldview (Tashakkori and Teddlie, 2010). Mixed methods have been identified as an appropriate design for studying complex medical education topics and are "*increasingly relevant to medical education*" (Maudsley, 2011, Schifferdecker and Reed, 2009)

An explanatory sequential study design entailed conducting a quantitative strand first (MMI- see below) followed by a qualitative strand (focus group interviews- see below). Data from both strands were analysed independently then considered together in the subsequent interpretation of findings. In general the quantitative data evaluated the predictive validity and fairness of the MMI while the qualitative data explored acceptability. Equal weighting was given to both sources of data (QUAN →QUAL) (Creswell and Clark, 2011)   Ethical approval was granted by NUI Galway Research Ethics Committee. Written consent to participate in the study was obtained from students who sat the MMI and additionally from all focus groups participants.

Quantitative Strand

All students enrolled, in First Year Medicine, NUI Galway, September 2012 were invited to participate. Volunteers were entered into a raffle for an iPad. A previously validated MMI was used (Chapter 8 Appendix 1). It consisted of ten, 7

minute stations. Full details of station breakdown, blueprinting of stations, assessor recruitment & training and student and assessor feedback are previously published (Kelly et al., 2014a). Cronbach's Alpha of MMI items was 0.78.

Post MMI, students were provided with individual written feedback highlighting their three best and weakest performing stations and information on how each station mapped to the eight domains of professional practice as per the Irish Medical Council (Medical Council, 2010a).

Demographic data collected were student age, gender, nationality, first language and parental occupation.   EU student selection data comprised school exit exam (predominately Leaving Certificate Examination-LCE) and HPAT-Ireland score. Non-EU student selection data comprised IELTS score, Grade Point Average (GPA) and Interview score. Experimental data consisted of MMI scores which were collected for all participants.

Outcome variables were First Year examination results (Chapter 8-Appendix 2) comprising First Med Score (a continuous variable representing each student's overall performance on the First Year Examinations) and First Med OSCE score (a continuous variable representing performance on a five station OSCE assessing communication and clinical skills).

Quantitative data were entered into PASW (formerly SPSS) (PASW Statistics 18, 2010). We report descriptive statistics, student t tests, chi-square, Pearson's product moment correlation coefficient, Spearman's Rho and regression analysis. Strength of correlations were compared using Cohen's effect size interpretations (small ≥.10, medium ≥.30, large ≥.50) (Cohen, 1992).

Qualitative Strand

Focus groups were utilised as they are effective in accessing a broad range of views; offer participants an opportunity to consider their own views in the context of others and are particularly appropriate for culturally sensitive issues (Mack et al., 2005, Patton, 2002, Pope and Mays, 2006). All student participants and MMI assessors were invited by email to take part in focus group interviews. Four separate homogenous focus groups were conducted in order to capitalise on people's shared experiences; two MMI Assessor focus groups (7 and 6

participants respectively), one EU student (7 participants) and one non-EU student focus group (8 participants) (Pope and Mays, 2006).

Each focus group lasted approximately an hour and was conducted on campus, by an independent experienced moderator. A second researcher took field notes and attended to flow. The topic guide was based on a post MMI evaluation questionnaire administered in the feasibility study (Kelly et al., 2014a) (Chapter 8 Appendix 3). Focus groups were audiotaped and transcribed verbatim. Debriefing took place between the focus group moderator, note taker and one of the authors (MK). Field notes were used to clarify and add contextual details to the transcribed interviews (Mack et al., 2005). Transcripts were independently open coded by three authors (AH, AWM, MK); codes were compared and discussed until agreement was reached. Axial and selective coding took place in an iterative fashion using the constant comparison technique. The final themes were agreed upon by all authors. N-Vivo10 software was used (QSR International, 2012).

Quality and rigour

To help ensure reflexivity a coding diary recorded reflections on how the researchers' own experiences may have shaped the collection, generation and analysis of data (Corbin and Strauss, 2008). Steps were taken throughout the research process to "bracket" any prior assumptions and experiences (Tufford and Newman, 2012). For example researchers were mindful that knowledge of the quantitative findings may influence their subsequent interpretation of focus group data and through careful re-reading of the data they sought to become aware of and manage these influences. Care was also taken to understand "deviant" or "negative" views and to actively look for opinions and thoughts that ran contrary to the researcher's own opinions. With respect to "fair dealing", attention was paid to not over emphasising the views of any one group of participants as if they represented the sole truth (Pope and Mays, 2006).

8.4 RESULTS

8.4.1 Quantitative results

Descriptive statistics

There were 241 eligible students in the year (54 % (n=130) Female; 43% (n=103) Non-EU). Of these 109 (45%) students participated of which 62.4% were female (n=68). See Table 17. Mean age was 19.64 years (SD 1.3; 95% CI 19.4-19.9 years). Of the sample 58.7 % (n=64) were EU origin and 41.3 % (n=45) were Non-EU. Similar proportions of EU and Non-EU students volunteered for the study (44% and 46% respectively). English was the first language of 64.2 % (n=70). Socioeconomic Class (SEC) was based on parental occupation as per Census 2011 guidelines with specific advice on applying these guidelines to Malaysian students sought from a medical academic in Kuala Lumpar (Central Statistics Office (CSO), 2011, Loh, 2014).

Table 18 Participants' Demographics

	Total Sample (100%, n=109)	EU origin* (58.7%, n=64)	Non-EU origin** (41.3%, n=45)
Age	Mean 19.64 yrs (SD 1.32; 95% CI 19.39-19.89)	Mean age = 19.5 yrs (SD 1.26)	Mean age = 19.84 yrs (SD 1.38)
Gender	Female 62.4% (n=68) Male 37.6 % (n=42)	Female 50% (n-32) Male 50% (n=32)	Female 80% (n=36) Male 20% (n=9):
Speaks English as First Language	64.2 % (n=70)	96.9% (n=62)	17.8% (n=8)
Socioeconomic Class			
SEC 1 = Professional workers	33% (n=36)	SEC 1& 2 combined 67% (n=43)	SEC 1& 2 combined 84% (n=38)
SEC 2 = Managerial and technical	41% (n=45)		
SEC 3 = Non-manual	16% (n=17)	SEC 3, 4, 5 combined 27% (n=17)	SEC 3, 4, 5 combined 16% (n=7)
SEC 4 = Skilled Manual	3% (n=3)		
SEC 5 = Semi skilled	4% (n=4)		
Missing data for SEC	4% (n=4)	6% (n=4)	0% (n=0)

*Footnote: * The EU group comprised 61 from Ireland (56% of the overall cohort) and 1 (0.9%) each from Great Britain, Finland and Germany*

***The Non-EU group comprised 37 from Malaysia (33.9%), 5 from Singapore (4.6%); 2 from Canada (1.8%), and 1 (0.9%) from USA*

Table 19 Predictor Variables Mean Scores and Relationship with Gender, Socioeconomic Group and Age

Selection tool	Mean score (SD, 95% CI for the mean) *except where noted^*	Female versus Male scores	SEC 1&2 versus SEC 3,4,5 scores	Correlation with Age *(Pearson's product moment correlation- except where noted^)*
LCE	98.8% (n=56) SD= 0.7; 95% CI 98.6-99.0%	Female 98.7% (n=26)SD 0.7; Male 99% (n=30) SD 0.7 p= 0.59	SEC 1&2 98.8% (n=40) SD 0.7; SEC 3, 4, 5 98.9% (n=12) SD 0.7 p= 0.83	***r=-0.28 (n=56) p= 0.04****
HPAT-Ireland	185.2 (out of max possible 300) (n=63) SD 10.1, 95% CI 182.6-187.7	Females (n=31) 184.6; Males (n=32) 185.8 p=0.64	SEC 1&2 (n=43) 186.2; SEC 3, 4,&5 (n=16) 182.7p=0.19	r=-0.03 (n=64) p=0.79
IELTS	7.2 (n= 38) SD 0.5; 95% CI 7.0-7.3	Females (n=31) 7.2; Males (n= 7) 7.1, p 0.98	SEC 1&2 (n=81) 7.2; SEC 3, 4, 5, (n=24) 7.1, p =0.97	***r=0.35; (n= 38) p=0.04****
Traditional Interview	76.5% (n=28) SD 10.4 95% CI 72.5-80.5%	Females (n=23) 76.7%; Male (n=5) 75.4%; p=0.83	SEC 1& 2 (n=23) 77.3%; SEC 3, 4, 5 (n=5) 72.8%, p=0.31	r=0.13 (n=28) p=0.5
GPA	^Median out of max possible score of 4 was 3.9 (n=45) min= 2.8, max=4; interquartile range 0.3	Females (n=36) 3.9, SD 0.3, Min 3 , Max 4_ Males(n=9) 3.9, SD 0.3, Min 2.8, Max 4_ Mann Whitney U p= 0.81	SEC 1&2 (n=38) median GPA 3.9; SEC 3, 4, 5 (n=7), median GPA 3.9, Mann Whitney U p =0.96.	Spearman's rho -0.05, p=0.73
MMI	67.1% (n=109) SD 9.7, 95% CI 65.2-68.9	Females (n=68) 66.9%; Males (n=41) 67.3% p=0.83	SEC 1&2 (n=81) 67.1%; SEC 3, 4, 5 (n=24) 66%, p=0.66	r=0.15 (n=109) p=0.12

Footnote *Significant results highlighted in bold italics*

Between Group comparisons

Between groups comparisons revealed no statistical difference between mean age of the EU and Non-EU students (T-test (two tailed) p= 0.18) nor the proportion of students in SEC 1 & 2 and SEC 3, 4, 5 in the different groups (Chi Square p=0.19 (df=1)). There was a significantly higher proportion of females and students who did not have English as a first language in the Non-EU group (Non-EU group Females 80% (n=36): EU students Females 50% (n=32) Chi Square p=0.003 (df=1): Non-EU group with English as a First Language 17.8% (n=8): EU group with English as First Language 96.9% (n=62) Chi Square p< 0.01 (df=1)). These differences reflected the First Year Medical class norms.

Selection and Experimental data

Table 18 presents the selection tool scores, including MMI, and indicates the relationship between candidates' scores, gender, SEC and age.   Mean MMI score 67.1% (n=109) SD 9.7, 95% CI 65.2-68.9.   EU students scored significantly higher on MMI than Non-EU students (EU (n=64) mean 72.0%; Non-EU (n=45) mean 60.1%; p<0.001; Difference in mean 11.9%, 95% CI 8.8-14.9). Likewise students with English as a first language scored significantly higher on MMI (First Language English (n= 70) mean MMI 71.4%; First language not English (n=39) mean MMI 59.2%; p<0.001; Difference in mean 12.2%; 95% CI 9.0-15.4).

Outcome variables

Table 19 presents students' scores on the overall First Medical Year Assessments and the OSCE. Cronbach Alpha of OSCE was 0.70. There was no significant difference between EU and Non-EU students in terms of mean First Med Overall Score however students who had English as a first language out performed those who did not (Difference in mean First Med Overall Score 3.4%; 95% CI 0.4- 6.3, p=0.03). With respect to OSCE scores there was no significant difference in mean scores between EU and Non-EU students nor between those with English as a First Language or not.

Correlations

Table 20 shows the correlations between selection data and the outcome variables for the whole sample. Significant positive correlations emerged between MMI and IELTS (large). When the sample were split according to origin

(EU and Non-EU separately), a further significant correlation emerged between EU students' MMI results and OSCE results (r=0.27, n=64, p=0.03) (small).

Regression analysis

A linear regression model was fitted separately for EU and Non EU students, initially with all predictors (full model) then used variable selection to identify potentially useful predictors. For EU students LCE was the only useful predictor (R^2 27%; p<0.0005), while for Non-EU students GPA was the only significant predictor (R^2 53%; p<0.0005).

Table 20 Outcome Variables and Relationship with Gender, SEC, Age, EU / Non-EU Background and English as First Language

Outcome Variable	Mean Score, SD, 95% CI of mean	Gender	SEC	Age	EU versus Non-EU	English First Language Yes/No
First Med Overall Score	65.5% (n=109) SD 8.1, 95% CI 63.9-67	Female (n=68) 65.46%; Male (n=41)65.48%, p=0.99	SEC 1&2 (n=81) 65.4%; SEC 3, 4, 5 (n=24) 65.8%, p=0.82	r=-0.02; p=0.84	EU students (n=64) 66.3% SD 8.4; Non-EU students (n=45) 64.2%, SD 7.5, p=0.17	*First Language English (n=70) 66.7% SD 8.5, First Language not English (n=39) 63.3% SD 6.9, p=0.03**
OSCE Chronbach alpha 0.70	81.7% (n=109) SD 5.1, 95% CI 80.7-82.6	*Females (n=68) 82.8% SD 4.2; Males (n=41) 79.9% SD 5.8 p=0.007***	SEC 1&2 (n=81) 81.9% SD 4.8; SEC 3, 4, 5 (n=24) 81.3% SD 5.5, p=0.64	r=-0.15; p=0.03	EU students (n=64) 81.5% SD 5.2; Non-EU student (n=45) 82% SD 4.9, p=0.61	First Language English (n=70) 81.6% SD 5, First Language not English (n=39) 81.8% SD 5.2, p=0.83

Footnote: Significant results highlighted in bold italic

Table 21 Correlations Between Selection and Outcome Variables

	MMI	OSCE	First Med-Score	LCE	HPAT	IELTS	GPA	Trad-I
MMI	_	.09	.11	-.07	.21	**.50****	-.23 (rho)	.27
OSCE	.09	_	**.33***	.19	**-.25***	.02	-.10 (rho)	-.16
First Med Score	.11	**.33****	_	**.52****	**-.27* (rho -.36**)**	**.44****	.10(rho)	-.21
LCE	-.07	.19	**.52****	_	**-.28* (rho -.37**)**	n/a ^	n/a ^	n/a ^
HPAT	.21	**-.25***	**-.27* (rho .36**)**	**-.28* (rho -.37**)**	_	n/a ^	n/a ^	n/a ^
IELTS	**.50****	.02	**.44****	n/a ^	n/a ^	_	**.42*(rho)**	.02
GPA	-.23 (rho)	-.13 (rho)	.10 (rho)	n/a ^	n/a ^	**.42*(rho)**	_	-.10 (rho)
Trad-I	.27	-.16	-.21	n/a ^	n/a ^	.02	-.10	_

Footnotes All correlations calculated using both parametric and non-parametric tests. Pearson's product moment correlation value (r) listed except where there were differences between the findings and in this case both Pearson's r and Spearman's rho reported. GPA was not normally distributed hence Spearman's Rho is used throughout for this variable. Significant results highlighted in ***bold italics.*** ** Correlation is significant at the 0.01 level (2 tailed) * Correlation is significant at the 0.05 level (2 tailed). n/a ^ Some correlations are not appropriate as students from the EU and Non-EU streams sat different selection tests

8.4.2 Qualitative results

Two overarching themes emerged. Authenticity describes participants' views on the trustworthiness of MMI and factors that impacted on this in both a positive and negative way. Cultural Awareness captures participants' understanding of how cultural values, beliefs and perceptions influenced both candidate and assessor performance at MMI. Quotes are identified as follows: Non-EU=Non-EU Student Focus Group; EU = EU Student Focus Group; MMI A1= MMI Assessor Focus Group 1; MMI A2= MMI Assessor Focus Group 2

Theme 1– Authenticity

1.1 Deeper understanding

Participants believed that selection " *is really high stakes. You either get into the career of your choice or you don't*" (MMI A2). EU and Non-EU student reaction was overwhelmingly positive. They viewed MMI as a very authentic assessment of high value which was "*more thorough*" (Non-EUS) than alternative tools. MMI offered "*insight into how you cope and handle things that wouldn't be apparent in a regular interview*" (Non-EUS) and provided "*more of a chance to show who you were*" (EUS). As one candidate stated "*I kind of felt that I was really being forced to think on my feet*" (EUS). There was broad consensus amongst students that MMI was a welcome addition to selection.

Assessors viewed the main benefit of MMI was the opportunity to "*get a feel for the person themselves and how that could translate into the [medical] course*" (MMI A1). It provided "*a deeper understanding*" of candidates (MMI A1). One assessor viewed the purpose of MMI to "*..make sure people who are totally not suitable for medicine don't go into it rather than you know ranking people who are good*" (MMI A2). In comparison with other selection tools, assessors felt that MMI was "*better than the HPAT-[Ireland]*" (MMI A2) and "*certainly better than just academic performance*" (MMI A2).

1.2 Relevance to clinical practice

A big advantage was that MMI station content was seen as " very *relevant*" (EU) to "*the skills that we need to have when we are doctors*" (Non-EU) and in this way students felt it offered authentic insight into "*what you're going to be doing*

ten years down the line" (EU).   An unexpected advantage of this was that students felt primed towards the skills that were important for them to develop in their undergraduate career: " *I feel that MMI actually help[ed] us in developing those skills*" (Non-EU). A small number of assessors also viewed this *"formative"* (MMI A2) role important especially with respect to students' *"communication skills"* (MMI A2) and *"professionalism"* (MMI A1).

Assessors felt that preparation for the MMI stations ensured that students " *..really had a better feel for what …. the day to day working as a medical doctor is. It actually forces them to put themselves in those situations*" (MMI A2).

1.3 Reservations and recommendations

Assessors expressed more reservations than students. A number of assessors (approximately 3) were concerned that MMI felt "*like a bit of a performance*" (MMI A1). They worried that students "*were trying to work out what*" was expected from them and "*give you exactly that*" (MMI A1). Coaching was seen as a threat that could "*undermine the whole process*" (MMI A2) and reduce its ability to be "*discriminatory "(*MMI A1). Commercial coaching could have a negative effect, because of the associated cost: "*it's just you end up with a lot of people from the highest socio-economic classes*" (MMI A2). Students suggested that live MMI stations be set up on medical school "*open days"*(EU) where applicants could practice and that sample MMI stations be available "*online so that people could look them up if they couldn't afford to go to a [coaching institution*]" (EU).

Some assessors admitted "*responding more to the candidates who were confident humorous and warm*" (MMI A1). This led to a concern that "*students who are more nervous…. a little bit shyer*" (MMI A1) or "*less empathic*" (MMI A2) may be "*negatively discriminated against*" (MMI A1) by MMI. Some students also voiced this concern and the possible impact on selection

"if you have very good communication skills you can do very well on the MMI.. So it's almost singling out that group of people" (EU)

It was suggested to "*broaden the range of skills or attributes*" (MMI A1) assessed to counter the perception that MMI over emphasised "*communication skills and empathy*" (MMI A1). For example by including stations that tested

"*efficiency and problem solving*" (MMI A1), that could identify applicants who were "*dexterous, skilled and imaginative*" (MMI A2) and that did not penalise applicants who were "*a bit brisker in their communication styles*" (MMI A1).

Other reservations included recognition that MMI was "*time consuming … labour intensive*" (MMI A1) and "*very costly*" (MMI A2) to run and concern about using MMI for entry from second level education

"*These are very young students. We are bringing them into college to teach them some of these skills….. Is this the right time to be assessing it? I'm not sure*" (MMI A2).

Theme 2– Cultural Awareness

2.1 Culture, Attitudes and Station Content

Assessors and students observed that "*the cultural context was an issue through a number of the stations*" (MMI A1). Culture and attitudes were particularly relevant in stations where societal matters which were "*subtle and culture specific*" (MMI A1) such as when a student with an alcohol problem or organ donation were discussed. (See Chapter 8 Appendix 1).

"*I know the, Islam[ic] students in our class in particular would have been very taken aback by the alcohol station …*" (EU).

International students clearly recognised the challenge culture posed to their communication skills:

"*I couldn't put myself in her shoes …..it is so difficult because our cultures are different so I can't put empathy there*" (Non-EU- reflecting on the alcohol station).

Conversely one international student saw merit in challenging applicants' cultural views

"*drinking is not in our culture, so I think that situation is a must in selection tools because we can see, …..how they cope with the issue that is outside of their own life*" (Non-EU).

One solution was to use stations based on "*behavioural things that cut across cultures*" (MMI A2) such as "*cheating… for breaches of professionalism*" (MMI A2) which may be more accessible to all applicants. An alternative suggestion was "*…to bring in some diverse cultural issues into it, like witchcraft… things that are not the typical things you'd meet in Ireland*" (MMI A2). These stations would then prove equally challenging to host country and international applicants.

Another common suggestion for improvement was to increase the standard and scope of the information that was made available to applicants in advance of the MMI. Knowledge "*of the kinds of station [to expect]*" (Non-EU) would allow students to "*prepare beforehand*" (Non-EU).

2.2 Culture and Assessor Subjectivity

Assessors were acutely conscious of the link between culture and assessor subjectivity:

"*I'm assessing from my viewpoint, cultural viewpoint which can be very, very different to others and I would think that I'm probably not a very* reliable [assessor]" (MMI A2).

Culture impacted significantly on how assessors viewed candidates' interpersonal communication. This was evident both in the stations with a role player:

"*Different cultures comfort people in different ways and so my perspective of someone comforting a friend, it's completely different to some international students who sat beside them and had their arm around them, which is culturally acceptable in their cultures..*" (MMI A2).

And the one to one interview type stations:

"[in the] *one to one interviews…..some people discuss things with humour but humour doesn't necessarily translate…. so then it just came off as weird… you could lose all of your marks.*" (MMI A2).

Detailed assessor training on cultural awareness was seen as essential;

"…*if I'm not informed of a different person, different culture, cultural practice, how can I assess them reliably?*" (MMIA2).

So too was recruiting *"multiple examiners from different cultures"* (MMIA2). In addition assessors strongly recommended that MMI *"pilots"*(MMI A2) should be part of normal advance MMI preparations and should include students from a variety of cultural backgrounds to ensure *"a representative core…of the core sitting the exam*" (MMI A2).

2.3 English Language Proficiency

International students recognised the important influence English language proficiency had on performance at MMI "*if you didn't understand the questions how are you going to perform very well?*"(Non-EU).   They also observed that "*not having it[English] as your first language …could make you more nervous*" (Non-EU) which again could impact negatively on performance.

Assessors too considered English language proficiency "*a very big barrier*" (MMI A1). However some EU students felt that the MMI was "*a useful screening tool for the English language*" (EU) which was important because

"*….if they can't understand the MMI, it's going to be very difficult for them to understand lectures and therefore they're not going to do well in Ireland*" (EU).

A contributing factor was that "*the time limitation is really short*" (Non-EU) and students recommended lengthening the allotted time.

8.5 Discussion

The reputation and use of MMI as a selection tool for medicine is growing (Pau et al., 2013). At the same time migration patterns of medical students internationally continue to rise. This study aimed to establish the fairness, predictive validity and stakeholder acceptability of MMI in an internationally diverse student population.

In this sample MMI performed as well, or better than, the other selection tools in terms of fairness with respect to age, gender and socioeconomic class (see Table18). We found no evidence that performance on any of the selection tools was influenced by the SEC of the candidates although this has been

demonstrated elsewhere (Broquet and Rockey, 2004, McManus and Wakeford, 2014). It is possible this may be a type 2 error due to our small sample size.

Widening diversity to medicine has become an important consideration in the choice and on-going use of selection tools (Cleland et al., 2011, Higher Education Authority, 2008). *Equal opportunity* is an important aspect of distributive justice. As MMI becomes more widely used it is likely that commercial coaching will become more prevalent, which may negatively impact lower socioeconomic applicants; a concern raised by our assessors. As suggested in the focus groups medical schools could mitigate this by providing open-source online access to preparatory materials, including mock stations and marking grids, links to resources outlining the relevance of MMI stations to clinical practice and guidelines for professional behaviours and standards in medicine. Ensuring that essential cultural and linguistic information was included in this resource would help extend equal opportunity to international applicants. These and other recommendations from the study have been summarised in Table 22.

The predictive validity of MMI with respect to First Med assessments was weaker in this study than reported predictions of medical course assessments elsewhere but somewhat better in terms of prediction of OSCE scores of EU students (Husbands and Dowell, 2013). Why MMI was more predictive of EU rather than Non-EU students' OSCEs is unclear. The First Medical Year in NUI Galway is largely devoted to the pre-clinical sciences and knowledge based assessments; these observed predictive validity patterns may change as students progress through the course and spend more time in the clinical domain.

Table 22 Best Practice Recommendations

1.	**Consider what domains to test and blueprint these against the relevant medical school and nationally agreed learning outcomes, regulatory standards and on the job requirements**
2.	Ensure adequate diversity of domains tested to avoid over reliance on any one skill set
3.	Consider cultural issues in the design and development of stations. Opt for culturally neutral material or adequately diverse cultural issues to avoid giving an advantage to any one group of candidates.
4.	Pilot stations with candidates and assessors from a range of cultures, where possible mapped to the cultural backgrounds of the relevant applicant pool
5.	Provide adequate cultural awareness training for assessors and recruit assessors from a range of cultural backgrounds
6.	Use clear unambiguous language, avoiding colloquialisms, for candidate and assessor instructions and role players' script
7.	Ensure station duration provides adequate time for candidates who are being assessed in a language other than their first language
8.	Provide free preparatory information to applicants in advance via a variety of media – e.g. Medical School Open days, online and printed materials. Material should include sample stations, a description of the MMI process, justification for the range of domains tested in terms of job relatedness, links to professional standards and medical school learning outcomes.
9.	Regularly audit applicant and successful candidates for demographics including age, gender, socioeconomic group and cultural backgrounds to monitor for fairness.
10.	Monitor for evidence of predictive validity on an ongoing basis.
11.	Provide adequate supports and formative feedback to international students throughout their training
12.	Draft globally agreed minimum standards for selection processes

Recently there has been significant controversy over evidence that IMGs perform less well than home graduates in several higher professional training examinations in the UK, USA and Canada (Boulet et al., 2006, Esmail and Roberts,

2013, McManus and Wakeford, 2014, Peile, 2014, Tiffin et al., 2014b). In light of this controversy a challenging finding from this study was that Non-EU students' mean MMI score was significantly below that of EU students. The same pattern emerged for students without English as a first language. In terms of *equity of outcome* these data require detailed consideration. Arguably selection is the most significant exam in one's medical career hence understanding the mediating and moderating influences of these finding is important. It assumes even greater importance should both home and international applicants be competing for the same places.

Similar proportions of EU and Non-EU students volunteered to participate in this study and there is no evidence to suggest that levels of preparation or motivation were lower amongst international students. The fact that EU and Non-EU students did not differ in First Med Overall Score, or OSCE score implies that levels of ability in both groups was similar, at least by the end of Year 1 and rules out differences in cognitive ability as a likely explanation. The quantitative and qualitative data however support the centrality of culture and language to MMI performance. Comparing MMI with currently used selection tools highlighted that the strongest association, a significant large strength correlation, lay between MMI and English language ability, with higher IELTS scores associated with enhanced MMI scores. Ireland uses a minimum IELTS subsection and overall score of 6.5 for entry which is a point lower than many UK schools and may go some way to explain the discrepancy in MMI scores (Peile, 2014). Tiffin et al (2014b) have established that IELTS scores significantly predict IMGs' clinical competence as measured by performance on the Annual Review of Competence Progression: each IELTS point scored above 7 increased the odds of achieving a more satisfactory appraisal by 69%.

English language proficiency was seen as a very significant barrier by both assessors and students. Apart from simply raising the eligible IELTS entry score, which may have the undesirable effect of disqualifying some potentially good applicants, a number of measures could be taken to improve the *characteristics of the test.* Further work is required to establish if these measures would have a beneficial effect in reducing the language hurdle. If shown to be helpful then these measures may enhance the procedural and ultimately distributive justice of

the MMI. The EU student focus group recognised that English language proficiency is also associated with ability to understand lectures and perform as a medical student. This was supported by quantitative data showing that First Med Overall mean score for students with English as a first language was over 3% higher than those without. This highlights the duty of responsibility medical schools have to provide adequate monitoring, feedback, support and access to additional training in conversational and medical English for non-native speakers.

Issues of language were closely aligned with culture and assessor subjectivity. Assessors were particularly concerned that the cultural context of some of the MMI stations, or their marking of them, may be unfair and potentially diminish *equity of outcome* for international students. They suggested additional training in cultural sensitivity. Cultural sensitivity means being aware that cultural differences and similarities exist and have an effect on values, learning and behaviour (Stafford et al., 1997). Cultural training is the norm for examiners on many postgraduate training programmes (McManus et al., 2013d). In addition to training in the design of culturally sensitive stations assessors recommended purposefully recruiting assessors from different backgrounds and involving international students in the piloting of MMI stations. All of these measures increase the resources and cost of MMI but in terms of best practice guidelines for MMI developers they are worthy of due consideration.

This study focused on the selection of students from an international background. However the issues raised with respect to culture are broadly pertinent to selection as a whole. Approximately 30% of "home" UK medical students (those with UK nationality) are from ethnic minorities and applicants from these backgrounds may be equally as challenged by some of the culturally specific stations (McManus et al., 2008). As the use of MMI becomes main stream it is important to ensure that ongoing audit and evaluation of the performance of both international applicants and those from different cultural and ethnic minority backgrounds is undertaken and made publically available.

In our study stakeholders' perceptions of the procedural justice of MMI were enhanced by its relevance to future clinical studies and the adequate opportunity it provided to candidates to demonstrate their ability. Non-EU students

highlighted that the MMI experience primed them for the skills they would need to acquire in their undergraduate education and likewise assessors saw an important role for MMI in giving formative feedback. It could be argued that there is a responsibility on medical schools to ensure that information outlining successful candidates' strengths and weaknesses is fed back to them to guide their learning. This would be particularly relevant for issues of professionalism, culture, attitude and language. Says et al have piloted the use of MMI in Saudi Arabia and, although MMI is not yet a formal part of the selection process, they use it to identify both outstanding and below average performing students targeting the latter for additional supportive workshops (El Says et al., 2013). This is a model that could be applied elsewhere and likely to be welcomed by students.

Study limitations include that this is a single study, situated in one medical school with a small sample and one year of follow-up.   It is possible that our findings were influenced by other confounding variables within the MMI, or systematic examiner bias. Students were already selected to medicine and along with assessors they volunteered to take part. This self-selected group may differ in meaningful and underdetermined ways to their counterparts. A further limitation is that this MMI was experimental hence the usual pressures, motivations and preparation undertaken by candidates in advance may not have applied in this circumstance. By contrast the main strength of this study is that the methodology provided for a rich and deep understanding of both the numerical data and facts relating to MMI as well as the meaning it had for both assessors and students.

Further considerations - Recognising the growing patterns of migration in medical school applicants and the heterogeneity of selection practices internationally perhaps the time has come to draft globally agreed minimum standards of selection practice similar to those for medical education programmes (World Federation for Medical Educators (WFME), 2012). International medical students are an important source of revenue for Medical Schools. They are however not a homogenous group. This study highlights the duty of Medical Schools in terms of social responsibility, to ensure that international students are provided with adequate supports, particularly with respect to culture and

language, tailored to their individual needs.   In these circumstances MMI could provide helpful formative feedback.

8.6 Conclusions

In conclusion MMI has proved a welcome addition to assessment armamentarium. This study found that MMI demonstrated good job relatedness and acceptability, particularly amongst candidates. Understanding the mediating and moderating influences of differences in performance of international candidates is essential to ensure that this selection tool complies with the metrics of good assessment practice and principles of both distributive and procedural justice for all applicants, irrespective of nationality and cultural background.

Chapter 9: Drawing it all Together

9.1 Thesis Strengths and Limitations

Strengths

The strengths and limitations of this thesis are considered at the start of this chapter because they have important implications for the integration of findings and drawing of conclusions.

The greatest strength of this thesis is that it is of a mixed methodology. This provides for a thorough exploration of the issues relating to both predictive validity and stakeholder acceptability. This led to a richer understanding of how both relate to the utility of HPAT-Ireland and MMI. Both measures are critical to the political validity of these selection tools (Cleland et al., 2012, Patterson et al., 2011).

This thesis commences with a detailed literature review on aptitude tests and multiple mini interview. It provides the first published evidence of the predictive validity of HPAT-Ireland. In doing so it responds to a call for further predictive validation studies of aptitude tests highlighted in a seminal review identifying best practice in selection (Cleland et al., 2012).

It addresses a gap in the literature by means of a systematic review of stakeholders' views of selection to medicine. The review highlighted a dearth of well-designed qualitative studies of stakeholder views. This thesis responds with two studies which use qualitative means to explore in detail the views of medical students, doctors and MMI assessors.

It explores the views of the medical profession, an overlooked stakeholder group, on HPAT-Ireland. A key strength of this study was that the participants first sat HPAT-Ireland hence they had direct knowledge of the tool to draw on. This was particularly important to the exploration of job relatedness of the tool.

It examines issues of fairness and predictive validity of MMI in an international student population. This addresses a demand for more research to demonstrate the transferability of MMI outside of a North American context and responds to a recommendation in a key recent review of MMI for research on stakeholder acceptability in "*different cultural contexts*" (Husbands and Dowell, 2013, Pau et

al., 2013). Additionally it provides stakeholder recommendations for improving the utility of this tool for an internationally diverse applicant pool.

The feasibility and outline cost of running a Multiple Mini Interview are established for the first time in an Irish context. Rosenfeld et al (2008) argue that it is necessary to determine the costs of MMI, in different contexts, to ensure that they can be feasibly implemented.

Limitations

Many of the thesis limitations have been already discussed in Chapters 5 to 8. This section discusses some additional important limitations that relate to integrating findings from the different strands of this thesis.

Issues Relating to Sample and Duration of Follow Up

Both Studies 1 and 4 report on predictive validity. However, there are important differences between the studies that impact on interpretation of findings, especially with respect to the predictive validity of HPAT-Ireland.

Study 1 was conducted on a sample from UCC and NUI Galway Medical Schools, whereas the sample in Study 4 was exclusively from NUI Galway. A sample drawn from multiple sites is likely to be more generalisable, as it can balance out context specific issues such as local teaching and learning styles, curricular issues and examination differences.

Study 1 was based on all successful applicants to UCC and NUI Galway. By contrast Study 4 was based on a volunteer sample. Volunteer samples in medical education research have been found to differ from their peers by generally performing better on examinations during and after medical school (Callahan et al., 2007). Using a volunteer sample can result in additional range restriction and underestimation of predictive validity (Hunter et al., 2006).

The size of the sample is also important. To check for limitations arising from inadequate sample size, post hoc power calculations were performed (Faul et al., 2007, Mayr et al., 2007). These confirmed that Study 1 had a sample size sufficient to achieve 86% power to detect a correlation effect size of 0.20, at the

0.05 probability level. In Study 4 the sample size was sufficient to achieve 68% power to detect a correlation effect size of 0.20, at the 0.05 probability.

Hence, when comparing predictive validity findings from Study 1 and 4, it should be borne in mind that Study 4 is limited by relying on a volunteer sample, from one medical school with one year of follow up and due to limited sample size has the potential for a Type II error.

A limitation of Study 1 was the inclusion of Foundation Year students. The rationale was to capture data on all students who enrolled in medicine in the two medical schools in the first year of use of HPAT-Ireland, and thus maximise the sample size. In NUI Galway, depending on scores in Leaving Certificate science subjects, students enrol in either First Year or Foundation Year. Although between groups comparisons did not show significant differences, it is possible that these groups varied in undocumented ways. The communication skills training and OSCE was largely identical for First Years and Foundation Year students. Importantly however, Foundation Year students were not included in the second year of follow up. They would have had the benefit of communication skills teaching in their First Year, which could have confounded the results and they were therefore excluded from analysis in Year 2.

Reliability Problems in the Predictor Variables

Sackett et al (2008) describe how unreliability in either the predictor or predicted variables leads to an underestimation of correlation. Reliability measures for the selection tools used in this thesis were not available. MMI was an exception. Cronbach alpha for the MMI was 0.78 which is within the typical range for in-course assessments (0.60-0.85), but would be considered slightly low for high stakes testing where alpha co-efficient of 0.8 or above are preferred (Linn and Gronlund, 2000, Streiner and Norman, 2008). Although exact figures were not available, it is likely that the reliability of IELTS was also good, given that previous records have put it at approximately 0.80. The lack of data regarding the reliability of the other selection tools in the thesis lessens the confidence we can place in the predictive validity results. This emphasises the need for transparency on behalf of companies/ agents responsible for providing selection tools for medicine with respect to making these data publicly available.

Reliability Problems in the Predicted Variables

The mean Cronbach alpha for Study 1 OSCE was 0.63 and in Study 4 it was 0.70, both of which lie within the typical range for in-course assessments (Linn and Gronlund, 2000, Streiner and Norman 2008). There was no measure of reliability for the overall weighted average First Year assessment score in Study 4, as it was based on a combination of marks from ten individual modules, and these data were not available. However, these assessments are all subject to annual external examiner review, while the entire medical programme undergoes regular review and accreditation by the Medical Council. These provide some degree of assurance that the reliability of these assessments is likely to have been reasonable.

Restriction of range

In the context of selection, range restriction refers to the fact that successful applicants necessarily have scored higher on the selection tool than those who were not selected. The effect of range restriction is to reduce the subsequent correlation between the predictor and predicted variable (Cleland et al., 2012, McManus et al., 2013a, Sackett et al., 2008, Sackett and Yang, 2000).

Although the practice is changing, there are numerous examples of peer reviewed published papers, in leading medical education journals which have not corrected for range restriction as commented on by Coates (2008). Some view the correction an unnecessary complication, particularly in cases where no minimum threshold score is set for the selection tool, and selection scores are normalised based on the entire applicant pool(Shulruf, 2015). Others argue that it is best to report conservative estimates (Wright, 2012). By contrast the field of psychology has over seventy five years of research and experience demonstrating the importance of correcting for range restriction in order to provide a truer estimate of the predictive validity of selection tools (Schmidt et al., 1992).

McManus et al (2013a) recommend applying the Hunter, Schmidt, Le formula (HSL) to correct for instances of indirect range restriction (Hunter et al., 2006). Medical students who participated in this thesis were selected based on a composite score, derived from the weighted performance on the various selection tools. Offers of places were made via a top down approach, according to the

applicant rank order list. This is an example of indirect range restriction. The HSL formula has an added benefit of adjusting for measurement errors in both the predictor and predicted variables. Amongst other variables, this formula requires the reliability coefficient and standard deviation of the selection tool in the applicant pool. Neither of these variables were available to me, for any of the selection tools reported on in this thesis.

Therefore, the correlations reported in this thesis between selection tools and outcome measures, should be treated as underestimates. In Study 4, the issue of range restriction will not apply to the predictive validity of MMI. This is because the entire sample, who sat the test was enrolled in the study and scores were not used for selection purposes.

Best practice in predictive validation studies would be to present both the restricted and unrestricted correlation coefficients (American Educational Research Association et al., 1999). Again this emphasises the important role that companies and agents who administer selection tools play in facilitating access to these data to allow for accurate estimations of predictive validity.

Fairness and predictive bias of MMI

It is a limitation of Study 4 that, when examining issues of fairness, the possibility of predictive bias with respect to MMI was not explored in depth in the published paper. This shortcoming is addressed here. There are a number of ways to test for predictive bias. One of the principal ways is to look for evidence of differential prediction (Kyei-Blankson, 2005). It occurs when the best prediction equations and/or the standard errors of estimate are significantly different for different groups of applicants (Young and Kobrin, 2001). Over-prediction is said to occur when the student does not do as well on in-course assessments as predicted by the selection tool. Under-prediction is the opposite of this. It is a more serious concern because it runs the risk of applicants being de-selected based on predictions of their future performance that underestimate their likely scores (Kyei-Blankson, 2005).

The data in Study 4 were re-examined for evidence of differential prediction. Chapter 9 Appendix (1) explains the methodology and presents the results (Table 29, Figures 10 & 11). In summary when the data from Study 4 were further

analysed there was no evidence of under-prediction of average weighted First Year Medical Scores for Non-EU students or students without English as a first language, despite findings of a significant difference in mean MMI scores. The findings overall do not present a convincing pattern of consistent under-prediction for any one group of students, based on gender, socioeconomic group, language or background. It is important though to note that the sample size of the subgroups is small and there is a possibility of a type II error.

Thesis by publication

It is both a strength and a limitation of the thesis that it is by publication. The strength is that this approach ensures that research findings are disseminated during the four years as opposed to the end. In fields such as selection with high levels of research activity, study findings have a finite period when they will be of interest to a journal, hence timely publication is essential. The benefit of having undergone regular peer review brought insight and provided feedback which helped strengthen future publications. It also ensured that I had material to present at conferences. This introduced me to a community of experts in the field. The ensuing collaborations and exchange of ideas again strengthened the thesis.

The limitation of this approach is that it necessitates publishing from very early on in the thesis journey, ahead of the wealth of subject matter expertise that one develops throughout the course of a doctorate. An example of this relates to interpreting the strength of correlation between selection tools and future student performance. Whereas Cohen's effect size is the most widely utilised measure reported in medical selection publications, it doesn't convey the value that even small effect sizes can bring to selection processes (Cohen, 1992). A more sensitive tool may be the U.S. Department of Labour, Employment and Training Administration's (1999) guidelines for interpreting correlation coefficients in predictive validity studies. These state correlations less than 0.11 are unlikely to be useful, between 0.11–0.20 usefulness is dependent on circumstances, between 0.21–0.35 are likely to be useful and those over 0.35 are considered very beneficial.

9.2 Summary and Combination of Key Findings

Predictive Validity of HPAT-Ireland - (Research Aim 1: Objective 1)

Study 1 found that medical students' total HPAT-Ireland scores significantly correlated with OSCE communication skills score, OSCE clinical scores and OSCE total scores. In addition HPAT-Ireland Section 2 scores correlated significantly with OSCE communication skills scores and total OSCE scores. These moderate strength correlations were evident in Year 2 but not Year 1. On regression total HPAT-Ireland scores were found to independently predict total OSCE score, accounting for 7% of the variance in students' performance. HPAT-Ireland and LCE combined score significantly, moderately correlated with knowledge of communication/clinical skills as measured in the First Medical MCQ.

Study 4 found that HPAT-Ireland significantly negatively correlated with First Medical OSCE scores and with First Medical Year Assessment total weighted average score. The strength of this correlation was also moderate. However on regression analysis, HPAT-Ireland was not a significant independent predictor of either outcome measure.

Stakeholder Perspectives of HPAT Ireland (Research Aim 1: Objective 2).

Qualified doctors from a range of clinical disciplines perceived HPAT-Ireland to have good levels of job relatedness (Study 2). Sections 1 and 2, examining logical reasoning and interpersonal understanding respectively, were both considered to resonate well with the demands of clinical practice. In particular Section 2 was strongly endorsed and there were calls to expand this section to incorporate fuller examination of communication skills. However, Section 3, non-verbal reasoning, was widely criticised for lacking clinical relevance.

Impact on diversity was a strong consideration. Doctors felt that differences in HPAT-Ireland subsection performances according to gender would balance out on the overall test. But they expressed concerns re HPAT-Ireland's susceptibility to coaching and the arising potential for socioeconomic bias.

HPAT-Ireland was considered acceptable and useful as a tool to help distinguish between excellent applicants, but participants had little faith in its ability to predict future performance as a doctor. The minority of participants,

who found it unacceptable, were influenced principally by perceptions of limited job relatedness and potential for negative impact on student diversity.

Feasibility of Multiple Mini Interview (Research Aim 2: Objective 3).

The feasibility of running a MMI in an Irish setting was established in Study 3. It demonstrated that the familiar experience of examining medical students using an OSCE format, including the use of simulated patients and computer software for marking stations, served well to facilitate the smooth running of a MMI. The estimated cost of hosting (as opposed to developing) the MMI was estimated at €145, per student, which is comparable with HPAT-Ireland.

Predictive Validity and Fairness of MMI (Research Aim 2: Objective 4)

Study 4 found that MMI did not significantly predict average weighted First Medical Year Assessment scores or OSCE scores for the study sample as a whole. There was no evidence of difference in students' performance on MMI based on gender, socio-economic groups or age. However, the mean scores of students from outside of the EU, and those without English as a first language, were significantly lower than their counterparts (differences in mean 12%). However, subsequent further exploration of the data, presented in Chapter 9 Appendix 2, did not find any consistent pattern of MMI predictive bias.

Stakeholders' views of MMI (Research Aim 2 Objective 5)

Quantitative feedback from Study 3 indicated that 90% of students and 75% of assessors felt that the content of MMI was relevant to medicine. Over half of students and 71% of assessors agreed that it would be a suitable and useful addition to the selection of medical students. Assessors were positive about the opportunity MMI offered to objectively test non-cognitive skills. However, they expressed reservations regarding the potential for quieter or international students to underperform. Another perceived barrier to using MMI was the logistic and resource requirements.

Qualitative feedback from Study 4 revealed that MMI was considered, by both assessors and students, a highly authentic assessment with immediate relevance to clinical practice. However, the content of some MMI stations in particular was viewed as culturally specific and this, along with limited English language proficiency, were seen to disadvantage international students. Assessors were

acutely aware of the link between their own culture and their reliability as interviewers. Assessors and candidates made a number of key recommendations for improving the cultural sensitivity of MMI including careful choice of station content so as not to offer advantage to any one group of applicants, cultural awareness training for assessors, and the importance of piloting MMI with a culturally diverse panel

Other reservations regarding MMI included concerns about workload, possible susceptibility to coaching, potential for slightly more reserved students to be discriminated against and an overemphasis of communication skills to the exclusion of other important non-academic skills.

9.3 Integration and Synthesis of Key Findings

The findings from each study have been discussed individually in the published papers (Chapters 5-8). The principal findings from all four studies, as they relate to each other, are drawn together here, and discussed with reference to organisational justice theories. Issues appraised in detail in the discussion sections of the published papers are not re-considered here.

HPAT-Ireland

Findings from Studies 1 and 4, with respect to the predictive validity of HPAT-Ireland for Year 1 OSCEs are somewhat conflicting. This discrepancy is not entirely unusual and similar differences between cohorts have been reported by others (Groves et al., 2003, Groves et al., 2007, Husbands and Dowell, 2013). This mirrors the contradictory findings reported in general regarding predictive validation of aptitude tests (See Chapter 2). A number of possible sources of error have been considered in the section on thesis limitations. While respecting that both studies have strengths and weaknesses on balance, the findings from Study 1, due to the larger sample size, longer duration of follow up and inclusion of data from two separate medical schools, can be considered more robust and generalisable than those reported in Study 4.

Based on Study 1, the predictive validity findings suggest that HPAT-Ireland measures something relevant to future doctor patient communication, and this emerges as students progress from Year 1 to Year 2 of the programme. This

224

finding concurs with a small number of recent studies that have demonstrated that UKCAT has similar levels of predictive validity for communication and clinical skills (Husbands and Dowell, 2013, Husbands et al., 2014b).

HPAT-Ireland Section 2 was found to significantly correlate with Year 2 OSCE scores. This contrasts with the literature on UMAT Section 2, despite the strong similarities between this subsection of both tools (Mercer and Puddey, 2011, Shulruf et al., 2012, Simpson et al., 2014, Wilkinson et al., 2011). It is not immediately apparent why this difference should have arisen, unless cultural contexts may play a role. From a procedural justice perspective, job relevance is a critically important influence on stakeholder acceptability (Gilliland and Hale, 2005, Patterson et al., 2011). The quantitative and qualitative data converge with respect to the perceptions and evidence of job relevance of HPAT-Ireland Section 2. It is not surprising therefore that it emerged as the section most acceptable to stakeholders in this thesis.

By contrast HPAT-Ireland Section 3 was not considered clinically relevant. In terms of communication and clinical skills it was not found to be predictive. Abstract items, such as those contained in HPAT-Ireland Section 3, are generally considered poorly job related by applicants. It is recommended that one way to incorporate procedural justice into the design of cognitive tests is to use comparatively concrete item types(Fodchuk and Sidebotham, 2005). Many participants called for this section to be removed. However as Section 3 provides a measure of reasoning independent of language, it would be important to consider any possible negative consequence on applicants without English as a first language. It is also possible that Section 3 predicts performance outside of communication and clinical skills. The findings with respect to UMAT Section 3 are mixed and it has been found to negatively predict annual assessment scores in a small number of medical schools (Edwards et al., 2013, Mercer and Puddey, 2011).

Like Section 2, HPAT-Ireland Section 1 was also considered quite relevant to clinical performance. This thesis did not find evidence that Section 1 was predictive of future performance. This contrasts with reports on UMAT Section 1- which is generally found to be the most consistently predictive subsection of

UMAT (Edwards et al., 2013, Mercer and Puddey, 2011, Shulruf et al., 2012). However, this is likely to be due to the fact that Section 1 mostly predicts knowledge based assessments, which were outside of the field of this research.

On the whole, perceptions of the job relevance of HPAT-Ireland were quite good and levels of acceptability were very reasonable. This is in keeping with findings from elsewhere that indicate approximately 70% of medical students, applicants and General Practitioners support aptitude tests in principle (Kelly et al., 2014c). At the same time, stakeholders in this thesis remained sceptical of its ability to predict future good doctors. This is supported by procedural justice theory which differentiates perceptions of the relevance and beliefs about predictive validity as two discrete dimensions of job relatedness (Baeur et al., 2001). These stakeholders felt that other important influences such as the quality of medical education, student's motivation and personality factors, were so important, that the ability of HPAT-Ireland to predict future performance, as a doctor, would be negligible. This position finds some support in the work of Yates and James (2013) who determined that once measures of academic attainment acquired during medical undergraduate training were taken into account in statistical modelling, the predictive power of UKCAT was dwarfed.

From a distributive justice stance, fairness was a critical consideration in the overall acceptability of HPAT-Ireland to stakeholders, particularly for those who found it unacceptable. Concerns regarding the possibility that commercial coaching could lead to unfair advantage represent a breach of the principle of equal opportunity and mirror concerns expressed by other stakeholders (Kelly et al., 2014c). Research has shown that justice rules can be more influential, and weigh more heavily on overall estimation of acceptability, when they are violated rather than when they are satisfied (Gilliland, 1993, Greenberg and Colquitt, 2005). There is evidence that the raw scores of those who attend commercial coaching, or repeat HPAT-Ireland do slightly better with most of the gain in Section 3 (O'Flynn et al., 2012). However, research on UMAT showed that when other confounding variables such as academic record, gender, age and student motivation are controlled for, the gains are minimal if any (Griffin et al., 2012, Griffin et al., 2008, Wilkinson and Wilkinson, 2013).

Despite the media attention to issues of HPAT-Ireland and gender, this did not emerge as an area of concern in this thesis. There was convergence of evidence from quantitative and qualitative data. Stakeholders' perceptions that performance on subsections of HPAT-Ireland, according to gender would balance out overall, was supported by the data which showed no significant difference in overall scores between males and females. However further research is required to establish whether or not there is gender based prediction bias.

Multiple Mini Interview

We do not report any significant correlation between MMI scores and either weighted average First Medical Year Assessment or First Year OSCE results (Study 4). This contrasts with reports from others who have found MMI significantly predicts OSCE scores in Year 1 and 2 of the medical degree programme (Eva et al., 2004c, Eva et al., 2012, Eva et al., 2009, Husbands and Dowell, 2013). The section on thesis limitations considered potential sources of error in this study that may impact on the generalisability of these findings.

The evidence for the predictive validity of MMI is in its infancy. The fact that our study conflicts with others highlights the need for more work in this field and emphasises that the transferability of MMI cannot be assumed. For example the predictive validity of MMI may differ between applicants who have a prior undergraduate degree and those who are school leavers or it may be more predictive for problem based curricula rather than systems based or more traditional curricula.

Issues of fairness are essential to the acceptability of new selection tools. Stakeholders' perceptions reported in Study 3 that international students may be disadvantaged by MMI were developed in depth in Study 4. Assessors and students shared the view that cultural specificity and reduced English language proficiency served as barriers to international students, in violation of the distributive justice rule of equity. These findings have been discussed in detail in Chapter 8. There was convergence between these views and the quantitative findings that students from outside the EU, and those for whom English was not a first language, achieved lower mean MMI scores. The quantitative data also shows a very strong correlation between students' IELTS score and MMI.

However, on deeper exploration we did not find consistent patterns of MMI prediction bias with respect to these students or any of the subgroups that we examined. However this is a small study and there is a risk of a type two error. Further work is required to explore this potential source of bias.

There was a convergence between the views expressed quantitatively in Study 3 and those generated qualitatively in Study 4 relating to the advantages of MMI. Students believed MMI to be both an appropriate (Study 3) and highly authentic assessment of their suitability to medicine (Study 4). Participants in both studies agreed on the value and suitability of MMI to medical student selection.

One possible interpretation of this positive reaction relates to the procedural justice concept of "voice" (Colquitt et al., 2005, Gilliland, 1993). Students responded positively to the experience of voice, and recognised that the MMI process gave them "*more of a chance to show who you were*" (Study 4). Selection tools that offer an adequate chance for the applicant to make a case for themselves as well as sufficient time to do so are generally considered to meet the procedural justice rule of opportunity to perform (Gilliland and Hale, 2005). International studies have repeatedly shown that both of these aspects of MMI meet with stakeholder approval (Brownell et al., 2007, Dore et al., 2010, Eva et al., 2004a, Hofmeister et al., 2008).

MMI circuits also provide multiple opportunities for applicants to demonstrate abilities and the evidence suggests that this is particularly influential on positive reactions (Brownell et al., 2007, Dowell et al., 2012, Eva et al., 2004a, Kumar et al., 2009, Uijtdehaage et al., 2011). This may be partially mediated through provision of opportunity for reconsideration. The chance provided by MMI to "*redeem*" oneself has been positively noted elsewhere (Kumar et al., 2009).

Gilliland and Hale (2005) contend that co-workers are so invested in the outcome of selection, they feel the need to have direct input into the decision making process. Likewise MMI includes the assessor directly in selection judgements and provides an opportunity for them to have voice. By contrast assessors are removed and distant from decisions made by aptitude tests which

may contribute to relatively poorer ratings of this tool. Gilliland and Cherry (2000) stress that additionally in order to increase acceptability, all stakeholders should have an opportunity for voice in the design and development stage of selection processes. The list of twelve recommendations generated from focus groups with stakeholders in Study 4, is an example of such voice. It is important that efforts are made to act on these recommendations and report back to the stakeholders.

The positive views of the job relatedness of MMI were highly influential on assessors' and students' perceptions of its acceptability and fairness. Furthermore, students in particular felt that the MMI primed them for subsequent learning. Feedback on MMI performance was given to each student in writing, indicating their strongest and weakest stations and how they could use this information to identify learning needs. Timely and informative feedback is influential on perceptions of interactional justice (Gilliland and Hale, 2005). While feedback can often be thought of as necessary only for unsuccessful applicants, this thesis suggests that MMI feedback is of benefit to successful candidates and provides a feasible model for doing so.

Although stakeholders' views of MMI expressed in this thesis are generally positive, they were not universally so. Participants' views tended to be somewhat more critical than the international body of research endorsing this selection tool. This may reflect the fact that interview is not routinely part of the selection process in Ireland. Concerns were expressed by assessors and students in Studies 3 and 4 that extroverted candidates may have an advantage. Elsewhere, research comparing students' scores on MMI with scores on the Big Five personality factors have been equivocal. Two studies found significant correlation between extraversion and MMI scores (Griffin and Wilson, 2012, Jerant et al., 2012) with the former also finding a significant association with conscientiousness. Whereas Kulasegaram et al (2010) found no such association. The bigger question of whether these personality factors represent significant attributes relevant to future on the job performance remains to be answered and likely to impact on stakeholder acceptability .

Lastly the possibility that commercial coaching could lead to enhanced performance for those who could afford it was raised in both Studies 3 and 4.

Again this would contravene the rules of distributive justice. To date the evidence suggests that MMI is relatively unsusceptible to coaching, but this is based on the findings from a limited number of studies and requires further exploration (Griffin et al., 2008, Reiter et al., 2006) .

To summarise, the findings from all four studies resonate with the constructs of organisational justice theories. There was a particularly strong resonance with both procedural and distributive justice. Figure 9 shows how the overall thesis findings map onto organisational justice theories.
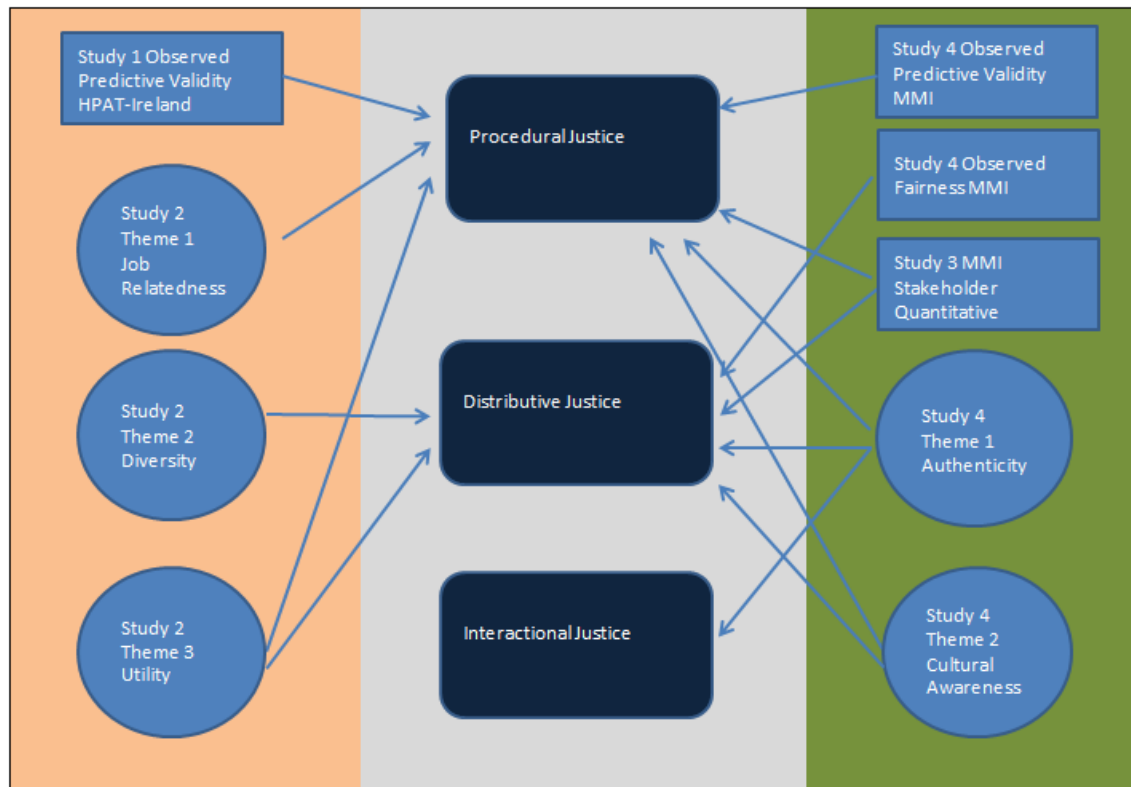
Figure 9 Mapping Findings to Organisational Justice Theories

Legend: Peach background denotes Phase 1: HPAT-Ireland Studies. *Green background denotes* Phase 2: MMI Studies. *Quantitative studies/strands are shown in boxes. Qualitative studies/strands in circles.*

9.4 Conclusions and Recommendations

This thesis concludes that, by comparison with other aptitude tests in its class, there is sufficient evidence of the predictive validity and stakeholder acceptability of HPAT-Ireland to justify its continued use in the selection of medical students, subject to the following provisos:

- Its ongoing use should be contingent on, the outcome of a national predictive validation study. Ideally this study would involve all five medical schools, following up three cohorts of students across all years of the undergraduate degree programme and into higher professional training. A follow-up study of the UCC and NUI Galway cohort is underway, while the National Research Group Evaluating Revised Entry Mechanisms to Medicine is giving consideration to the larger study.

- A structured programme of regular consultation and feedback with a wide group of stakeholders is required to ensure the political validity of HPAT-Ireland. To date the National Research Group Evaluating Revised Entry Mechanisms to Medicine have conducted four other stakeholder acceptability studies. However, all these relate to stakeholder views in the timeframe immediately after introduction of HPAT-Ireland. These need to be revisited and expanded.

- In order to be accountable to stakeholders an annual report should be made publicly available including details, not only of HPAT-Ireland, but all aspects of the selection process to medicine. It is essential that this includes details of applicant scores and standard deviation, and test psychometric information such as measures of reliability and fairness.

- Evidence for the job relatedness and predictive validity of Section 3 are lacking, and it appears somewhat susceptible to coaching. Adequate consideration needs to be paid to stakeholders' recommendations to remove this section and replace it with a more robust and acceptable measure that relates directly to the on the job requirements of the medical graduate.

- Further research be undertaken to identify any evidence of negative consequences relating to the changes to selection in Ireland and HPAT-Ireland. These include examining the impact on student diversity,

particularly with respect to socioeconomic diversity and the relationship with future career intentions of successful applicants. Socio-economic diversity could be estimated by mapping applicants' postal address or designated school status to markers of deprivation. Evidence for the impact of repeating HPAT-Ireland on test validity, diversity and on dropouts from other university programmes needs to be continually monitored.

With respect to MMI this thesis concludes that:

- Issues of culture and language are crucial to stakeholders' perceptions of the fairness of MMI. The growing use of MMI internationally means that it will increasingly be used in cultural contexts beyond that in which it was originally developed. We found no consistent objective evidence to suggest that MMI is unfair however there is a need for further adequately powered studies to establish this for applicants from a variety of cultural, ethnic and social backgrounds. Further research should also seek to establish if following the recommendations suggested by stakeholders in this thesis for culture proofing the MMI impacts positively on perceptions of fairness or actual performance of international applicants on MMI.

- MMI is feasible in an Irish setting, generally acceptable to stakeholders and worthy of further consideration. However, its introduction to the mainstream selection process would represent a very significant departure from current practice and would entail a substantial degree of input and resources from the medical schools. This is not warranted without further evidence that it demonstrates predictive validity in our cultural and educational context. Furthermore, were MMI to be introduced it should be developed based on a thorough job analysis of the role of the medical graduate. Due consideration would have to be taken of the fact that a high proportion of Irish medical graduates return to practice medicine in their country of origin. This represents a significant body of work, however it would be an essential step in the development of an MMI that would be fit for purpose. Finally, given that some reservations were expressed by stakeholders in this thesis, consultation

with a wider group of stakeholders, in particular potential applicants, would be necessary before a decision to introduce this new tool could be justified.

9.5 Directions for Future Research

This thesis highlights the conflicting evidence regarding the utility of aptitude tests in the selection of medical students. Future research should aim to conduct a systematic review and meta-analysis of the predictive validity of the newer aptitude tests.

Organisational justice theories proved a very helpful conceptual lens through which to integrate and synthesise the findings from the various strands of this thesis. Further research should consider the use of these theories to inform a theory driven mixed methods exploration of stakeholder views. This would entail using organisational justice theories to inform the research question, design, data collection, analysis and interpretation. For example, quantitative data collection could incorporate use of a validated scale such as the Selection Procedural Justice Scale (Bauer et al., 2001), while qualitatively the topic guides could be closely mapped to the three dimensions of organisational justice theories.

A number of key stakeholder groups have been overlooked in the literature with respect to medical student selection. Patients, parents of applicants, teachers, potential applicants and future employers are important groups, whose insights are likely to benefit selection processes. Future research should aim to incorporate the views of these key stakeholders.

An underexplored key consideration is deciding the appropriate weighting and relative order of use of selection tools. The order in which tools are used has the potential to exert powerful influence on the selection process. One way to examine this would be to conduct a multicentre international randomised controlled trial to assess the effect of varying the order and weighting on issues of predictive validity, stakeholder acceptability, fairness and impact on widening diversity. This research may be feasible to conduct via the newly established International Network for Researchers in Selection into Healthcare (INReSH).

Research identifying the actual constructs tested by MMI is in its infancy. Further research is required to establish the skills and attributes applicants draw on to deal with the issues at each MMI station. One way to examine this would be to insert debriefing stations after each live station, where candidates would talk about what they understood to be the task at the station, how they approached it and what skills they used to do so.

9.6 Concluding Comments

The selection of medical students has come a long way in the past 95 years. Real advances have been made with respect to transparency and critical evaluation of selection systems. We have a moral and social responsibility to continue to strive for improvements to our selection systems not just for the sake of our applicants, but for all our stakeholders most importantly our future patients.

The changes introduced in Ireland in 2009, although unpopular in many ways, have provided a unique opportunity to concentrate attention on the process of selecting medical students. For the first time sincere debate about the merits of various selection tools, be they aptitude tests, academic record or interview, are being had by many different stakeholders. Real consideration is being given to new and innovative methods such as MMI. There is a growing realisation that the old reliable systems were not without their shortcomings. Widening diversity is on the agenda. Irish research is being conducted that contributes in a meaningful way to the international debate on selection.   This thesis is part of a story of medical student selection research that is just beginning.

Appendices

Appendices

Appendix A Outputs and Initiatives

This appendix lists the conference presentations, fellowship and awards, professional collaborations, research funding and initiatives arising from this thesis.

Table 23 Conference Presentations

Title / Co-authors / Presenting author underlined	Conference (Oral/ Moderated Poster)
Using Multiple Mini Interview in an internationally diverse medical student population Kelly Maureen E, Dowell Jon, Husbands Adrian, Newell John, O 'Flynn Siun, Kropmans Thomas, Dunne Fidelma P, Murphy Andrew W.	Accepted for oral presentation AMEE, Glasgow, Scotland 2015
Medical Student Selection – An international perspective. Maureen E Kelly	Association of University Departments of General Practice in Ireland (AUDGPI), Belfast March 2015. Invited Conference Closing Plenary
The role of Aptitude Testing in Selection Maureen E Kelly	Irish Network of Medical Educators - INMED University of Limerick, Feb 2015 (Invited short oral presentation)
Using Multiple Mini Interview in an internationally diverse student population. Kelly Maureen E, Dowell Jon, Husbands Adrian, Newell John, O 'Flynn Siun, Kropmans Thomas, Dunne Fidelma P, Murphy Andrew W.	International Network for Researchers in Selection into Healthcare (INReSH) Woburn House Conference Centre, London November 2014 (moderated poster)
Stakeholder views on selection to undergraduate medicine in Ireland. Kelly Maureen E, Arnett Richard, Dennehy Thomas, Duggan Eileen, Dunne Fidelma, Gallagher Niamh, Hennessy Martina, Lapthorne Susan, Last Jason, Mc Grath Deirdre, Murphy Andrew W, Stevens Laragh, O'Flynn Siun	International Network for Researchers in Selection into Healthcare (INReSH) Woburn House Conference Centre, London November 2014 (moderated poster)
Aptitude testing may negatively predict student performance Kelly Maureen E, Dowell Jon, Husbands Adrian, Newell John, O'Flynn Siun, Kropmans Thomas, Dunne Fidelma P, Murphy Andrew W	Ottawa conference. Convention Centre, Ottawa April 2014 (oral).

Appendices

Doctors' views on the Health Professions Admission Test Ireland (HPAT Ireland) as a selection tool for medicine- A qualitative study Maureen E Kelly, Niamh Gallagher, Fidelma Dunne, Andrew W Murphy	Association of University Departments of General Practice in Ireland (AUDGPI), University College Cork, March 2014 (oral)
Widening access to medicine- a community outreach pilot programme ME Kelly, CM Connolly, C Mac Liathain, S Neary, T Ó Flartharta, L Nicholson, S Lally, Y Finn, A Hynes, D Ashe, H Hamilton, G Flaherty	Irish Network of Medical Educators - INMED Queen's University, Belfast, Feb 2014 (Moderated Poster)
To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills? Maureen E Kelly, Daniel Regan, Fidelma Dunne, Patrick Henn, John Newell, Siun O'Flynn	Irish Network of Medical Educators - INMED University College Dublin, Feb 2013 (Moderated Poster)
Medical Student Selection: Can Multiple Mini Interviews work in an Irish setting? – A feasibility study Maureen E Kelly, Jon Dowell, Adrian Husbands, Thomas Kropmans, Aoife Jackson, Fidelma Dunne, Siun O' Flynn, John Newell, Andrew W Murphy	Irish Network of Medical Educators - INMED University College Dublin, Feb 2013 (Oral)
To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills? Maureen E Kelly, Daniel Regan, Fidelma Dunne, Patrick Henn, John Newell, Siun O'Flynn	Association of University Departments of General Practice in Ireland (AUDGPI), University of Limerick, March 2013 (oral)
Medical Student Selection. Maureen E Kelly, Jon Dowell	ASME 20th July, 2012 Conference Centre, Brighton UK (Oral)

Table 24 Non-Conference Presentations

Title	Location
What has HPAT-Ireland brought to selection? Maureen E Kelly	Population Health Sciences Divisional Seminars- Royal College of Surgeons in Ireland May 2015 (Invited lecture)
What has HPAT-Ireland brought to selection? Maureen E Kelly	Medical School Grand Rounds, NUI Galway Jan 2015 (Invited lecture)
The role of HPAT-Ireland in selection? Maureen E Kelly	Medical School Board Executive Meeting, July 2014 (invited short presentation).

Appendices

Table 25 Awards / Fellowship

Title	Awarding Body
Irwin Prize 2014	AUDGPI – Best Educational Research Presentation AUDGPI Conference 2014
Irwin Prize 2013	AUDGPI – Best Educational Research Presentation AUDGPI Conference 2013
Educator Development Group (EDG) Travelling Fellowship 2011	Association for the Study of Medical Education

Table 26 Associated Publications

Title	Authors/ Reference
2015 Report –School Leaver Entrants - National Research Group Evaluating Revised Entry Mechanisms to Medicine Report to the Council of Deans	Dr. Siun O Flynn, Dr Maureen Kelly, Prof Deirdre McGrath, Dr Martina Hennessy, Dr Jason Last, Dr Richard Arnett, Dr Paul Corcoran, Dr Tony Fitzgerald, Mr Ivor Gleeson, Ms Celeste Golden , Dr Sean Dinneen, Prof Mike Larvin
Irish Medical school applicant views on Selection Tools for Medicine	Siun O'Flynn, Eileen Duggan, Susan Lapthorne, Maureen Kelly , Martina Hennessy , Jason Last, Richard Arnett, Deirdre Mc Grath. Submitted to Irish Medical Journal Nov 2014
Medical Students' Views on Selection Tools for Medical School - A mixed methods study"	Stevens L, Kelly ME, Hennessy M, Last J, Dunne F, O'Flynn S. Irish Medical Journal Vol.107. No.8 Sept 2014
General Practitioners' perspectives on revised entry and selection mechanisms to medicine and the HPAT	T Denehy, M Kelly, S O'Flynn Irish Medical Journal 2013: *106;* 4:113-116.
Medical Student Selection	Kelly ME, Dowell J. . Association for the Study of Medical Education 2012:51 (abstract)

Appendices

Table 27 Research Funding

Funder and Award	Purpose of Award	Amount (Date)
Irish Network of Medical Educators (INMED)	Travel Bursary— Ottawa Conference	*€1,000 (2014)*
WestREN – Small grant award	Dissemination of research findings Study 3	€1,000 (2014)
ICGP Research and Education Grant	Follow-up to Study 1	*€1,500 (2013)*
WestREN – Research Bursary	Study 2 and 3	€10,000 (2013)
WestREN – Small grant award	Pilot phase of Study 2	€1,000 (2012)
ASME Educator Development Grant	Travelling Fellowship	€1,950 (2011)
NUI Galway Millennium Fund	Study 1	€7,305 (2010)
Further Educational Policy Scheme, NUI Galway	PhD Fees	€2,785 (2011) renewed annually for the four years

Professional Collaborations

This PhD thesis has provided excellent opportunities to invest in productive professional collaborations with recognised leaders in the field of selection. Attendance at an ASME run "Meet the experts" Special Interest Group – Medical Student Selection in York in 2011 led me to Dr Jon Dowell from Dundee Medical School. This collaboration has resulted in two peer reviewed publications with colleague Adrian Husbands (Studies 3 and 4) and plans for continued collaborations.

I have also had the opportunity to build a strong link with Dr Siun O'Flynn, Medical School, University College Cork and she has contributed to three of the papers in this thesis. Our collaboration has also resulted in three additional publications on stakeholder acceptability of HPAT-Ireland outside of the parameters of this thesis (listed below- associated publications). A follow-up study on the predictive validity of HPAT-Ireland in the senior clinical years is underway.

Through collaborations with a colleague, Anne Marie Regan, completing her doctoral thesis on selection, I have had the opportunity to contribute to the development of a Situational Judgement Test, for selection to postgraduate GP training. I participated in a two day residential workshop facilitated by Professor Fiona Patterson and colleagues from Work Psychology.

Appendices

In November 2104 I attended and chaired a session at the inaugural meeting of the International Network for Researchers in Selection into Healthcare (INReSH). This new international network for researchers in selection should provide a supportive platform for an on ongoing research agenda. I look forward to building and expanding on the connections I made at this meeting.

Lastly I have been appointed to the National Research Group Evaluating Revised Entry Mechanisms to Medicine which is a consortium of medical educators, researchers and statisticians who meet under the auspices of the Council of Deans of the Medical Faculties of Ireland. Our role is to conduct research on selection to medicine in Ireland and advise future policy decisions.

Initiatives

Working in this thesis has made me keenly aware of the need to widen diversity in medicine. It has also made me realise that one of the critical steps to achieve this is to ensure that non-traditional, disadvantaged and minority candidates get in to the applicant pool in the first instance. I have set up and oversee a number of Medical School initiatives to help foster widening diversity within this institution.

Clár Ambasadóireachta Scoil an Leighis –(Medical School Ambassador Programme)

This is a school outreach programme that I established as part of NUI Galway Medical School's strategy to widen access. The Gaeltacht is a collection of geographical areas where Irish is the principal spoken language. The areas are generally rural and remote. Some areas experience significant disadvantage due to limited employment options, high levels of emigration and poor services. The Galway Gaeltacht is the largest in the country and covers an area of over 1,200 km^2. Students from Gaeltacht secondary schools are underrepresented amongst the undergraduate student population. Focus groups with medical students from Gaeltacht backgrounds revealed that they experienced a range of obstacles in their path to successfully securing a place in medical school. The aim of this programme was to encourage Gaeltacht school leavers to apply to medicine.

The programme was developed as a community partnership between Irish speaking undergraduate medical students, School of Medicine staff and two

Gaeltacht school principals. This working group developed a series of workshops to address perceived barriers to students from the Gaeltacht applying to medicine. The workshops are designed to be age appropriate and give some basic information about what doctors do, what the life of a doctor is like, what is required to study medicine and how to apply. To date the workshops have been delivered in three national schools and seven secondary schools, reaching over 500 students. In a number of schools they have been linked with a work shop teaching cardio pulmonary resuscitation. The feedback from school children and teachers has been very positive. School children particularly valued being taught a medical skill as well as having some myths about what one needs to be a doctor dispelled. Medical students' written reflections demonstrated enthusiasm about "giving something back" and an increased awareness of their own professional values. However the true value of this intervention remains to be seen.

Mature Entry to Medicine

In 2013, with the support of then Head of the Medical School, Professor Fidelma Dunne, I established a new mature entry route to medicine. This route is aimed at non-traditional students. We now offer two places per annum via this route and it has the ongoing support of the Medical School Board and new Head Dr Sean Dinneen. To qualify as eligible for this route applicants need to be aged over 23 and to have an honour in Leaving Certificate Chemistry.

Eligible mature applicants are rank-ordered based on their HPAT-Ireland score and the top performing (approximately 15) candidates are invited to attend a MMI. The MMI stations are blueprinted against the eight domains of professional practice by the Irish Medical Council. We use 10-12 stations and they last 7 minutes. Interviewers are drawn widely from across the spectrum of disciplines and departments in the medical school. Unofficial feedback from candidates indicates that the process appears acceptable. To date four applicants took up their place in medical school via this route.

Appendix B Appendices Relating to Thesis Chapters

Chapter 1 Appendix

Chapter 1 Appendix 1: Copyright permission to reproduce Yale Photograph

Appendices

Chapter 1 Appendix 2: Permission to Reproduce Table of Evaluative Standards
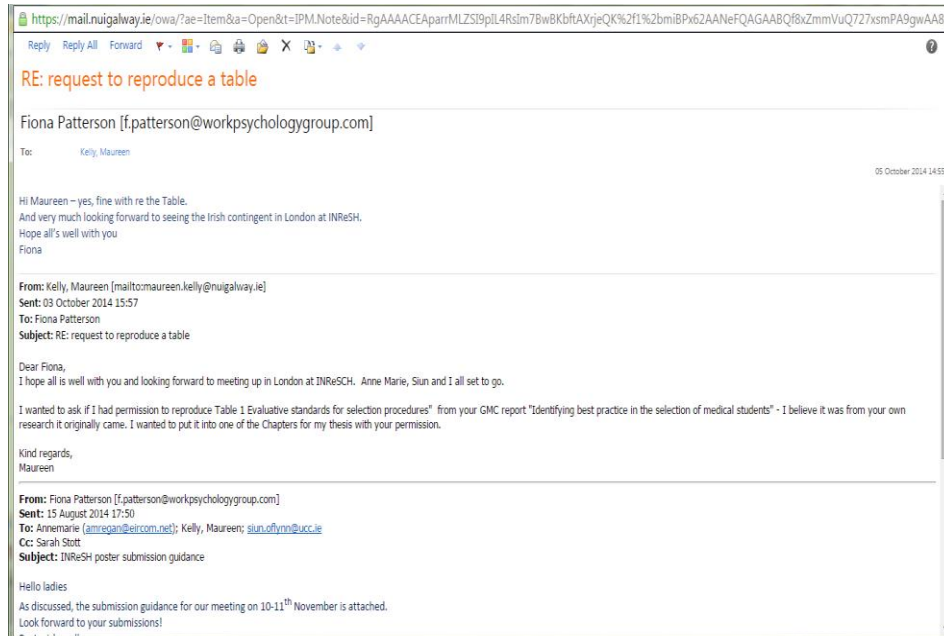
Chapter 2 Appendix

Chapter 2 Appendix 1 Search Strategy: Aptitude Tests for Selection to Medicine
Ovid Medline

	Search Statement	
1	(MCAT or "medical colleges admission test").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	505
2	(UKCAT or "UK Clinical Aptitude Test").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	27
3	(UMAT or "undergraduate medical and health professions admissions test").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	31
4	(BMAT or "BioMedical Admissions test").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	27
5	(HPAT or "health professions admission test" or "HPAT-Ireland" or "health professions admission test- Ireland").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	38
6	(GAMSAT or "Graduate Medical Schools Admissions Test").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	11
7	Aptitude tests.mp. or exp Aptitude Tests/	19578
8	general mental ability.mp.	53
9	School admission criteria.mp. or exp School Admission Criteria/	4246
10	college admission test.mp. or exp College Admission Test/	674

11	Entrance examinations.mp.	58
12	exp Education, Medical/	132711
13	9 or 10 or 11	4728
14	12 and 13	1602
15	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8	20222
16	14 and 15	178
17	**limit 16 to (english language and humans and yr="2000 - 2014")**	**97**

Appendices

Chapter 2 Appendix 2: Stakeholder Views Search Strategy Ovid Medline

Searches	Results	Search Type
41	**limit 40 to yr="2000 - 2014"**	**268**
40	18 and 25 and 31 and 39	393
39	32 or 33 or 34 or 35 or 36 or 37 or 38	17243
38	("residency selection$" or "select$ to residency" or "select$ for residency" or "recruit$ to residency" or "recruit$ for residency" or "training selection" or "select$ to training" or "select$ for training" or "training recruit$" or "recruit$ to training" or "recruit$ for training").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	1609
37	("medical school$ admission" or "admission$ polic$" or "medical school$ entr$").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	520
36	("medical student$ selection" or select$ medical student$*" or "medical student$ admission$" or "admit$ medical student$").mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	17
35	exp Personnel Selection/	11107
34	Entrance examinations.mp.	58
33	exp College Admission Test/	499
32	exp School Admission Criteria/	4244
31	26 or 27 or 28 or 29 or 30	145000
30	exp "Internship and Residency"/	36732
29	exp Education, Medical, Graduate/	22420
28	exp Education, Medical, Undergraduate/	18619
27	exp Education, Medical/	132711

26	exp Schools, Medical/	21241
25	19 or 20 or 21 or 22 or 23 or 24	942374
24	(view* or opinion* or perspective* or belie* or attitude*or accept*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	707830
23	student attitude.mp.	44
22	job applicant attitude.mp.	0
21	exp Public Opinion/	15867
20	exp "Attitude of Health Personnel"/	124030
19	exp Attitude/	273753
18	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	616894
17	public.mp.	355221
16	"school leaver*".mp.	235
15	applicant*.mp.	5507
14	stakeholder*.mp.	13274
13	"Overseas student*".mp.	67
12	"international student*".mp.	279
11	"international medical graduate*".mp.	528
10	exp Patients/	69769
9	exp Physicians, Primary Care/	1475
8	exp Foreign Medical Graduates/	3058
7	exp General Surgery/	34841
6	exp General Practitioners/	2350
5	exp Physicians/	89482
4	exp Medical Staff/	24324

Appendices

3	exp "Internship and Residency"/	36732
2	exp Faculty, Medical/	10248
1	exp Students, Medical/	22673

Appendices

Chapter 2 Appendix 3- Excerpt from Data Extraction Form

Author /Year/ Type of article	Location / Setting	Study Design /Aim	Selection tool(s)	Stakeholder characteristics	Data Collection method	Outcome variable	Overall findings	Quality Mark
Adams 2009 * **Research paper**	USA Medical Schools	Quantitative Aim to establish the acceptability of online courses as criteria for admission to medical school	Academic record – either earned online or in a traditional setting	Medical school administrators who evaluated applications n=58, response rate 49%. 88% were over 45yrs. 38% had taken an online course themselves	Postal questionnaire survey.	Each participant was given a three situations each with 2 applicants; identical except for one had an online course the others traditional course at either university or community college. Participants were asked to rate them and select one for interview, plus give reasons why.	100% of Medical school administrators selected for interview the applicant with the university course over that with the online course, 93% selected the applicant with the university course over that with a community college course, and 79% selected the applicant with a community college course over that with an online course.	7
Agrawal et al 2005* **Research paper**	USA Medical Schools	Quantitative Aim to assess current initiatives at U.S. medical schools to recruit underrepresented minorities (URM) and to identify perceived barriers to enrolment of URM students.	MCAT/ GPA / Other factors	Deans of Student Affairs or Medical Admission (schools who identified as selecting high URM already were excluded) N= 86 response rate 59%.	Postal questionnaire survey developed and piloted after extensive literature review and work by the American Medical Student Association Diversity Coalition, piloted in 4 medical schools	Percentage responses to a list of 37 potential barriers to URM recruitment Responses to a list of 11 possible recruitment programmes – indicate if they had it / rate its effectiveness	Mean 10.4% of students were URM. Educational barriers- 90% of respondents perceived low MCAT score as a barrier to URM selection, 60% low GPA, poor science preparation 55% Sociocultural absence of role models 77% Financial – lack of financial aid 48% Recruitment/Admission – Not enough minority faculty members 71%	8.5

Chapter 3 Appendix

Chapter 3 Appendix 1 – Points Calculation

a) *How Leaving Certificate points are calculated*

For entry to medicine applicants' scores on six subjects are taken into account in a system that awards a certain number of *points* for each grade achieved. Points are awarded as follows: 100 points awarded for an A1 grade in a higher level subject (equating with a score of 90-100%): 90 points awarded for an A2 grade (equating with a score of 85-89.99%): 85 points awarded for a B1 grade (equating with a score of 80-84.99%) and so forth (theleavingcert.com 2014).

For EU applicants outside of Ireland an equivalent academic record is accepted such as the A level and there is a system of converting to the appropriate points. Applicants are required to score a minimum of 480 points in the LCE to be considered eligible; however to secure a place in medicine applicants must score substantially higher than this.

b) *Moderating LCE points*

Moderation of the LCE score entails an adjustment being made to the raw LCE score. When moderated LCE points in excess of 550 are adjusted so that every 5 points earned on the raw count are allocated just 1 point. Therefore a raw count score of 555 points is moderated to 551; a score of 560 points moderated to 552 and so forth until a raw count score of 600 points is worth 565. These moderated scores are known as *Adjusted Leaving Certificate Scores*.

c) *Weighting of Academic Points and HPAT-Ireland Score*

The weighting of scores between LCE and HPAT-Ireland was originally approximately two to one 560 (Maximum Adjusted LCE points): 300 (Maximum HPAT-Ireland points). In 2012 a further change to entry and selection to *all* higher institution course saw that higher level mathematics was awarded a bonus of 25 points, hence the maximum points available for the Adjusted LCE is now 565 (allowing for moderation of the bonus for applicants earning in excess of 550 points). The minimum entry requirement of combined LCE and HPAT-Ireland score differs very slightly between medical schools but is approximately 720 points.

Chapter 4 Appendix

Chapter 4 Appendix 1 Ethical Approval Study 1
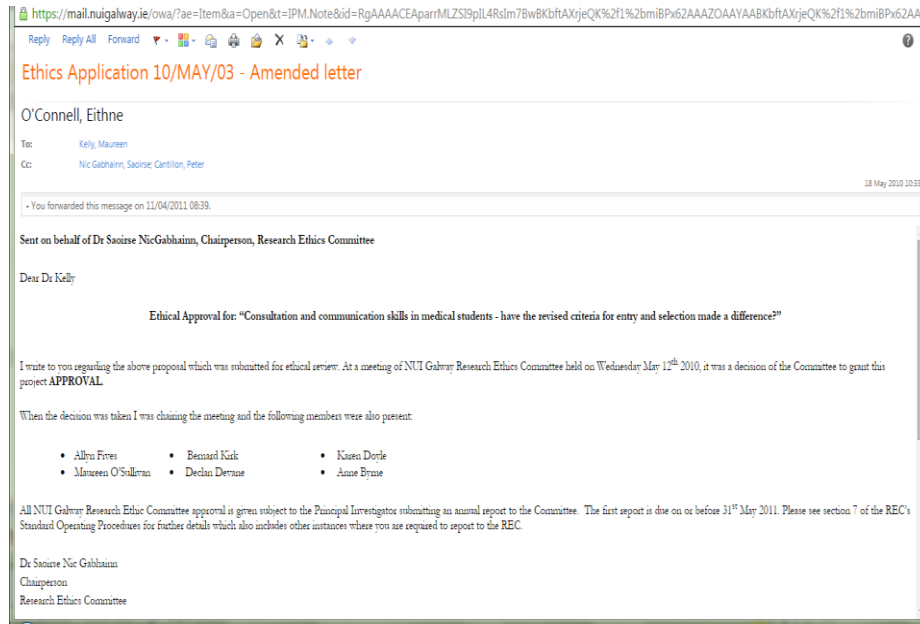
Appendices

Chapter 4 Appendix 2 Ethical Approval Pilot Phase Study 2

Reply Reply All Forward ▼ ▾ ▦ ▾ ✎ ⎙ ⌦ ✕ ⧉ ▾ ⬆ ➡   ❓

Ethics Ref: 11/JAN/08 – Full approval

O'Connell, Eithne

To: Kelly, Maureen
Cc: Nic Gabhainn, Saoirse; Cantillon, Peter
                                                                                                        16 March 2011 11:20

Sent on behalf of Dr Saoirse NicGabhainn, Chairperson, Research Ethics Committee.

Dear Dr Kelly

RE: Ethical Approval for "Establishing the Face Validity of the Health Professions Aptitude Test (HPAT): A pilot Study."

I write to you regarding the above proposal which was submitted for Ethical review. Having reviewed your response to my letter, I am pleased to inform you that your proposal has been granted **APPROVAL**.

All NUI Galway Research Ethic Committee approval is given subject to the Principal Investigator submitting an annual report to the Committee. The first report is due on or before 31st January 2012. Please see section 7 of the REC's Standard Operating Procedures for further details which also includes other instances where you are required to report to the REC.

Yours Sincerely

Dr Saoirse NicGabhainn
Chairperson - Research Ethics Committee

(Should you require a signed hard copy of this letter please contact me, details below)

From: O'Connell, Eithne
Sent: 04 February 2011 16:23
To: Kelly, Maureen
Cc: nic Gabhainn, Saoirse; Cantillon, Peter
Subject: Ethics Ref: 11/JAN/08

Appendices

Chapter 4 Appendix 3 Ethical Approval Study 2

Appendices
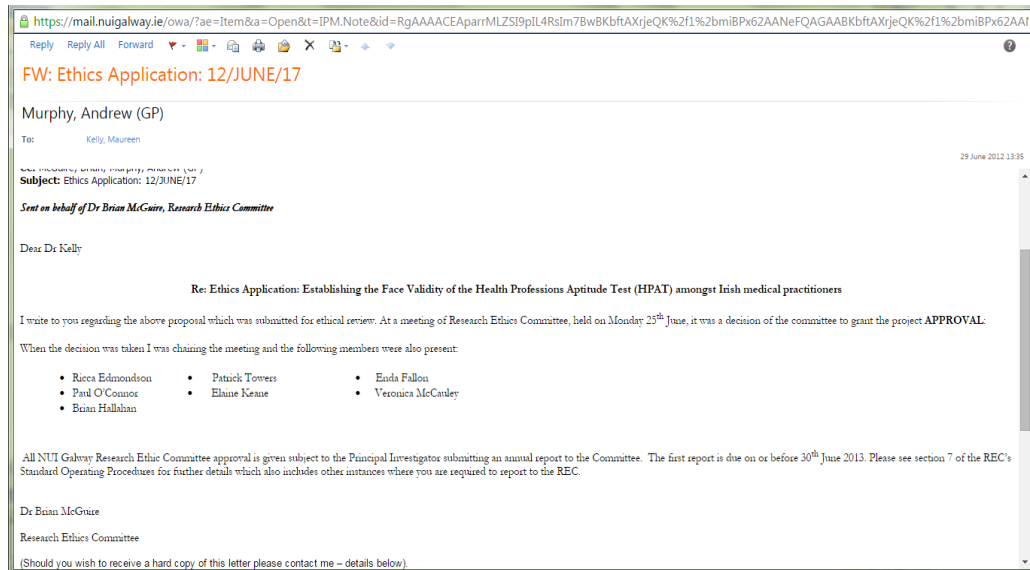
Chapter 4 Appendix 4 Ethical Approval Studies 3 and 4
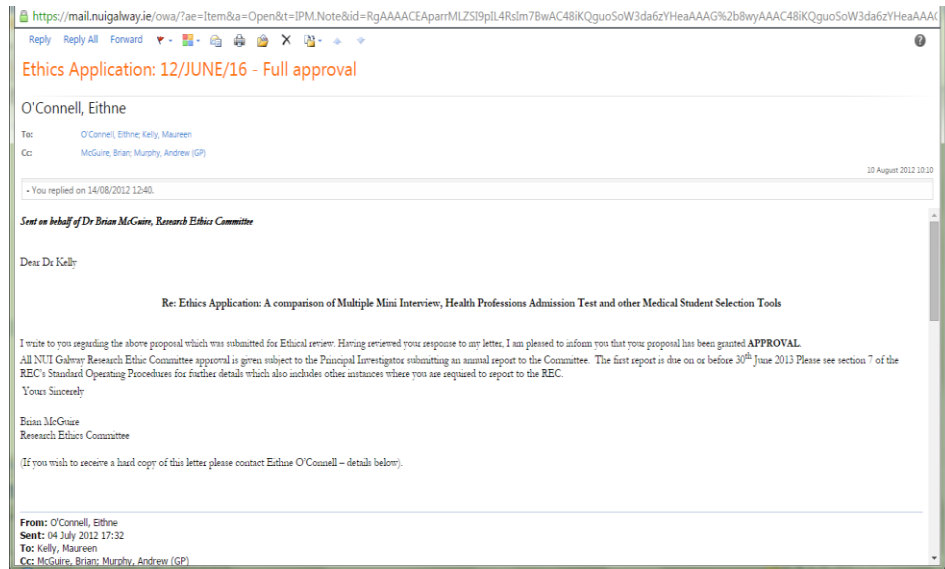
Chapter 4: Appendix 5- Review of Application for WestREN funding

WestREN Research Bursary 2012
External review. Prof Susan Smith. April 2012.

Application 1: Establishing the role of the Health Professions Admission Test (HPAT) as a selection tool for Irish undergraduate Medical Schools.
PI: Dr Maureen Kelly
Score: 8/10
This proposal involves an educational research project. It deals with an issue of national and international importance and can feed into policy in this area, which is due to change in coming years. Assessment for medical school entry is of interest to all the stakeholders mentioned in the application. The research findings will have broader implications for assessment of entry into postgraduate training programmes as well. The aims are clear and the methodology is sound. The track record of the applicant and her supervisor is appropriate. There are excellent institutional supports in place and there is added value in that the PI will undertake a higher degree using this proposed project thus building academic capacity in Irish general practice.

Chapter 4 Appendix 6 – Excerpts from Reflective Research Diary

Excerpt from reflective diary Study 2- (Diary Entry 6th March 2012)

 "I need to actively look for evidence that contradicts the open code "Sense of disillusionment with HPAT" as my own viewpoint might lead me to concentrate on the negative views held by interviewee rather than positive ones. CREATE AN OPEN CODE "Sense of affirmation with the HPAT" and actively look for evidence of this" –

Excerpt from reflective diary Study 2- (Diary Entry April 3rd 2012)

 "I am picking up that the interviewees are commenting a bit on maturity and life experience, being an advantage in doing well in the HPAT. That the age and maturity of the HPAT candidate is an important factor in doing well…… … Need to check for counter arguments in case I am being too receptive to this notion".

Excerpt from reflective diary Study 4 – (Diary Entry 31st Jan 2014)

 "Read and re-read Andrew's , Adrian's and my coding of the international students focus group. Remarkable agreement between the three independent coders for about ¾ of the content. Andrew was more clued into the testing of communication skills and how this featured in the MMI and also into the sense that MMI was not as easy to coach candidates for as the traditional interview. Adrian picked up more on the "realism" that the MMI offered and that it appeared consistent with life as a doctor, I picked up more on the fact that MMIs should be more culturally challenging for all candidates as opposed to just international students".

Excerpt from reflective diary Study 4 – (Diary Entry 7th Feb 2014)

 "– recoded Irish / EU student focus group – having read over the original interview and re-read the international student interview with Adrian and Andrew's comments. Changed a number of codes. Overall very positive reaction from Irish students- lists of advantages and disadvantages / lots of comparisons with HPAT. Summarised the main findings in note form. Did and undid a list of advantages and disadvantages of MMI as might be too soon to do this as could blinker vision of examiner interviews ".

Chapter 5 Appendix

Chapter 5 Appendix 1 –Confirmation of copyright approval to reproduce papers 1 and 4

Dr Maureen E Kelly,
Discipline of General Practice,
Clinical Science Institute,
NUI Galway

Mr. Maxwell Allen-Benton,
Production Editor,
BioMed Central,
Floor 6,
236 Grays Inn Road,
London, WC1X 8HB

March 10th 2015

Dear Mr Allen-Benton,

I am completing a doctoral thesis at National University of Ireland, Galway and I am contacting you to request permission to include the following materials within the electronic version of my PhD:

1)To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills? – An observational cohort study

Maureen E Kelly, Daniel Regan, Fidelma Dunne, Patrick Henn, John Newell, Siun O'Flynn *BMC Medical Education 2013, :68 (10 May 2013)*

2) The fairness, predictive validity and acceptability of Multiple Mini Interview in an internationally diverse student population- a mixed methods study Kelly ME, Dowell J, Husbands A, Newell J, O 'Flynn S, Kropmans T, Dunne FP, Murphy AW BMC Medical Education 2014 (1), 267

I would like to make an electronic version of my thesis available in ARAN which is the National University of Ireland, Galway's online repository http://aran.library.nuigalway.ie/. Once available in digital format, access to the thesis will be freely available via the Web. All users of the thesis will be bound by a Creative Commons Attribution Non-Commercial No Derivatives license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

I would be grateful if you, or the company you represent, could grant me permission to include the Material in my thesis and to use the Material, as set out above, royalty free in perpetuity. Please be aware that the Material will be archived in ARAN in perpetuity.

As BMC is an Open Access publisher, all manuscripts are published under the Creative Commons Attribution License which allows re-distribution and re-use of a licensed work on the condition that the creator is appropriately credited (you can find information at this link - http://opendefinition.org/licenses/cc-by/). You do not require approval or confirmation to re-distribute or re-use the works in any way. The only

requirement when doing this is that BMC is cited correctly as the original publisher of the manuscript.

.....M. Allen-Benton........

Maxwell Allen-Benton

Date ...11/03/2015

Yours sincerely,

Dr. Maureen Kelly MICGP, FRCGP, M Med Ed.,

Lecturer in General Practice,Discipline of General Practice,

College of Medicine, Nursing and Health Sciences,

National University of Ireland Galway.

Associate Programme Director, Western Training Programme in General Practice

Tel: (353) 091 494107 (Monday and Tuesday)

Fax: (353) 091 494559

Email: maureen.kelly@nuigalway.ie

Webpage: www.nuigalway.ie/dgp/

Chapter 5 Appendix 2: Agreed National Selection Criteria for Undergraduate Medical Schools in Ireland

The Leaving Certificate Examination adjusted and Leaving Certificate Examination/HPAT-Ireland combined scores are based on the agreed national selection criteria for undergraduate medical schools in Ireland. These criteria dictate the relative weighting of the LCE and the HPAT-Ireland. Candidates must achieve a Leaving Certificate performance of at least 480 points, where an A1 or score of greater than 90% represents 100 points, an A2 or score of 85-89 represent 90 points, B1 or score of 80-84 represents 85 points.

In 2009 the maximum possible combined score was 860; 560 attributable to the LCE and 300 to the HPAT-Ireland (Central Applications Office, 2012b). The "LCE adjusted" is a moderated LCE score after a sliding scale has been applied to the raw LCE score. Above a cut-off of 550 points, every additional 5 points achieved on the LCE is awarded 1 point; so that a raw LCE score of 600 becomes 560 after adjustment.

Appendices

Chapter 5 Appendix 3 OSCE Details

Cronbach alphas are reported for Total OSCE, OSCE Clinical and OSCE Communication station items.

GY1 Total OSCE = .49 (95% CI=.32 - .62)

GY1 OSCE Clinical = .41 (95% CI = .21 - .56)

GY1 OSCE Communication = .59 (95% CI =.45 - .70)

GY2 Total OSCE = .67 (95% CI =.56 - .76)

GY2 OSCE Clinical= .70 (95% CI = 61 - .77)

GY2 Comm= .78 (95% CI = .72 - .83)

CY1 OSCE Total: .55 (95% CI=.39 - .67)

CY1 OSCE Clinical= .80 (95% CI = .75 - .85)

CY1 OSCE Communication = .65 (95% CI = .54 - .74)

CY2 OSCE Total: .50 (95% CI=.32 - .64)

CY2 OSCE Clinical= .81 (95% CI = .76 - .86)

CY2 OSCE Communication= .57 (95% CI = .43 - .68)

Chapter 5 Appendix 4: Result Section Notes

[i] There was a significant difference in the proportion of those sitting/ not-sitting the LCE and HPAT-Ireland exams between NUI GALWAY (e.g., % HPAT-Ireland: No-HPAT- Ireland: 49:51) and UCC (e.g., % HPAT- Ireland: No-HPAT- Ireland: 73:27).

[ii] Gender difference on HPAT- Ireland 2 approached significance: t (182) = -2.79, p = .006, d = -.41, with males having lower scores).

[iii] Some results are connected only with certain parts of the sample. For example, Galway Foundation year had no OSCE clinical station results provided, and therefore the OSCE Clinical results, refer to 1st Med Galway and 1st Med Cork only.

[iv] It should be noted that some researchers suggest a more conservative r = .3 minimum threshold for inclusion within regression analyses.

[v] A sensitivity analysis of the conclusions made to the presence of missing data was performed where each final model was refitted for each imputed data set and the results pooled over the five sets. The same predictors were identified for each analysis suggests that final conclusions are not sensitive to missing data.

Chapter 6 Appendix

Chapter 6 Appendix 1: Confirmation of copyright approval to reproduce paper as part of PhD

Appendices
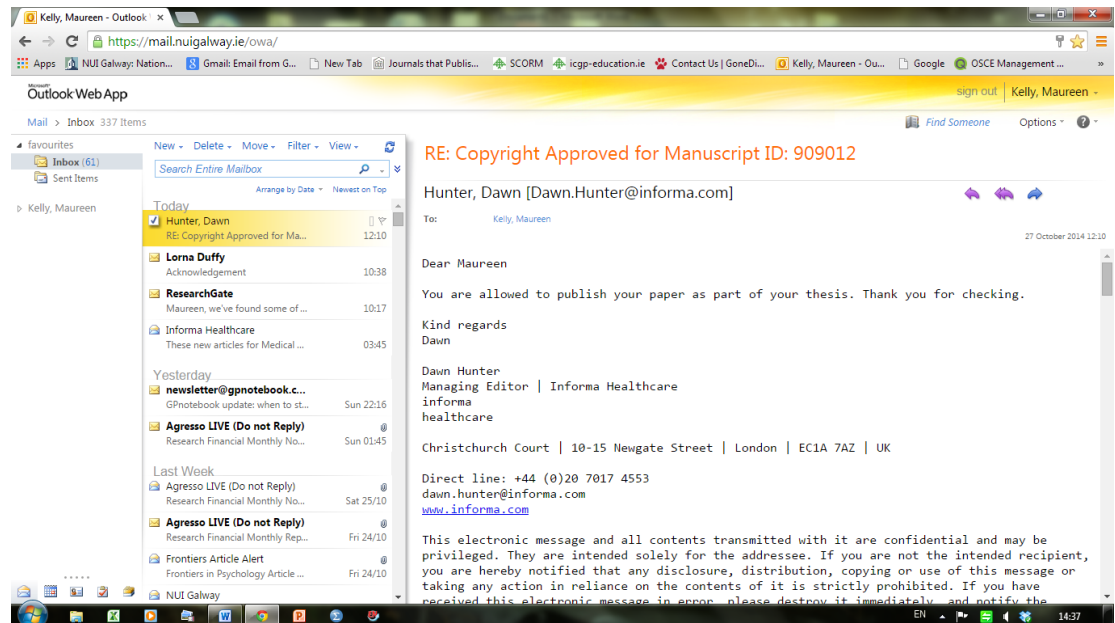
Chapter 6 Appendix 2 Topic Guide

Topic 1: Exploration of General Thoughts

Key questions: What are your thoughts having sat the HPAT-IRELAND?

Topic 2: Establishment of pre –test expectations and any changes to these expectations

Key questions: What were your expectations of the HPAT-IRELAND before sitting it?
Have your expectations changed having sat the HPAT-IRELAND? If so talk to me about this

Topic 3: Resonance with Clinical Practice

Key Questions: To what extent did the HPAT-IRELAND tap in to skills that you use in your practice as a doctor?   Can you expand on this?

Specific Probes: Can you give me an example of a skill that you used in the HPAT-IRELAND and where you would use that same skill in clinical practice?

Topic 4: Exploration of skills used in Section 1: 'logical reasoning'

Key Questions: What were your thoughts on this section?
 What skills do you think you used to answer this section?

Specific Probe: In what way, if at all, does this section correlate with your medical practice? (If not addressed already)

Topic 5: Exploration of skills used in Section 2 'Interpersonal Understanding'

Key Questions: What were your thoughts on this section?
 What skills do you think you used to answer this section?

Specific Probe: In what way, if at all, does this section correlate with your medical practice? (If not addressed already)

Topic 6: Exploration of skills used in Section 3:'Abtract Reasoning"

Key Questions: 1) What were your thoughts on this section?
 2)What skills do you think you used to answer this section?

Specific Probe: In what way, if at all, does this section correlate with your medical practice? (If not addressed already)

Topic 7: Exploration of the meaning and interpretation of the HPAT-IRELAND

Appendices

<u>Key questions:</u> What do you think the HPAT-IRELAND brings to medical student selection?

Can you tell me why you have this view?

What message do you think the HPAT-IRELAND sends out to medical school applicants and their families?

What do you think, is the value of the HPAT-IRELAND?

<u>Specific Probes:</u> What are your thoughts on the usefulness of the HPAT-IRELAND in predicting how well someone would perform as a doctor later on in life?

What are your thoughts on the usefulness of the HPAT-IRELAND in predicting how well someone would perform as a medical student?

Would you suggest any changes to the HPAT-IRELAND?

What are your thoughts on the influence the HPAT-IRELAND might have on student diversity?

What are your thoughts on whether or not the HPAT-IRELAND is specific to health professions?

Final Questions: Is there anything else you would like to say about the HPAT-IRELAND?

Is there anything I haven't asked you that you feel I should have?

Chapter 6 Appendix 3: Member Checking: Cover Letter

Dear _____

   Many thanks for participating in this study to date. The time that you gave to both sit the HPAT-IRELAND and do the interview was most appreciated. The analysis of the data from the post HPAT-IRELAND one-to-one interviews proved very interesting. In order to ensure that I have captured and interpreted the views of the study participants accurately I am "member checking" with approximately half of the group. To this end I have summarised what I understood from your interview and include it here for your verification. I am also attaching the original verbatim transcript from your audio taped interview.

In my summarised account of your interview the order of topics may vary slightly, for the purpose of clarity, from the way it is in the actual interview.   I ask you to please look over your original interview and then read through my two page summary to see if I have captured the essence of what you said at the time of the interview.

You are not being asked to offer any additional views or to go through the original interview in great detail. What I would like you to comment on is whether or not I have correctly summarised the overall sense of the interview, and whether my summary is a true reflection of what you thought and felt at that interview. Please also know that I am happy to be corrected should you need to do so.

You will see that your original interview is transcribed verbatim, including any repetitions, hesitations and grammatical errors that are usual in everyday spoken English. It is common place for some Doctors to find this unsettling or even slightly embarrassing to read over. However please be assured that this is normal for unprepared conversation and a regular feature of qualitative data. In the final write-up I will take care when using quotes to select them, or where appropriate edit them, so that they will read easier. The utmost care will be taken to retain the exact meaning of the phrase.

Please respond to me by email at maureen.kelly@nuigalway.ie

   Yours sincerely,

   Dr Maureen Kelly,

Chapter 7 Appendix

Chapter 7 Appendix 1: Confirmation of copyright approval to reproduce paper as part of PhD

Dr Maureen E Kelly,
Discipline of General Practice,
Clinical Science Institute,
NUI Galway

Dr. JFA Murphy, FRCPI,
Editor,
Irish Medical Journal,
IMJ Editorial Office,
IMO House,
10 Fitzwilliam Place,
Dublin 2

March 10th 2015

Dear Dr. Murphy,

I am completing a doctoral thesis at National University of Ireland, Galway and I am contacting you to request permission to include the following material within the electronic version of my PhD:

Can Multiple Mini Interviews work in an Irish setting? A feasibility study Kelly ME, Dowell J, Husbands A, Kropmans T, Jackson A, Dunne F, O' Flynn S, Newell J, Murphy AW. Irish Medical Journal 2014;107:7;210-212.

I would like to make an electronic version of my thesis available in ARAN which is the National University of Ireland, Galway's online repository http://aran.library.nuigalway.ie/. Once available in digital format, access to the thesis will be freely available via the Web. All users of the thesis will be bound by a Creative Commons Attribution Non-Commercial No Derivatives license [http://creativecommons.org/licenses/by-nc-nd/3.0/).

I would be grateful if you, or the company you represent, could grant me permission to include the Material in my thesis and to use the Material, as set out above, royalty free in perpetuity. Please be aware that the Material will be archived in ARAN in perpetuity.

I/We hereby grant permission for use of the Material for the purposes and on the terms identified above

(Signature) 30/3/15
Dr Colm F. A. Murph   Editor, I.M.I

[insert name of addressee] John FA Murphy, Editor

Date ...... 30/03/15 ....

Yours sincerely,

Dr. Maureen Kelly MICGP, FRCGP, M Med Ed.,

Lecturer in General Practice,

Discipline of General Practice,

College of Medicine, Nursing and Health Sciences,

National University of Ireland Galway.

Associate Programme Director,

Western Training Programme in General Practice

Tel: (353) 091 494107 (Monday and Tuesday)

Fax: (353) 091 494559

Email: maureen.kelly@nuigalway.ie

Webpage: www.nuigalway.ie/dgp/

Chapter 8 Appendix

Chapter 8 Appendix 1 MMI Description

There were ten MMI stations in total each lasting seven minutes, with one examiner per station. Station content was provided by Dundee Medical School.   Modifications to station content were made where necessary to ensure authenticity in an Irish setting. Five stations involved an interviewer, a role-player and the candidate. The other five stations were interview based (one interviewer: one candidate).

Each station was scored across three domains and one global rating scale.   Domain scores ranged from 0-5 (0= poor; 5 = excellent) with detailed written descriptors of excellent and poor performances. Global score were on a five point scale ranging from unacceptable to excellent performance.

Table 28 MMI Stations

	Station Content	Mapped to Irish Medical Council Eight Domains of Professional Practice
1	Counselling conversation with a distressed medical student	**Relating to patients** Management
2	Discussion of possible breaches of professionalism by a medical student	**Professionalism** Communication and Interpersonal skills
3	Conversation with a medical student who was drinking alcohol to excess	**Relating to patients** Professionalism
4	Interview about motivation and preparation to study medicine	**Scholarship** Communication and Interpersonal skills
5	Candidate and helper complete a puzzle station	**Collaboration and Teamwork** Relating to patients
6	Interview about the experience of team work, achievements and social responsibility	**Scholarship** Communication and Interpersonal skills
7	Discussion about what makes a fair health care system	**Patient safety and quality of patient care** Professionalism
8	Giving advice to a non-English speaking patient	**Communication and Interpersonal skills** Relating to patients
9	Discussion about the issues relating to organ donation	**Professionalism** Communication and Interpersonal skills
10	Completing a complex card sort with a helper	**Management** Collaboration and teamwork

Chapter 8 Appendix 2

Description of First Year Medical Course/ Outcome Variables

First Year Modules and Assessments

The medical undergraduate degree programme is an integrated modular, systems based curriculum. The First Year Medical Course comprises 10 modules which together equal 60 ECTS. The modules and ECTS credit weighting are listed below. In general knowledge based assessments (e.g. MCQs/ written papers) account for 70% of the mark per module, while practical / skills based assessments (e.g. laboratory exams/OSCEs) account for 30% of the mark.

Semester 1

Musculoskeletal System and Peripheral Nerves 10 ECTS

General Principles of Human Body Structure 5 ECTS

Principles of Physiology 5 ECTS

Biomolecules, Metabolism and Energy 5 ECTS

Medical Professionalism (1) 10 ECTS (This module includes ethics, law, health promotion, health informatics, communication and clinical examination skills and a special study module).

Semester 2

Cardiovascular System 5 ECTS

Respiratory System 5 ECTS

Renal System 5 ECTS

Gastrointestinal System 5 ECTS

Metabolism, Nutrition and Health 5 ECTS.

Details of the outcome variables:

First Med Score was a continuous variable representing each student's percentage score overall across all first year modules. It is calculated based on their percentage mark out of 100 for each of ten modules and is adjusted to reflect the relative weightings of each of the module in terms of its allocated credits according to the formula: Sum of (Percentage mark* credit weighting of module) / total number of credits).

First Med OSCE score, a continuous variable representing performance on a five station OSCE that contributes (one third) to the professionalism score. The OSCE stations examine communication and clinical skills and are of 5 minute duration. The OSCE is conducted in

two parallel cycles, with one examiner and one simulated patient per station, per cycle. OMIS software is used to mark the OSCE stations online (Kropmans et al., 2011). OSCE marking is based on a combination of a checklist and clinical / communication descriptor bands. The Calgary Cambridge Model forms the basis of the communication skills checklist (Silverman et al., 2005).

Example OSCE stations are

Asking a patient for consent to measure vital signs, measuring them and explaining the findings to the patient.

Performing urinalysis on a sample of urine and explaining the findings to the patient.

Measuring a patient's blood pressure and discussing the results with the patient.

Calculating a patient's BMI and discussing results.

Performing Basic Life Support and using an automated external defibrillator on a mannequin

Appendices

Chapter 8 Appendix 3 FOCUS GROUP TOPIC GUIDE

Topic 1: Exploration of General Thoughts

<u>Key questions:</u> What are your thoughts having examined at / sat the MMI?

<u>Specific probes:</u> From your experience how do you think MMI compares to other tools used for the selection of medical students?

Topic 2: Potential role for MMI in the selection of medical students in Ireland

<u>Key questions:</u> Do you think the MMI has a potential role in the selection of medical students in Ireland?

<u>Specific probes:</u> What do think the barriers to its use might be? What advantages might it offer over current selection tools? Do you have any suggestions that might make MMI a viable option for selection in Ireland?

Topic 3: Establishment of general views on the use of MMI for selecting students from international backgrounds

<u>Key Questions:</u> What are your views on the utility / appropriateness of MMI as a selection tool for medical students from international backgrounds?

Can you expand on this? / Do you see any advantages to using MMI in this context? / Do you see any barriers to using MMI in this context?

How do you think MMI would compare to traditional interview in the selection of international students?

Topic 4: Impact of English language proficiency on MMI performance (If not already explored in Topic 2)

<u>Key Questions:</u> To what extent do you think English language proficiency can impact on a candidates' performance in the MMI? Can you expand on this?

Topic 5: Impact of cultural issues on MMI performance (If not already explored in Topic 2)

<u>Key Questions:</u> To what extent do you think cultural issues could impact on a candidates' performance in the MMI? Can you expand on this?

(If the interviewee is unsure of what is meant by this then prompt with- for instance one of the MMI stations referred to a medical student drinking alcohol to excess- this may be a particular issue for Irish medical students and as such is quite culturally specific).

Topic 6: Recommendations for best practice

<u>Key Questions:</u> Can you suggest any recommendations for designing and running MMIs that would be appropriate to use in the selection of medical students from international background?

<u>Specific Probes:</u>

In the development of station content? / In the running of the MMI? /

Final Questions: Is there anything else you would like to say about the MMI?

Is there anything I haven't asked you that you feel I should have?

Chapter 9 Appendix

Chapter 9 Appendix 1: Study 4 Examining MMI for Evidence of Differential prediction

Three steps were conducted as recommended by Kyei-Blankson (2005). I regressed the OSCE and average weighted First Year Medical assessment scores on MMI for the total sample. Then I generated a prediction equation for a) MMI and OSCE b) MMI and average weighted First Med assessment score according to the formula: $y=ax+c$, where y= predicted examination score, a = regression coefficient, x the observed MMI score and c the constant generated from the linear regression intercept. Residuals were calculated by subtracting the predicted value from the observed value for each student in the various subgroups. Mean residuals were calculated by getting the average of the differences. Negative values indicate that the MMI over-predicted, and positive values that it under-predicted.

Results

Table 29 presents the results of analysis of the data from Study 4 with respect to the MMI. Figures 9 and 10 depict the plots of the residuals for the predicted overall First Year weighted average score and OSCE. There was no evidence of under-prediction of average weighted First Year Medical Scores for Non-EU students or students without English as a first language, rather the data suggest that the MMI may have slightly over-predicted for these students. Whereas students from Socioeconomic groups 3, 4 & 5 appear to have been under-predicted. Findings with respect to the prediction of OSCE scores are presented here too for completeness. The figures suggest that MMI may have under-predicted OSCE scores for Non-EU students, students without English as a first language and female students. However, examining the plot of residuals for each of the sub-groups does not suggest that there was any pattern of under or over predication for any one group. Additionally the only significant difference in mean residuals lay between males and females for OSCE performance.

Discussion

Evidence for predictive bias is an important consideration in the overall fairness of a selection tool. In spite of the significant difference in mean MMI scores between Non-EU and EU students and students without English as a first language and those with English,

there is no evidence of under-prediction of average weighted First Year Medical Scores for either group based on these data. It is important to note that this analysis has a number of important limitations. The total sample size is small, the sample size of sub-groups even smaller, the overall correlations between MMI and OSCE and First Year Medical score are very weak and not statistically significant. Hence these results have to be interpreted very cautiously and further research would be required before confident statements could be made.

Table 29 Evidence of Differential Prediction

Group	Observed minus predicted OSCE scores	Mean residual observed minus predicted OSCE^	Student t test for difference in mean/ median residual OSCE Score between groups	Observed minus predicted average weighted First Year Med Scores	Mean residual observed minus predicted average weighted First Year Med Scores ^	Student t test for difference in mean/median residual First Year Med Score between groups
Non EU Students	21 over-predicted 24 under-predicted	0.64	t=1.11, df 107, p=0.27	28 over-predicted 17 under-predicted	-0.64	t=-0.675, df 107, p=0.501
EU Students	29 over-predicted 36 under-predicted	-0.45		32 over-predicted 32 under-predicted	0.42	
Without English as first language	19 over-predicted 20 under-predicted	0.53	t=0.82, df 107, p=0.42	26 over-predicted 13 under-predicted	-1.5	t=-1.54, df 94, p=0.13
With English as first language	30 over-predicted 40 under-predicted	-0.29		34 over-predicted 36 under-predicted	0.81	
Female	26 over-predicted 42 under-predicted	1.1	**t=2.82, df 65.32, p=0.006**	36 over-predicted 32 under-predicted	-0.01	t=0.01, df 107, p=0.99
Male	23 over-predicted 18 under-predicted	-1.8		24 over-predicted 17 under-predicted	-0.02	
Socioeconomic group 3,4,5	12 over-predicted 12 under-predicted	-0.31	Independent samples compare median test p=0.86	11 over-predicted 13 under-predicted	0.44	Independent samples compare median test p= 0.45
Socioeconomic group 1,2	35 over-predicted 46 under-predicted	0.23		47 over-predicted 34 under-predicted	-0.07	

Footnote ^Negative residual indicates over-prediction; positive residual indicates under-prediction. Significant results highlighted in bold.
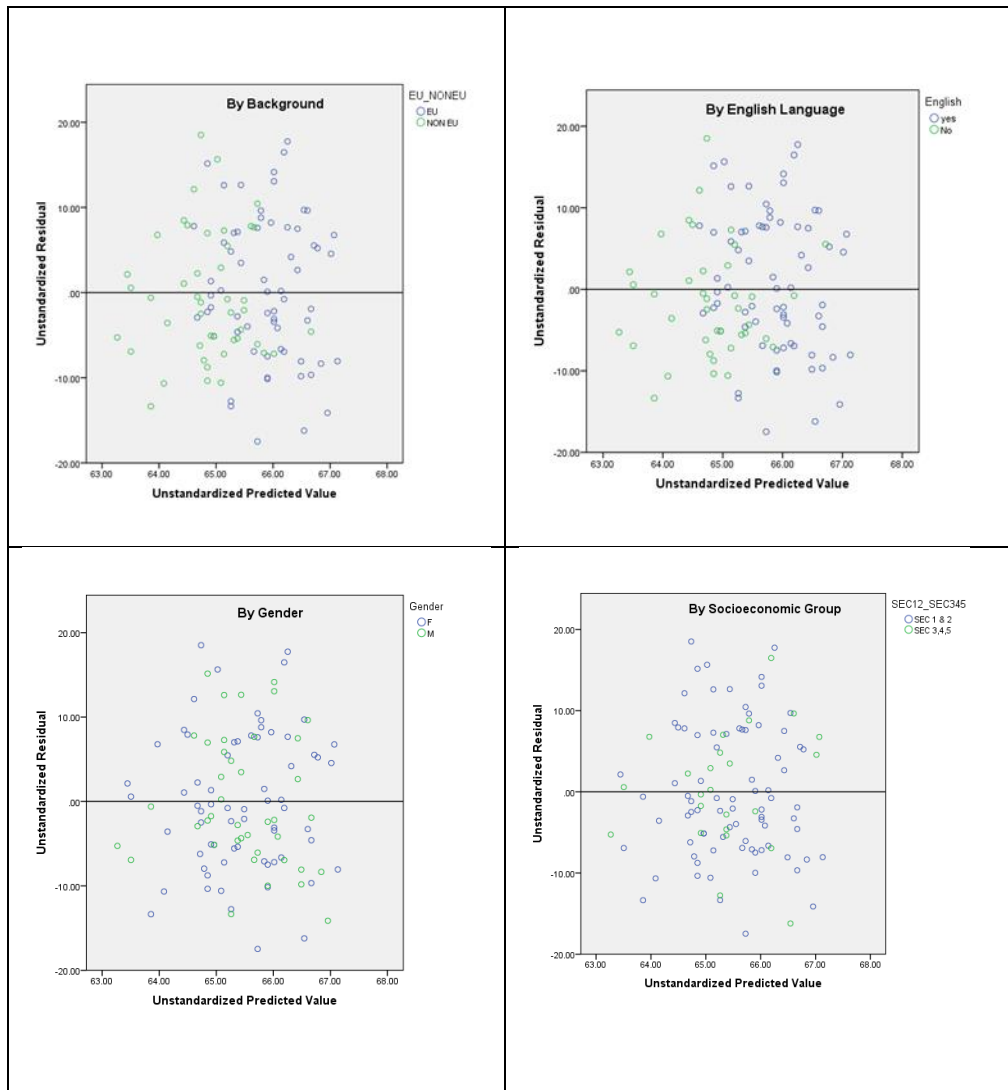
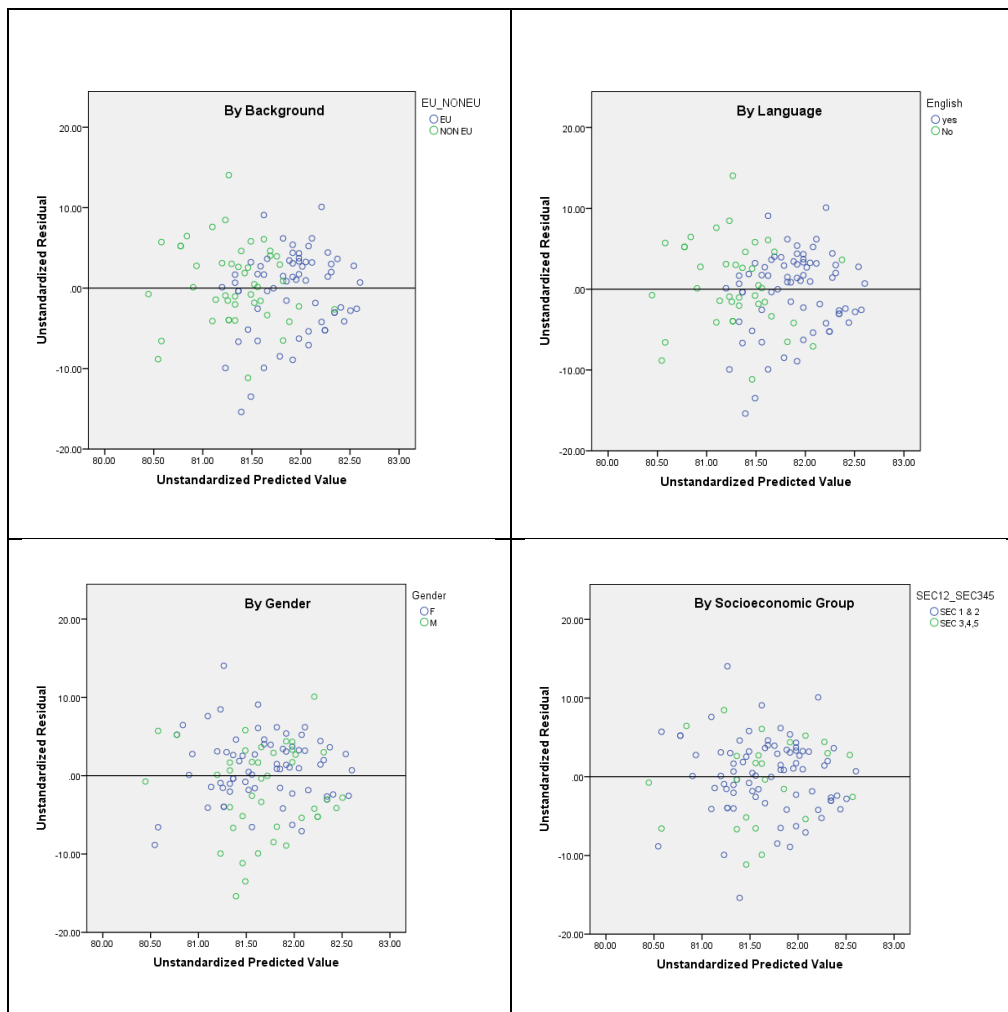**Figure 10 Plot of the Residuals First Med Assessments by Group**

Figure 11 Plot of the Residuals OSCE by Group

References

Aamodt, M. 2012. Industrial/Organizational psychology: An applied approach. California, USA: Cengage Learning.

Adam, J., Bore, M., McKendree, J., Munro, D. & Powis, D. 2012. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Medical Education,* 12**,** 69.

Adam, J., Dowell, J. & Greatrix, R. 2011. Use of UKCAT scores in student selection by UK medical schools, 2006-2010. *BMC Medical Education,* 11, 98.

Adams, J. 2009. The acceptability of online courses as criteria for admission to medical school. *Ochsner Journal,* 9**,** 4-10.

Admissions Testing Service. 2014. *About the BioMedical Admissions Test* [Online]. Cambridge, UK: Admissions Testing Service,. Available: http://www.admissionstestingservice.org/for-test-takers/bmat/about-bmat/ [Accessed July 8th 2014].

Agrawal, J. R., Vlaicu, S. & Carrasquillo, O. 2005. Progress and pitfalls in underrepresented minority recruitment: Perspectives from the medical schools. *Journal of the National Medical Association,* 97**,** 1226-1231.

Ahmed, A., Qayed, K. I., Abdulrahman, M., Tavares, W. & Rosenfeld, J. 2014. The multiple mini-interview for selecting medical residents: First experience in the Middle East region. *Medical Teacher,* 36**,** 703-709.

Albanese, M. A., Snow, M. H., Skochelak, S. E., Huggett, K. N. & Farrell, P. M. 2003. Assessing personal qualities in medical school admissions. *Academic Medicine,* 78**,** 313-321.

Ambrose, M. & Arnaud, A. 2005 Are procedural justice and distributive justice conceptually different? . *In:* Greenberg, J. & Colquitt, J. A. (eds.) *Handbook of Organisational Justice* New York: Psychology Press

American Board of Medical Specialities. 2014 *Standards for the ABMS Program for Maintenance of Certification* [Online]. Available: http://www.abms.org/media/1109/standards-for-the-abms-program-for-moc-final.pdf [Accessed 12/02 2015].

American Educational Research Association, American Psychological Association & National Council on Measurement in Education 1999. Standards for Educational and Psychological Testing Washington DC, USA: American Education Research Association

American Medical Association. 2014 *Opinion 9.121 - Racial and Ethnic Health Care Disparities* [Online]. USA: American Medical Association. Available: http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion9121.page [Accessed 12/12/ 2014].

American Medical Student Association. 2014. *Enriching Medicine Through Diversity* [Online]. USA. Available: http://www.amsa.org/AMSA/Homepage/About/Priorities/Diversity.aspx [Accessed 10/03/2015

Andriole, D. A. & Jeffe, D. B. 2010. Prematriculation variables associated with suboptimal outcomes for the 1994-1999 cohort of US medical school matriculants. *Journal of the American Medical Association,* 304**,** 1212-1220.

Archer, L., Hutchings, M. & Ross, A. 2005. *Higher education and social class: issues of exclusion and inclusion,* London Routledge Falmer

References

Association of American Medical Colleges. 2012. *Medical College Admission Test* [Online]. http://www.aamc.org/mcat. Available: https://www.aamc.org/students/applying/mcat/ [Accessed 17/02/ 2012].

Association of American Medical Colleges 2013 MCAT and Medical School Admissions: Research to Support the Design of the MCAT2015 Exam.

Association of American Medical Colleges. 2014. *Medical College Admission Test* [Online]. Washington: The Association of American Medical Colleges. Available: https://www.aamc.org/students/applying/mcat/ [Accessed July 6th 2014].

Australian Council for Educational Research (ACER). 2007. *UMAT A validity Study* [Online]. Available: http://umat.acer-edu.au/umat-research [Accessed January 27th 2014].

Australian Council for Educational Research (ACER). 2012a. *About us* [Online]. ACER. Available: http://www.acer.edu.au/ [Accessed 26/10/ 2012].

Australian Council for Educational Research (ACER). 2012b. *Undergraduate Medicine and Health Sciences Admission Test* [Online]. http://umat.acer.edu.au/. Available: http://umat.acer.edu.au/ [Accessed 17/02/ 2012].

Australian Council for Educational Research (ACER). 2012c. *Health Professions Admission Test - Ireland* [Online]. ACER. Available: http://www.hpat-ireland.acer.edu.au/ [Accessed 17/02/ 2012].

Australian Council for Educational Research (ACER). 2013a. *Graduate Australian Medical School Admissions Test Information Session* [Online]. ACER Available: http://www.flinders.edu.au/medicine/fms/sites/gemp_promo/GAMSAT%20Information%20Session%20(2013).pdf [Accessed 14/12/ 2014].

Australian Council for Educational Research (ACER). 2013b. *Health Professions Admission Test-Ireland (HPAT-Ireland)* [Online]. Available: http://www.hpat-ireland.acer.edu.au/ [Accessed October 1 2013].

Australian Council for Educational Research (ACER). 2014a. *HPAT-Ireland- Health Professions Admission Test Ireland* [Online]. Available: http://www.hpat-ireland.acer.edu.au/ [Accessed June 16th 2014].

Australian Council for Educational Research (ACER). 2014b. *Graduate Medical School Admissions Test* [Online]. Available: http://www.gamsat-ie.org/ [Accessed July 9th 2014].

Australian Council for Educational Research (ACER). 2014c. *Undergraduate Medicine and Health Sciences Admission Test* [Online]. Available: https://umat.acer.edu.au/ [Accessed July 9th 2014].

Axelson, R. D. & Kreiter, C. D. 2009. Rater and occasion impacts on the reliability of pre-admission assessments. *Medical Education,* 43**,** 1198-1202.

Bardes, C. L., Best, P. C., Kremer, S. J. & Dienstag, J. L. 2009. Perspective: Medical school admissions and noncognitive testing: Some open questions. *Academic Medicine,* 84**,** 1360-1363.

Barlett, F. C. 1946. Selection of Medical Students *British Medical Journal,* 2**,** 2, 665-666.

Bauer, T., Truxillo, D., Sanchez, R., Craig, J., Ferrara, P. & Campion, M. 2001. Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology,* 54**,** 387-419.

Beckett, J. 2008. *The University Admission Process is Still Deeply Flawed* [Online]. The Gaurdian Education (Internet), 13 Feb. Available:

http://www.thegaurdian.com/education/2008/feb/13/highereducation.n
ews [Accessed October 1 2013].

Benbassat, J. & Baumal, R. 2007. Uncertainties in the Selection of Applicants for
Medical School. *Advances in Health Sciences Education,* 12**,** 509-521.

Bielenberg, K. 2009. Barred from Medical School – Why Grade A rejects feel sick
Independent.ie, August 22. Available:
http://www.independent.ie/lifestyle/barred-from-medical-school-why-
gradea-rejects-feel-sick-26560351.html [Accessed June 24 2014]

Bies, R. & Moag, J. 1986. Interactional justice: Communication criteria of fairness.
Research on negotiation in organizations, 1, 43-55.

Bodger, O., Byrne, A., Evans, P. A., Rees, S., Jones, G., Cowell, C., Gravenor, M. B.
& Williams, R. 2011. Graduate entry medicine: selection criteria and
student performance. *PLoS ONE,* 6**,** e27161.

Bore, M., Munro, D., Kerridge, I. & Powis, D. 2005. Selection of medical students
according to their moral orientation. *Medical Education,* 39**,** 266-275.

Bore, M., Munro, D. & Powis, D. 2009. A comprehensive model for the selection of
medical students. *Medical Teacher,* 31**,** 1066-72.

Boulet, J. R., Cooper, R. A., Seeling, S. S., Norcini, J. J. & Mckinley, D. W. 2009. US
citizens who obtain their medical degrees abroad: An overview, 1992–
2006. *Health Affairs,* 28**,** 226-233.

Boulet, J. R., Swanson, D. B., Cooper, R. A., Norcini, J. & Mckinley, D. 2006. A
comparison of the characteristics and examination performances of US
and Non-US citizen international medical graduates who sought
educational commission for foreign medical graduate certificates: 1995-
2004. *Academic Medicine,* 81, S116-S119.

Brant Toni [@Tonibrant]. 2014. *Goodbye medicine #hpat -[Tweet].* Available:
https://twitter.com/search?src=typd&q=%23hpat [Accessed June 23
2014].

Brennan, N. & Mattick, K. 2013. A systematic review of educational interventions
to change behaviour of prescribers in hospital settings, with a particular
emphasis on new prescribers. *British Journal of Clinical Pharmacology,* 75**,**
359-72.

British Medical Association 2009. Equality and diversity in UK medical schools.
London. British Medical Association

Broquet, K. E. & Rockey, P. H. 2004. Teaching Residents and Program Directors
about Physician Impairment. *Academic Psychiatry,* 28**,** 221-225.

Brown, C. A. & Lilford, R. J. 2008. Selecting medical students. *British Medical
Journal,* 336**,** 786.

Brown, J. & Griffin, B. 2012. Attitudes to Medical Student Selection - analysis of
key stakeholders' attitudes and perceptions of student selection *ASME
Annual Scientific Meeting - Delivering a high quality learning experience in
changing times* Brighton, UK. 18-20[th]July. Association for the Study of
Medical Education

Brownell, K., Lockyer, J., Collin, T. & Lemay, J. F. 2007. Introduction of the multiple
mini interview into the admissions process at the University of Calgary:
Acceptability and feasibility. *Medical Teacher,* 29**,** 394-396.

Buttimer, J. 2006. *Preparing Ireland's Doctors to Meet the Health Needs of the
21st Century: Report of the Postgraduate Medical Education and Training
Group,* Dublin, Ireland, Department of Health and Children.

Byrne, A. 2000. Cold comfort on medical points race. *The Irish Times* Jan 20, 2000,
p.4.

References

Callahan, C. A., Hojat, M. & Gonnella, J. S. 2007. Volunteer bias in medical education research: an empirical study of over three decades of longitudinal data. *Medical Education,* 41**,** 746-753.

Callahan, C. A., Hojat, M., Veloski, J., Erdmann, J. B. & Gonnella, J. S. 2010. The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: A longitudinal study of 36 classes of jefferson medical college. *Academic Medicine,* 85**,** 980-987.

Cambridge Assessment Group. 2008. *BioMedical Admission Test.* Available: http://www.admissionstests.cambridgeassessment.org.uk/adt/bmat [Accessed 23/02/ 2012].

Campagna-Vaillancourt, M., Manoukian, J., Razack, S. & Nguyen, L. H. 2014. Acceptability and reliability of multiple mini interviews for admission to otolaryngology residency. *Laryngoscope,* 124**,** 91-6.

Carlson, J. A. 2010. Avoiding traps in member checking. *The Qualitative Report,* 15**,** 1102-1113.

Carrasquillo, O. & Lee-Rey, E. T. 2008. Diversifying the medical classroom: Is more evidence needed? *Journal of the American Medical Association,* 300**,** 1203-1205.

Cassidy, J. 2008. UKCAT among the pigeons. *British Medical Journal* 336**,** 691-692.

Central Applications Office. 2012a. *CAO Board of Directors Report 2010* [Online]. Dublin: CAO Available: http://www2.cao.ie/dir_report/pdf/caoreport2010.pdf [Accessed 29/03/ 2012].

Central Applications Office. 2012b. *Selection criteria for Undergraduate Entry to Medicine For EU Applicants 2012* [Online]. Available: http://www2.cao.ie/downloads/documents/2012UGMedEntry.pdf [Accessed 02/06/ 2012].

Central Applications Office. 2014 *Guideline Entry Requirements for EU/EFTA Applicants (other than Leaving Certificate)* [Online]. Available: http://www2cao.ie/downloads/documents/Guidelines-EU-EFTA.pdf [Accessed June 16th 2014].

Central Applications Office (CAO). 2014 *Selection Criteria for Undergraduate Entry to Medicine EU Applicants* [Online]. Available: http://www2cao.ie/downloads/documents/UGMedEntry2014.pdf.

Central Statistics Office (CSO). 2011. *Census:This is Ireland Part 2* [Online]. Central Statistics Office Available: http://www.cso.ie/en/census2011reports/census2011thisisirelandpart2/ [Accessed June 26th 2014].

Chambers, B. A. 2002. Applicant reactions and their consequences: Review, advice, and recommendations for future research. *International Journal of Management Reviews,* 4**,** 317-333.

Christakis, P. G., Christakis, T. J., Dziura, J. & Christakis, J. T. 2010. Role of the interview in admissions at the University of Toronto ophthalmology program. *Canadian Journal of Ophthalmology,* 45**,** 527-30.

Citizens Information Board. 2015 *Third-level student fees and charges* [Online]. Available: http://www.citizensinformation.ie/en/ [Accessed 23/03 2015].

Clarkson, M. 2007. Hanafin proposals on Medical School Entry, Letter to the Editor *The Irish Times* June 9th, p.17.

References

Cleland, J., Dowell, J., Mclachlan, J., Nicholson, S. & Patterson, F. 2012. Identifying best practice in the selection of medical students. London, United Kingdom. General Medical Council.

Cleland, J., French, F. & Johnston, P. 2011. A mixed-methods study identifying and exploring medical students' views of the UKCAT. *Medical Teacher,* 33**,** 244-249.

Coates, H. 2008. Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT). *Medical Education,* 42**,** 999-1006.

Cohen-Charash, Y. & Spector, P. E. 2001. The Role of Justice in Organizations: A Meta-Analysis. *Organizational Behavior and Human Decision Processes,* 86**,** 278-321.

Cohen, J. 1992. A power primer. *Psychological Bulletin,* 112**,** 155-9.

Collin, T. V., Violato, C. & Hecker, K. 2009. Aptitude, Achievement and Competence in Medicine: A Latent Variable Path Model. *Advances in Health Sciences Education,* 14**,** 355-366.

Collins, M. 2009. Admissions test helps to level the playing field. *Irish Examiner*, August 27. Available: http://www.irishexaminer.com/viewpoints/yourview/admissions-test-helps-to-level-the-playing-field-99518.html [Accessed March 2nd 2012]

Colquitt, J. A. 2001. On the dimensionality of organizational justice: a construct validation of a measure. *Journal of Applied Psychology,* 86**,** 386.

Colquitt, J. A., Greenberg, J. & Zapata-Phelan, C. P. 2005. What is organizational justice? A historical overview *In:* Greenberg , J. & Colquitt, J. A. (eds.) *Handbook of Organisational Justice* New York Psychology Press

Columbia University. 2015 *Latent Variable Modeling* [Online]. New York, USA Columbia University Available: http://www.columbia.edu/~mmw2177/LVcourse/slides1handout.pdf [Accessed 01/04/2015 2015].

Cook, D. A. & Beckman, T. J. 2006. Current concepts in validity and reliability for psychometric instruments: theory and application. *The American Journal of Medicine,* 119**,** 166. e7.

Corbin, J. & Strauss, A. 2008. *Basics of Qualitativ Research 3e- Techniques and Procedures for Developing Grounded Theory* Thousand Oaks, California, Sage Publications Inc

Cotter, K. 2009. Gender Balance in Medicine. *The Irish Times*, Aug 25th. Available: http://www.irishtimes.com/opinion/letters/gender-balance-in-medicine-1.725369 [Accessed October 12, 2013]

Coyne, I. & Cowley, S. 2006. Using grounded theory to research parent participation. *Journal of Research in Nursing,* 11**,** 501-515.

Coyne, I. T. 1997. Sampling in qualitative research. Purposeful and theoretical sampling; merging or clear boundaries? *Journal of Advanced Nursing,* 26**,** 623-630.

Cresswell, A. 2011. *Selection exam shown to be poor guide to university performance* [Online]. The Australian (Internet). Available: http://www.theaustralian.com.au/news/nation/selection-exm-shown-to-be-poor-guide-to-university-performance/story-e6frg6nf-1226032925914 [Accessed October 1 2013].

Creswell, J. W. 2007. *Qualitative inquiry & research design: Choosing among five approaches,* Thousand Oaks, CA, Sage Publications, Inc.

Creswell, J. W. 2013. *Qualitative inquiry and research design : choosing among five approaches,* Los Angeles, Los Angeles : Sage Publications.

281

References

Creswell, J. W. & Clark, V. L. P. 2011. *Designing and conducting mixed methods research,* Thousand Oaks CA, Sage Publications, Inc.

Cropanzana, R., Bowen, D. E. & Gilliland, S. W. 2007. The management of organizational justice. *The Academy of Management Perspectives*, Nov; 34-48.

Crossley, J., Humphris, G. & Jolly, B. 2002. Assessing health professionals. *Medical Education,* 36, 800-804.

Crotty , M. 1998. *The foundations of social research: Meaning and perspective in the research process.,* London, Sage.

Cushing/Whitney Medical Library. 2010. *Medicine at Yale, 1910 - 1960* [Online]. CT, USA Medical Historical Library. Available: http://doc.med.yale.edu/historical/bicentennial/1910/winternitz.html [Accessed 21/06/ 2014].

Daram, S. R., Wu, R. & Tang, S. J. 2014. Interview from anywhere: feasibility and utility of web-based videoconference interviews in the gastroenterology fellowship selection process. *American Journal of Gastroenterology,* 109, 155-9.

Davis, D., Dorsey, J. K., Franks, R. D., Sackett, P. R., Searcy, C. A. & Zhao, X. 2013. Do racial and ethnic group differences in performance on the MCAT exam reflect test bias? *Academic Medicine,* 88, 593-602.

De Cremer, D., Stinglhamber, F. & Eisenberger, R. 2005. Effects of Own versus Other's Fair Treatment on Positive Emotions: Field Study. *The Journal of Social Psychology,* 145, 741-744.

Dempsey. 2003. Taking the heat out of points pressure: Education Minister, Noel Dempsey, explains why he favours changes in entry to health science courses. *The Irish Times* August 26th, p.4.

Dennehy, T., Kelly, M. & O'Flynn, S. 2013 General Practitioners' Perspectives on Revised Entry and Selection Methods to Medicine and the HPAT. *Irish Medical Journal,* 106, 113-115.

Department of Education and Skills. 2006. *Government Announces €200 million package of reforms of Medical Education and Training* [Online]. Dublin, Ireland. Available: http://www.education.ie/en/Press-Events/Press-Releases/2006-Press-Releases/PR06-02-01.html [Accessed 20/08 2012].

Department of Education and Skills. 2009. *Details of new selection criteria for undergraduate medical education goes out to all post primary schools* [Online]. Dublin, Ireland: Department of Education and Skills. Available: http://www.education.ie/en/Press-Events/Press-Releases/2008-Press-Releases/PR08-05-19.html [Accessed June 22/ 2014 2014].

Dhar, D., Perry, W. R. G. & Poole, P. 2012. Students' perceptions of the Undergraduate Medicine and Health Sciences Admissions Test (UMAT). *New Zealand Medical Journal,* 125, 29-36.

Dodson, M., Crotty, B., Prideaux, D., Carne, R., Ward, A. & De Leeuw, E. 2009. The multiple mini-interview: How long is long enough? *Medical Education,* 43, 168-174.

Donnellan, E. 2006. Aptitude Test Likely for aspiring medical students *The Irish Times*, Feb 2. Available: http://www.irishtimes.com/news/aptitude-test-likely-for-aspiring-medical-students-1.1009685 [Accessed October 5/2013]

Donnelly, K. & Heffernan, B. 2011. Medicine points still out of reach as aptitude test deemed a failure. *Independent.ie*, Monday August 22. Available: http://www.independent.ie/lifestyle/education/latest-news/medicine-

References

points-still-out-of-reach-as-aptitude-test-deemed-a-failure-2854097.html [Accessed October 1/ 2013]

Donnelly, R. 2006. Graduate-entry medicine in the UK: GAMSAT-based selection and learning outcomes *International Workshop on Aptitude Tests for Admission to Medical School.* Tokyo.

Donnon, T., Paolucci, E. O. & Violato, C. 2007. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine,* 82**,** 100-106.

Dore, K. L., Reiter, H. I., Eva, K. W., Krueger, S., Scriven, E., Siu, E., Hilsden, S., Thomas, J. & Norman, G. R. 2009. Extending the interview to all medical school candidates-computer-based multiple sample evaluation of noncognitive skills (CMSENS). *Academic Medicine,* 84**,** S9-S12.

Dore, K. L., Kreuger, S., Ladhani, M., Rolfson, D., Kurtz, D., Kulasegaram, K., Cullimore, A. J., Norman, G. R., Eva, K. W., Bates, S. & Reiter, H. I. 2010. The Reliability and Acceptability of the Multiple Mini-Interview as a Selection Instrument for Postgraduate Admissions. *Academic Medicine,* 85**,** S60-S63.

Dowell, J., Lynch, B., Till, H., Kumwenda, B. & Husbands, A. 2012. The multiple mini-interview in the U.K. context: 3 years of experience at Dundee. *Medical Teacher,* 34**,** 297-304.

Downing, S. M. 2003. Validity: on meaningful interpretation of assessment data. *Medical Education,* 37**,** 830-7.

Downing, S. M. 2004. Reliability: on the reproducibility of assessment data. *Medical Education,* 38**,** 1006-1012.

Dunleavy, D. M., Kroopnick, M. H., Dowd, K. W., Searcy, C. A. & Zhao, X. 2013. The predictive validity of the MCAT exam in relation to academic performance through medical school: A National Cohort Study of 2001-2004 matriculants. *Academic Medicine,* 88**,** 666-671.

Dunne, J. 2007. Hanafin proposals on medical school entry- Letter to the Editor. *The Irish Times* June 9th, p.17.

Edwards, D., Friedman, T. & Pearce, J. 2013. Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Medical Education,* 13**,** 173.

El Says, F., Ayuob, N., Fahmy, A. R., El Fayez, F., Hasanian, M. & El Deek, B. 2013. Experience of Establishment of Multiple Mini structure Interview as part of student admission policy at Faculty of Medicine, King Abdulaziz University, 2011-2012. *Medical Teacher,* 35**,** S74-S77.

Emery, J. L. & Bell, J. F. 2009. The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance. *Medical Education,* 43**,** 557-564.

Emery, J. L., Bell, J. F. & Vidal Rodeiro, C. L. 2011. The BioMedical admissions test for medical student selection: Issues of fairness and bias. *Medical Teacher,* 33**,** 62-71.

Erdogan, B. 2003. Antecedents and consequences of justice perceptions in performance appraisals. *Human Resource Management Review,* 12**,** 555-578.

Esmail, A. & Roberts, C. 2013. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. *British Medical Journal,* 347,f5662.

References

Eva, K. W. & Macala, C. 2014. Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education,* 48**,** 604-13.

Eva, K. W. & Reiter, H. I. 2004. Where Judgement Fails: Pitfalls in the Selection Process for Medical Personnel. *Advances in Health Sciences Education, 9***,** 161-174.

Eva, K. W., Rosenfeld, J., Reiter, H. I. & Norman, G. R. 2004a. An admissions OSCE: The multiple mini-interview. *Medical Education,* 38**,** 314-326.

Eva, K. W., Reiter, H. I., Rosenfeld, J. & Norman, G. R. 2004b. The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Academic Medicine,* 79**,** 602-609.

Eva, K. W., Reiter, H. I., Rosenfeld, J. & Norman, G. R. 2004c. The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine,* 79**,** S40-S42.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J. & Norman, G. R. 2009. Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education,* 43**,** 767-775.

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 (JAMA),* 308**,** 2233-2240.

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 39**,** 175-191.

Feast, V. 2002. The impact of IELTS scores on performance at University. *International Education Journal,* 3**,** 70-85.

Ferguson, E., James, D. & Madeley, L. 2002. Factors associated with success in medical school: systematic review of the literature. *British Medical Journal,* 324**,** 952-7.

Fernando, N., Prescott, G., Cleland, J., Greaves, K. & Mckenzie, H. 2009. A comparison of the United Kingdom clinical aptitude test (UK-CAT) with a traditional admission selection process. *Medical Teacher,* 31**,** 1018-23.

Finlayson, H. C. & Townson, A. F. 2011. Resident selection for a physical medicine and rehabilitation program: Feasibility and reliability of the multiple mini-interview. *American Journal of Physical Medicine and Rehabilitation,* 90**,** 330-335.

Flexner, A. 1910. Medical Education in the United States and Canada Bulletin Number 4 (the Flexner report) New York . The Carnegie Foundation.

Flynn, S. 2007. Reforms will make it easier to study medicine. *Irish Times* Dec 17

Flynn, S. 2008. New entry criteria for medicine courses unveiled. *The Irish Times* Tuesday May 20th.

Fodchuk, K. M. & Sidebotham, E. J. 2005. Procedural Justice in the Selection Process: A Review of Research and Suggestions for Practical Applications. *The Psychologist-Manager Journal,* 8,105-120.

Fottrell, P. 2006. *Medical education in Ireland: A new direction. Report of the Working Group on Undergraduate Education and Training (the Fottrell Report)* [Online]. Dublin, Ireland Available: https://www.education.ie/en/Publications/Policy-Reports/Medical-Education-in-Ireland-A-New-Direction-Report-of-the-Working-Group-on-Undergraduate-Medical-Education-and-Training.pdf [Accessed Octiber 1 2013].

References

Foundation for Advancement of International Medical Education and Research. 2014. *Mapping the World's Medical Schools* [Online]. Available: http://www.faimer.org/resources/mapping.html [Accessed 16/07/ 2014].

Frank, J. R. 2005. *The CanMEDS 2005 physician competency framework: better standards, better physicians, better care* [Online]. Royal College of Physicians and Surgeons of Canada. Available: http://www.royalcollege.ca/portal/page/portal/rc/common/documents/c anmeds/resources/publications/framework_full_e.pdf [Accessed 12/02/ 2015].

Freeman, R. E. 2010. *Strategic management: A stakeholder approach,* New York Cambridge University Press.

Gale, T. C. E., Roberts, M. J., Sice, P. J., Langton, J. A., Patterson, F. C., Carr, A. S., Anderson, I. R., Lam, W. H. & Davies, P. R. F. 2010. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *British Journal of Anaesthesia,* 105**,** 603-609.

Gantley, M., Harding, G., Kumar, S. & Tissier, J. 1999. *An Introduction to Qualitative Methods for Health Professionals,* London, UK, Royal College of General Practitioners.

Garlick, P. B. & Brown, G. 2008. Widening Participation in Medicine. *British Medical Journal,* 336**,** 1111-1113.

General Medical Council. 2009. *Tomorrow's Doctors: Education Outcomes and Standards for Undergraduate Medical Education* [Online]. London: General Medical Council Available: http://www.gmc-uk.org/publications/undergraduate_education_publications.asp [Accessed 29/03/ 2012].

Gilliland, S. W. 1993. The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *The Academy of Management Review,* 18**,** 694-734.

Gilliland, S. W. 1994. Effects of Procedural and Distributive Justice on Reactions to a Selection System. *Journal of Applied Psychology,* 79**,** 691-701.

Gilliland, S. & Cherry, B. 2000. Managing "Customers" of Selection Processes *In:* Kehoe, J. F. (ed.) *Managing selection in changing organizations: Human resource strategies.* San Francisco CA: Jossey-Bass.

Gilliland, S. & Hale, J. 2005. How can justice be used to improve employee selection practices? . *In:* Greenberg, J. & Colquitt, J. A. (eds.) *Handbook of organizational justice.* New York Psychology Press

Ginsburg, S., Schreiber, M. & Regehr, G. 2004. The Lore of Admissions Policies: Contrasting Formal and Informal Understandings of the Residency Selection Process. *Advances in Health Sciences Education,* 9**,** 137-145.

Glaser, B. G. & Strauss, A. L. 2009. *The discovery of grounded theory: Strategies for qualitative research,* Chicago, USA, Transaction Publishers.

Goho, J. & Blackman, A. 2006. The effectiveness of academic admission interviews: an exploratory meta-analysis. *Medical Teacher,* 28**,** 335-340.

Goulston, K. & Oates, K. 2009. *Admission policy review :* Faculty of Medicine, University of Sydney.

Greenberg, J. & Colquitt, J. A. 2005. *Handbook of Organizational Justice,* New York, Psychology Press.

Greene, J. C. 2007. *Mixed Methods in Social Inquiry* San Francisco Jossey-Bass (A Wiley Imprint)

Greenhalgh, T. 2014. *How to read a paper: The basics of evidence-based medicine,* Sussex, UK, Wiley Blackwell

References

Greenhalgh, T., Seyan, K. & Boynton 2004. "Not a university type": focus group study of social class, ethnic, and sex differences in school pupils' perceptions about medical school. *British Medical Journal* 328, 1541-7

Griffin, B. & Wilson, I. 2012. Associations between the big five personality factors and multiple mini-interviews. *Advances in Health Sciences Education,* 17**,** 377-388.

Griffin, B., Harding, D. W., Wilson, I. G. & Yeomans, N. D. 2008. Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. *Medical Journal of Australia,* 189**,** 270-3.

Griffin, B., Carless, S. & Wilson, I. 2012. The effect of commercial coaching on selection test performance. *Medical Teacher,* 35**,** 295-300.

Groves, M., O'Rourke, P. & Alexander, H. 2003. The association between student characteristics and the development of clinical reasoning in a graduate-entry, PBL medical programme. *Medical Teacher,* 25**,** 626-31.

Groves, M. A., Gordon, J. & Ryan, G. 2007. Entry tests for graduate medical programs: Is it time to re-think? *Medical Journal of Australia,* 186**,** 120-123.

Gula, L., Kennard, M., Sadlar, K., Whelan, E., Chan, F., Garcia, B. & Hammond, R. Reliability of Standardised interviews at Schulich Medicine 16th Ottawa Conference- Transforming healthcare through excellence in assessment and evaluation April 25th -29th 2014 Ottawa Canada.

Haig, A. & Dozier, M. 2003a. BEME Guide No 3: Systematic searching for evidence in medical education--Part 1: Sources of Information. *Medical Teacher* 25**,** 352-363.

Haig, A. & Dozier, M. 2003b. BEME Guide No. 3: Systematic searching for evidence in medical education--Part 2: Constructing Searches. *Medical Teacher,* 25**,** 463-484.

Hallock, J. A., Mckinley, D. W. & Boulet, J. R. 2007. Migration of doctors for undergraduate medical education. *Medical Teacher,* 29**,** 98-105.

Halpenny, D., Cadoo, K., Halpenny, M., Burke, J. & Torreggiani, W. C. 2010. The health professions admission test (HPAT) score and leaving certificate results can independently predict academic performance in medical school: Do we need both tests? *Irish Medical Journal,* 103**,** 1-3.

Hamdy, H., Prasad, K., Anderson, M. B., Scherpbier, A., Williams, R., Zwierstra, R. & Cuddihy, H. 2006. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher,* 28**,** 103-116.

Harden, R. 2014. Welcome Address. 16th Ottawa Conference- Transforming healthcare through excellence in assessment and evaluation. April 25th - 29th. Ottawa, Canada

Harden, R., Stevenson, M., Downie, W. W. & Wilson, G. 1975. Assessment of clinical competence using objective structured examination. *British Medical Journal,* 1**,** 447-451.

Harris, I. 2002. Qualitative Methods *In:* Norman, G. R., Van der Vleuten, C. & Newble, D. I. (eds.) *International Handbook of Research in Medical Education.* Amsterdam: Springer Science & Business Media.

Harris, I. 2003. What Does "The discovery of grounded theory" Have to Say to Medical Education? *Advances in Health Sciences Education,* 8**,** 49-61.

References

Harris, S. & Owen, C. 2007. Discerning quality: Using the multiple mini-interview in student selection for the Australian National University Medical School. *Medical Education,* 41**,** 234-241.

Henry, M. 2009. Gender Balance in Medicine. *The Irish Times*, Aug 27th.

Henry, M. 2010. Securing a place in Medicine *The Irish Times* Nov 10, p.15.

Henry, P. 2006. Educational and Career Barriers to the Medical Profession: Perceptions of Underrepresented Minority Students. *College Student Journal,* 40**,** 429-441.

Higher Education Authority 2008. National Plan for Equity of Access to Higher Education Dublin, Ireland.

Hissbach, J. C., Sehner, S., Harendza, S. & Hampe, W. 2014. Cutting costs of multiple mini-interviews - Changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC Medical Education,* 14, 54.

Hofmeister, M., Lockyer, J. & Crutcher, R. 2008. The Acceptability of the Multiple Mini Interview for Resident Selection. *Family Medicine,* 40**,** 734-740.

Hofmeister, M., Lockyer, J. & Crutcher, R. 2009. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education,* 43**,** 573-579.

Hojat, M., Erdmann, J. B., Veloski, J. J., Nasca, T. J., Callahan, C. A., Julian, E. & Peck, J. 2000. A validity study of the writing sample section of the Medical College Admission Test. *Academic Medicine,* 75**,** S25-S27.

Hopson, L. R., Losman, E. D., Stansfield, R. B., Vohra, T., Turner-Lawrence, D. & Burkhardt, J. C. 2012. Use of the multiple mini interview (MMI) for emergency medicine resident selection: Acceptability to participants and comparison with application data. *Academic Emergency Medicine,* 19**,** S200 Abstract 374.

Hopson, L. R., Burkhardt, J. C., Stansfield, R. B., Vohra, T., Turner-Lawrence, D. & Losman, E. D. 2014. The multiple mini-interview for emergency medicine resident selection. *Journal of Emergency Medicine,* 46**,** 537-543.

Houses of the Oireachtas. 2010. *Other Questions - Health Professional Education* [Online]. Dublin, Ireland: Office of the Houses of the Oireachtas. Available: http://debates.oireachtas.ie/dail/2010/09/29/00012.asp [Accessed 20/09 2012].

Houses of the Oireachtas. 2012. *Health Professions Admission Test Administration* [Online]. Dublin, Ireland Office of the Houses of the Oireachtas. Available: http://oireachtasdebates.oireachtas.ie/debates%20authoring/debateswebpack.nsf/takes/dail2012092000032?opendocument# [Accessed 12/09 2013].

Houses of the Oireachtas. 2014. *Health Professions Admission Test Administration* [Online]. Dublin, Ireland: Office of the House of the Oireachtas. Available: http://oireachtasdebates.oireachtas.ie/debates%20authoring/debateswebpack.nsf/takes/seanad2014031200006?opendocument [Accessed 12/03 2015].

Houston, M. 2009. Welcome for more men doing medicine: Medical schools pleased by return of gender balance. *Irish Times* 2009/8/18, p.6.

Houston, M. 2010. Questions raised over need for entry exam to medical school *The Irish Times*, Dec 22, p.3.

Humphrey, S., Dowson, S., Wall, D., Diwakar, V. & Goodyear, H. M. 2008. Multiple mini-interviews: Opinions of candidates and interviewers. *Medical Education,* 42**,** 207-213.

References

Hunter, J., Schmidt, F. & Le, H. 2006. Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology* 91**,** 594-612.

Hurwitz, B. & Vass, A. 2002. What's a good doctor, and how can you make one? *British Medical Journal,* 325**,** 667-668.

Husbands, A. & Dowell, J. 2013. Predictive validity of the Dundee multiple mini-interview *Medical Education,* 47**,** 717-725.

Husbands, A., Dowell, J. & Patterson, F. 2014a. Validating an integrity focused situational judgement test for pre-interview selection to medical school *16th Ottawa Conference- Transforming healthcare through excellence in assessment and evaluation. April 25th -29th* Ottawa, Canada.

Husbands, A., Mathieson, A., Dowell, J., Cleland, J. & Mackenzie, R. 2014b. Predictive validity of the UK clinical aptitude test in the final years of medical school: A prospective cohort study. *BMC Medical Education,* 14, 88.

Hussey, G. 2009. Gender Balance in Medicine. *The Irish Times*, Aug 20th.

Hyatt, D. 2012. Stakeholders' perceptions of IELTS as an entry requirement for higher education in the UK. *Journal of Further and Higher Education***,** 37, 844-836.

International English Language Testing System 2014. *What is IELTS?* [Online]. British Council, IDP Education, Cambridge English Language Assessment. Available: http://www.ielts.org/institutions/about_ielts/what_is_ielts.aspx [Accessed June 16th 2014].

James, D., Yates, J. & Nicholson, S. 2010. Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *British Medical Journal,* 340**,** c478.

Jauhar, S., Bartos, M., Hacker, T. & Srireddy, P. 2008. Can trainees contribute to recruitment into specialty training? *Psychiatric Bulletin,* 32**,** 470-472.

Jayasuriya, R., Lightman, E. & Marshall, M. 2012. Applying to medical school- an informed decision? *AMEE - The continuum of education in medicine and the healthcare professions.* August 25-29. Lyon France: Association for Medical Education in Europe (AMEE).

Jerant, A., Griffin, E., Rainwater, J., Henderson, M., Sousa, F., Bertakis, K. D., Fenton, J. J. & Franks, P. 2012. Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Academic Medicine,* 87**,** 1250-1259.

Jesson, J. & Lacey, F. 2006. How to do (or not to do) a critical literature review. *Pharmacy Education,* 6**,** 139-148.

Johnson, M. & Elam, C. 2001. Letters of recommendation: comparing perceptions of premedical advisors and admission committee members. *Academic Medicine,* 76**,** 651.

Johnson, R. B. & Onwuegbuzie, A. J. 2004. Mixed methods research: A research paradigm whose time has come. *Educational Researcher,* 33**,** 14-26.

Julian, E. R. 2005. Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine,* 80**,** 910-7.

Kaffenberger, B. H., Kaffenberger, J. A. & Zirwas, M. J. 2014. Academic dermatologists' views on the value of residency letters of recommendation. *Journal of the American Academy of Dermatology,* 71**,** 395-6.

References

Kanerva, A., Lammintakanen, J. & Kivinen, T. 2010. Experiences of the fairness of recruitment from unsuccessful applicants in the field of nursing. *Journal of Nursing Management,* 18**,** 8.

Keily, B. & Byrne, S. 2009. Gender Balance in Medicine. *The Irish Times*, Aug 21st.

Kelleher, D., Powderly, B., Kerins, D., Dunne, F., Kelly, C. & Finucane, P. 2009. Gender balance in medicine. *Irish Times* August 28*.*

Kelly, M. E., Regan, D., Dunne, F. P., Henn, P., Newell, J. & O'Flynn, S. 2013. To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills?–An observational cohort study. *BMC Medical Education,* 13, 68.

Kelly, M. E., Dowell, J., Husbands, A., Kropmans, T., Jackson, A. E., Dunne, F., O'Flynn, S., Newell, J. & Murphy, A. W. 2014a. Can Multiple Mini Interviews work in an Irish setting? A feasibility study. *Irish Medical Journal,* 107**,** 210-2.

Kelly, M. E., Gallagher, N., Dunne, F. P. & Murphy, A. W. 2014b. Views of doctors of varying disciplines on HPAT-Ireland as a selection tool for medicine. *Medical Teacher,* 36**,** 775-782

Kelly, M. E., Arnett, R., Dennehy, T., Duggan, E., Dunne, F., Gallagher, N., Hennessy, M., Lapthorne, S., Last, J., Mc Grath, D., Murphy, A. W., Stevens, L. & O'Flynn, S. 2014c. Stakeholder views on selection to undergraduate medicine in Ireland. *International Network for Researchers in Selection into Healthcare (INReSH).* Nov 10-11. London.

Kennedy, T. J. & Lingard, L. A. 2006. Making sense of grounded theory in medical education. *Medical Education,* 40**,** 101-108.

Kerr, A. 2006. Will the increase in medical places change student profile? *The Irish Times*, Feb 7.

Kleshinski, J., Shriner, C. & Khuder, S. A. 2008. The use of professionalism scenarios in the medical school interview process: Faculty and interviewee perceptions. *Medical Education Online,* 13**,** 2.

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B. & Wilkinson, M. 2012. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education,* 46**,** 399-408.

Kothari, D., Gourevitch, M., Lee, J., Grossman, E., Truncali, A., Ark, T. & Kalet, A. 2011. Undergraduate medical education in substance abuse: a review of the quality of the literature. *Academic Medicine* 86**,** 98-112.

Kreiter, C. D. & Axelson, R. D. 2013. A Perspective on Medical School Admission Research and Practice Over the Last 25 Years. *Teaching and Learning in Medicine,* 25**,** S50-S56.

Kreiter, C. D. & Kreiter, Y. 2007. A validity generalization perspective on the ability of undergraduate GPA and the medical college admission test to predict important outcomes. *Teaching and Learning in Medicine,* 19**,** 95-100.

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,* 9**,** 147-159.

Kropmans, T. J., O Donovan, B. G., Cunningham, D., Murphy, A. W., Flaherty, G., Nestel, D. & Dunne, F. P. 2011. An online management information system for objective structured clinical examinations. *Computer and Information Science,* 5**,** 38-48.

Kulasegaram, K., Reiter, H. I., Wiesner, W., Hackett, R. D. & Norman, G. R. 2010. Non-association between Neo-5 personality tests and multiple mini-interview. *Advances in Health Sciences Education,* 15**,** 415-423.

References

Kulatunga Moruzi, C. & Norman, G. R. 2002. Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions *Teaching and Learning in Medicine,* 14**,** 34-43.

Kumar, K., Roberts, C., Rothnie, I., Du Fresne, C. & Walton, M. 2009. Experiences of the multiple mini-interview: a qualitative analysis. *Medical Education,* 43**,** 360-7.

Kumwenda, B., Dowell, J. & Husbands, A. 2013. Is embellishing UCAS personal statements accepted practice in applications to medicine and dentistry? *Medical Teacher,* 35**,** 599-603.

Kyei-Blankson, L. S. 2005. *Predictive validity, differential validity, and differential prediction of the subtests of the Medical College Admission Test.* Doctorate of Philosophy, Ohio University.

Lakhan, S. E. 2003. Diversification of U.S. medical schools via affirmative action implementation. *BMC Medical Education,* 3**,** 6.

Lambe, P., Waters, C. & Bristow, D. 2012. The UK Clinical Aptitude Test: is it a fair test for selecting medical students? *Medical Teacher,* 34**,** e557-65.

Leinster, S. 2013. Selecting the right medical student. *BMC Medicine,* 11, 245.

Lemay, J. F., Lockyer, J. M., Collin, V. T. & Brownell, A. K. 2007. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Medical Education,* 41**,** 573-579.

Leung 2005. How Generalizable Are Justice Effects Across Cultures? *In:* Greenberg, J. & Colquitt, J. A. (eds.) *Handbook of organizational justice.* New York: Psychology Press

Lievens, F. 2013. Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education,* 47**,** 182-189.

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**,** 1181-1188.

Lievens, F. & Sackett, P. R. 2012. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology,* 97**,** 460-468.

Lincoln, Y. S. & Guba, E. G. 1985. *Naturalistic Inquiry,* Beverly Hills CA, Sage.

Linn, R. L. & Gronlund, N. E. 2000. *Measurement and Assessment in Teaching* New Jersey, USA, Merrill, Prentice-Hall.

Loh, E. 2014. *Personal Communication- Email to* Kelly, M. December

Lubarsky, S. & Young, M. 2013. Assessing the multiple mini interview (MMI) for use in selecting applicants to a neurology residency program. *Neurology,* 80,P04-253.

Lynch, B., Mackenzie, R., Dowell, J., Cleland, J. & Prescott, G. 2009. Does the UKCAT predict Year 1 performance in medical school? *Medical Education,* 43**,** 1203-9.

MacFarlane, A. & O Reilly-De Brún, M. 2012. Using a theory-driven conceptual framework in qualitative health research. *Qualitative Health Research,* 22**,** 607-618.

Mack, N., Woodsong, C., Macqueen, K. M., Guest, G. & Namey, E. 2005. *Qualitative research methods: a data collectors field guide* [Online]. Available: http://www.fhi360.org/resource/qualitative-research-methods-data-collectors-field-guide [Accessed June 16th 2014].

Maher, B. M., Hynes, H., Sweeney, C., Khashan, A. S., O'Rourke, M., Doran, K., Harris, A. & O'Flynn, S. 2013. Medical school attrition-beyond the statistics a ten year retrospective study. *BMC Medical Education,* 13, 13.

References

Marrin, L. M., Mcintosh, A. K., Keane, D. & Schmuck, L. M. 2004. Use of the Paired-Comparison Technique to Determine the Most Valued Qualities of the McMaster Medical Programme Admissions Process. *Advances in Health Sciences Education,* 9**,** 7.

Mathers, J. & Parry, J. 2010. Older mature students' experiences of applying to study medicine in England: an interview study. *Medical Education,* 44**,** 1084-94.

Maudsley, G. 2011. Mixing it but not mixed-up: Mixed methods research in medical education (a critical narrative review). *Medical Teacher,* 33, e92–e104.

May, T. 2011. *Social research: Issues, methods and process,* Berkshire, England, McGraw-Hill Education (UK).

Mayr, S., Erdfelder, E., Buchner, A. & Faul, F. 2007. A short tutorial of GPower. *Tutorials in Quantitative Methods for Psychology,* 3**,** 51-59.

McAndrew, R. & Ellis, J. 2012. An evaluation of the multiple mini-interview as a selection tool for dental students. *British Dental Journal,* 212**,** 331-335.

McDonagh, S. 2010. Securing a place in medicine *The Irish Times* June 24, p.15.

McGaghie, W. C. 2002. Assessing readiness for medical education: evolution of the medical college admission test. *Journal of the American Medical Association,* 288**,** 1085-1090.

McManus, I.C. & Powis, D. 2007. Testing medical school selection tests. *Medical Journal of Australia* 186**,** 118-9.

McManus, I. & Wakeford, R. 2014. PLAB and UK graduates' performance on MRCP (UK) and MRCGP examinations: data linkage study. *British Medical Journal,* 348.

McManus, I., Smithers, E., Partridge, P., Keeling, A. & Fleming, P. 2003. A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *British Medical Journal,* 327**,** 139 - 142.

McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D. & Richards, P. 2005. Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal,* 331**,** 555-9.

McManus, I. C., Woolf, K. & Dacre, J. 2008. The educational background and qualifications of UK medical students from ethnic minorities. *BMC Medical Education,* 8, 21.

McManus, I. C., Ferguson, E., Wakeford, R., Powis, D. & James, D. 2011a. Predictive validity of the Biomedical Admissions Test: An evaluation and case study. *Medical Teacher,* 33**,** 53-57.

McManus, I. C., Ferguson, E., Wakeford, R., Powis, D. & James, D. 2011b. Response to comments by Emery and Bell, medical teacher 33(1). *Medical Teacher,* 33**,** 60-61.

McManus, I. C., Dewberry, C., Nicholson, S., Dowell, J. S., Woolf, K. & Potts, H. W. W. 2013a. 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, 243.

McManus, I. C., Woolf, K., Dacre, J., Paice, E. & Dewberry, C. 2013b. The Academic Backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors. *BMC Medicine,* 11, 242.

McManus, I. C., Dewberry, C., Nicholson, S. & Dowell, J. S. 2013c. The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome

in a cross-sectional collaborative study of 12 UK medical schools. *BMC Medicine,* 11, 244.

McManus, I., Elder, A. & Dacre, J. 2013d. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCGP(UK) PACES and nPACES examinations. *BMC Medical Education* 13, 103.

Medical Council. 2010a. *Eight Domains of Good Professional Practice* [Online]. : Medical Council. Available: http://www.medicalcouncil.ie/Education-and-Training/Good-Professional-Practice/Eight-Domains-of-Good-Professional-Practice-as-devised-by-Medical-Council.pdf [Accessed 01/05 2012].

Medical Council. 2010b. *Number and Type of Complaints Received by Preliminary Proceedings Committee in 2010* [Online]. http://www.medicalcouncil.ie/Professional-Standards/Statistics/Statistics-2010/Statistics-2010.html: Medical Council Available: http://www.medicalcouncil.ie/Professional-Standards/Statistics/Statistics-2010/Statistics-2010.html [Accessed 18/02/ 2012].

Medical Deans Australia and New Zealand. 2014 *Student Statistics Annual Tables 2013* [Online]. Available: http://www.medicaldeans.org.au/statistics/annualtables [Accessed 16/07/ 2014].

Medical Schools Council 2014a. Selecting for Excellence Final Report. London, UK: Medical Schools Council

Medical Schools Council. 2014b. *Number of places available* [Online]. UK. Available: http://www.medschools.ac.uk/Students/howtoapply/international/Pages/Limitedplaces.aspx [Accessed 16/07/ 2014].

Mercer, A. & Puddey, I. B. 2011. Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. *Medical Teacher,* 33**,** 997-1004.

Milburn, A. 2009. Unleashing aspiration: the final report of the panel on fair access to the professions. London, UK. Cabinet Office.

Milne, C. K., Bellini, L. M. & Shea, J. A. 2001. Applicants' perceptions of the formal faculty interview during residency recruitment. *Academic medicine Journal of the Association of American Medical Colleges,* 76**,** 501.

Mitchson, H. 2009. Assessment centres for core medical training: how do the assessors feel this compares with the traditional interview? *Clinical Medicine* 9**,** 147-150.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine,* 151**,** 264-269.

Molloy, P. 2010. Securing a Place in Medicine. *The Irish Times*, June 24, p.15.

Monroe, A., Quinn, E., Samuelson, W., Dunleavy, D. M. & Dowd, K. W. 2013. An overview of the medical school admission process and use of applicant data in decision making: What has changed since the 1980s? *Academic Medicine,* 88**,** 672-681.

Mookherjee, S., Pheatt, L., Ranji, S. R. & Chou, C. L. 2013. Physical examination education in graduate medical education—a systematic review of the literature. *Journal of General Internal Medicine,* 28**,** 1090-1099.

Mooney, B. 2012. Why the HPAT system has failed. *The Irish Times*, Sep 18, p. 14.

Mooney, B. 2012 Medicine Entry Test Fails to level the playing field. *The Irish Times*, June 6, p.14.

References

Moreau, K., Reiter, H. & Eva, K. W. 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**,** 58-61.

Murchan Emilie [@Emiliemurchan]. 2012. *My brain feels like its melting from doing hpat questions, actually kinda nervous now ahhh [Tweet]* [Online]. Available: https://twitter.com/search?q=HPAT%20lang%3Aen%20since%3A2012-01-01%20until%3A2014-06-24&src=typd [Accessed June 23 2014].

Murphy, D., Bruce, D. & Eva, K. 2008. Workplace based assessment for General Practitioners: using stakeholder perception to aid blueprinting of an assessment battery. *Medical Education,* 42**,** 96-107.

Murray, N. 2009. 3,000 Leaving Certs to sit aptitude test for medicine *Irish Examiner*, February 10. Available: http://www.irishexaminer.com/ireland/3000-leaving-certs-to-sit-aptitude-test-for-medicine-84166.html [Accessed 18 Nov 2013]

Murray, N. 2011. More students with lower Leaving Cert points get medicine. *Irish Examiner August* 23. Available: http://www.irishexaminer.com/ireland/more-students-with-lower-leaving-cert-points-get-medicine-165133.html [Accessed 30 April 2013]

Neligan, M. 2009. Selecting our future doctors: How can we ensure the right students get to study medicine? *The Irish Times August 25*[th], p.A4.

Nicholson, S. 2005. The benefits of aptitude testing for selecting medical students. *British Medical Journal,* 331**,** 559-561.

Niyomdecha, W., Sukthamrong, J. & Saeheng, S. 2012. The student selection interview in Faculty of Medicine, Prince of Songkla University *15th Ottawa Conference - Assessment of Competence in Medicine and the Healthcare professions.* March 11th - 13th Kuala Lumpur, Malaysia ASME

Norcini, J. J., Lipner, R. S. & Kimball, H. R. 2002. Certifying examination performance and patient outcomes following acute myocardial infarction. *Medical Education,* 36**,** 853-859.

Norman, G. 2004. The Morality of Medical School Admissions. *Advances in Health Sciences Education,* 9**,** 79-83.

Norman, G. 2010. Non-cognitive factors in health sciences education: from the clinic floor to the cutting room floor. *Advances in Health Sciences Education,* 15**,** 1-8.

O'Brien, A., Harvey, J., Shannon, M., Lewis, K. & Valencia, O. 2011. A comparison of multiple mini-interviews and structured interviews in a UK setting. *Medical Teacher,* 33**,** 397-402.

O'Flynn, S. 2010. Entry and selection to medical school-do we know what we should measure and how we should measure it? *In:* Salerno-Kenned, R. & O'Flynn, S. (eds.) *Medical Education: The State of the Art.* Hauppauge, NY: Nova Science Publishers.

O'Flynn, S., Mills, A. & Fitzgerald, T. 2012. *Interim Report School Leaver Entrants Ireland*. Report of the National Research Group Evaluating Revised Entry Mechanisms to Medicine. Irish Universities Association. Available: http://www.iua.ie/wp-content/uploads/2013/06/HPAT-report-July2012.pdf [Accessed July 10 2013]

O'Flynn, S., Power, S., Hennessy, M., Mills, A. & Lapthorne, S. 2013a. Guidance Counsellors perspectives on revised entry and selection mechanisms to

References

medicine and the HPAT. *Guideline - A Publication of the Institute of Guidance Counsellors*

O'Flynn, S., Fitzgerald, T. & Mills, A. 2013b. Modelling the impact of old and new mechanisms of entry and selection to medical school in Ireland: Who gets in? *Irish Journal of Medical Science,* 182**,** 421-427.

O'Flynn, S., Mills, A. & Fitzgerald, A. 2013c. Entry to medical school the gender question. What has happened? *Irish Medical Journal* 106, 230-232.

O'Neill, L., Vonsild, M. C., Wallstedt, B. & Dornan, T. 2013. Admission criteria and diversity in medical school. *Medical education,* 47**,** 557-561.

O'Regan, E. 2012. Minister wants to scrap test for medical schools. *Independent.ie* Feb 18. Available: http://www.independent.ie/irish-news/minister-wants-to-scrap-test-for-medical-schools-26822872.html [Accessed 1 March 2013]

O'Sullivan Ryan[@Jandyman95]. 2014. *Received my #Hpat results. Came out pretty high. The prognosis is good ;) Absolutely delighted! :) [Tweet]* [Online]. Available: https://twitter.com/search?src=typd&q=%23hpat [Accessed June 23 2014].

Organisation for Economic Co-operation and Development (OECD) 2014. How many students study abroad and where do they go? *Education at a Glance 2013: Highlights.* OECD Publishing.

Papadakis, M. A., Teherani, A., Banach, M. A., Knettler, T. R., Rattner, S. L., Stern, D. T., Veloski, J. & Hodgson, C. S. 2005. Disciplinary Action by Medical Boards and Prior Behavior in Medical School. *The New England Journal of Medicine,* 353**,** 2673-2682.

Parry, J., Mathers, J., Stevens, A., Parsons, A., Lilford, R., Spurgeon, P. & Thomas, H. 2006. Admissions processes for five year medical courses at English schools: Review. *British Medical Journal,* 332**,** 1005-1009.

Pasw Statistics 18 2010. Release Version 18.1 ed. Chicago

Patel, S., Singh, G., Zhang, S. L., Lee, T., Vettese, T. & Levine, D. Internal Medicine Residency Applicants' thoughts about group interviews 4th International Conference of Education, Research and Innovation (ICERI), Nov 14-16 2011 Madrid, Spain. 7133-7139.

Patterson, D. G., Baldwin, L. M. & Olsen, P. M. 2009. Supports and Obstacles in the Medical School Application Process for American Indians and Alaska Natives. *Journal of Health Care for the Poor and Underserved,* 20**,** 308-329.

Patterson, F. & Ferguson, E. 2007. Selection for Medical Education and Training *Understanding Medical Education* Edinburgh, Scotland. Association for the Study of Medical Education.

Patterson, F. & Ferguson, E. 2010. Selection for medical education and training. *In:* Swanwick , T. (ed.) *Understanding Medical Education* Sussex, UK: Wiley Online Library.

Patterson, F. & Ferguson, E. 2012. Testing non-cognitive attributes in selection centres: how to avoid being reliably wrong. *Medical Education,* 46**,** 240-2.

Patterson, F., Zibarras, L., Carr, V., Irish, B. & Gregory, S. 2011. Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education,* 45**,** 289-97.

Patterson, F., Lievens, F., Kerrin, M., Zibarras, L. & Carette, B. 2012a. 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**,** 486-496.

References

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M. & O'Neill, P. 2012b. Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education,* 46**,** 850-868.

Patterson, F., Ashworth, V., Fitzpatrick, S., Walker, K. & Good, D. 2013. Design of a new situational judgement test to assess the professional attributes of junior doctors *AMEE - Coluring outside the lines* August 24th -28th Prague, Czech Republic Association for Medical Education in Europe (AMEE).

Patton, M. Q. 2002. *Qualitative research and evaluation methods,* Thousand Oaks, CA, Sage Publications, Inc.

Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C. & Nadarajah, V. D. 2013. The Multiple Mini-Interview (MMI) for student selection in health professions training-A systematic review. *Medical Teacher,* 35**,** 1027-1041.

Pearson Vue. 2014. *About Pearson VUE* [Online]. Pearson Education Inc. Available: http://home.pearsonvue.com/Home.aspx [Accessed July 11th 2014].

Peile, E. 2014. Selecting an internationally diverse medical workforce. *British Medical Journal,* 348, 2696-68.

Perkins, A., Burton, L., Dray, B. & Elcock, K. 2013. Evaluation of a multiple-mini-interview protocol used as a selection tool for entry to an undergraduate nursing programme. *Nurse Education Today,* 33**,** 465-469.

Poole, P. & Shulruf, B. 2013. Shaping the future medical workforce: take care with selection tools. *J Prim Health Care,* 5**,** 269-75.

Poole, P., Shulruf, B., Harley, B., Monigatti, J., Barrow, M., Reid, P., Prendergast, C. & Bagg, W. 2012a. Shedding light on the decision to retain an interview for medical student selection. *New Zealand Medical Journal,* 125**,** 81-88.

Poole, P., Shulruf, B., Rudland, J. & Wilkinson, T. 2012b. Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study. *Medical Education,* 46**,** 163-171.

Pope, C. & Mays, N. 2006. *Qualitative Research in Healthcare,* Oxford, UK, Blackwell Publishing.

Powis, D.A. 1994. Selecting medical students. *Medical Education,* 28**,** 443-469.

Powis, D. A 1998. How to do it: Select medical students. *British Medical Journal,* 317**,** 2.

Powis, D.A. 2003. Selecting medical students. *Medical Education,* 37**,** 1064-1065

Powis, D. 2010. Improving the selection of medical students. *BMJ (Online),* 340**,** 432.

Powis, D. 2014. Selecting medical students: An unresolved challenge. *Medical Teacher***,** 1-9.

Powis, D., Hamilton, J. & Mcmanus, C. I. 2007a. Widening Access by Changing the Criteria for Selecting Medical Students. *Teaching and Teacher Education: An International Journal of Research and Studies,* 23**,** 1235-1245.

Powis, D., James, D. & Ferguson, E. 2007b. Demographic and socio-economic associations with academic attainment (UCAS tariff scores) in applicants to medical school. *Medical Education,* 41**,** 242-9.

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**,** 215-223.

Puddey, I. B. & Mercer, A. 2013. Socio-economic predictors of performance in the Undergraduate Medicine and Health Sciences Admission Test (UMAT). *BMC Medical Education,* 13**,** 155.

References

Puddey, I. B. & Mercer, A. 2014. Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education,* 14**,** 31.

QSR International 2012. NVIVO Qualitative Data Analysis Software. 10 ed.: QSR International Pty Ltd.

Quinlivan, J. A., Lam, L. T., Wan, S. & Petersen, R. W. 2010. Selecting medical students for academic and attitudinal outcomes in a Catholic medical school. *Medical Journal of Australia,* 193**,** 347-50.

Quinn, A., Corrigan, M. A., Broderick, J., Mcentee, G. & Hill, A. D. 2010. A comparison of performances of consultant surgeons, NCHDs and medical students in a modified HPAT examination. *Irish Medical Journal* 103**,** 172-3.

Raghavan, M., Burnett, M., Martin, B. D., Christensen, H., Young, D. G., Mackalski, B. & Aoki, F. 2013a. Utility of a writing station in the multiple mini-interview. *Journal of Veterinary Medical Education,* 40**,** 177-183.

Raghavan, M., Martin, B. D., Burnett, M., Aoki, F., Christensen, H., Mackalski, F., Young, D. G. & Ripstein, I. 2013b. Multiple mini-interview scores of medical school applicants with and without rural attributes. *Rural and Remote Health,* 13,2362.

Ramachandran, S. & Venkatesh, H. 2014. A call for greater transparency in the evidence base supporting the BMAT. *Medical Teacher,* 1,1.

Ranchin, J. & Rebolledo-Gómez, C. 2013. *Students – the migrants everyone wants* [Online]. Organisation for Economic Co-operation and Development Available: http://oecdeducationtoday.blogspot.fr/2013/07/students-migrants-everyone-wants.html [Accessed 23/03 2015].

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**,** 444-448.

Razack, S., Faremo, S., Drolet, F., Snell, L., Wiseman, J. & Pickering, J. 2009. Multiple mini-interviews versus traditional interviews: Stakeholder acceptability comparison. *Medical Education,* 43**,** 993-1000.

Reed, D. 2014. *Personal Communication- Email to Kelly M.* December 1st

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M 2007. Association between funding and quality of published medical education research. *Journal of the American Medical Association,* 298**,** 1002-1009.

Reibnegger, G., Caluba, H. C., Ithaler, D., Manhal, S., Neges, H. M. & Smolle, J. 2010. Progress of medical students after open admission or admission based on knowledge tests. *Medical Education,* 44**,** 205-14.

Reiter, H. I. & Eva, K. W. 2005. Reflecting the relative values of community, faculty, and students in the admissions tools of medical school. *Teaching & Learning in Medicine,* 17**,** 4-8.

Reiter, H. & Eva, K. W. 2011. Selection for Medicine *In:* Dornan, T., Mann, K., Scherpbier, A. & Spencer, J. (eds.) *Medical Education; Theory and Practice* Elsevier, Churchill Livingstone

Reiter, H. I., Salvatori, P., Rosenfeld, J., Trinh, K. & Eva, K. W. 2006. The effect of defined violations of test security on admissions outcomes using multiple mini-interviews. *Medical Education,* 40, 36-42.

Reiter, H. I., Eva, K. W., Rosenfeld, J. & Norman, G. R. 2007. Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education,* 41**,** 378-384.

References

Reiter, H. I., Lockyer, J., Ziola, B., Courneya, C. A. & Eva, K. 2012. Should efforts in favor of medical student diversity be focused during admissions or farther upstream? *Academic Medicine,* 87**,** 443-448.

Rich, E. 2011. Medical Students and Interviewers opinions of traditional interviews for selection to a widening access medical program *AMEE - Inspire and be inspired* August 29th-31st Vienna Austria Association for Medical Education in Europe (AMEE).

Rizo, C. A. & Jadad, A. R. 2002. Doctors should be good companions for people. *British Medical Journal,* 325**,** 711.

Roberto, K. A., Gigliotti, C. M. & Husser, E. K. 2005. Older women's experiences with multiple health conditions: Daily challenges and care practices. *Health Care for Women International,* 26**,** 672-692.

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**,** 396-404.

Roberts, C., Zoanetti, N. & Rothnie, I. 2009. Validating a multiple mini-interview question bank assessing entry-level reasoning skills in candidates for graduate-entry medicine and dentistry programmes. *Medical Education,* 43**,** 350-359.

Roberts, C., Rothnie, I., Zoanetti, N. & Crossley, J. 2010. Should candidate scores be adjusted for interviewer stringency or leniency in the multiple mini-interview? *Medical Education,* 44**,** 690-698.

Rodgerson, M., Husbands, A. & Dowell, J. 2013 Investigating the acceptability of Multiple Mini Interviews (MMIs) at Dundee *ASME Annual Scientific Meeting - Changes in healthcare and the effects on medical education* July 10th-12th Edinburgh, UK: ASME- The Association for the Study of Medical Education

Rosenfeld, J. M., Reiter, H. I., Trinh, K. & Eva, K. W. 2008. A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education,* 13**,** 43-58.

RTÉ News. 2012a. *HPAT-Medicine* [Online]. RTÉ News Bulletin Radio Telefis Éireann (RTÉ). Available: http://www.rte.ie/news/2012/0921/338533-hpat-medicine [Accessed October 1 2013].

RTÉ News. 2012b. *HPAT fails to broaden entry to medicine courses* [Online]. RTÉ News.ie. Available: http://www.rte.ie/news/player/2012/0921/3398259-hpat-fails-to-broaden-entry-to-medicine-courses/ [Accessed June 23 2014].

Rubin, H. J. & Rubin, I. S. 2004. *Qualitative Interviewing: The art of hearing data,* California, Sage.

Sackett, P. R. & Yang, H. 2000. Correction for range restriction: an expanded typology. *Journal of Applied Psychology,* 85**,** 112-118.

Sackett, P. R., Borneman, M. J. & Connelly, B. S. 2008. High stakes testing in higher education and employment: appraising the evidence for validity and fairness. *American Psychologist,* 63**,** 215-227.

Saha, S., Guiton, G., Wimmers, P. F. & Wilkerson, L. 2008. Student body racial and ethnic composition and diversity-related outcomes in US medical schools. *Journal of the American Medical Association,* 300**,** 1135-1145.

Salvatori, P. 2001. Reliability and Validity of Admissions Tools Used to Select Students for the Health Professions. *Advances in Health Sciences Education,* 6**,** 159-175.

References

Samarasekera, D., Tay, S. & Hooi, S. 2014. A new approach to selecting medical students-National University of Signapore School of Medicine Experience *16th Ottawa Conference- Transforming healthcare through excellence in assessment and evaluation* April 25th-29th Ottawa, Canada.

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.

Schifferdecker, K. E. & Reed, V. A. 2009. Using mixed methods research in medical education: basic guidelines for researchers. *Medical Education,* 43**,** 637-644.

Schmidt, F., Ones, D. & Hunter, J. 1992. Personnel-selection. *Annual Review of Psychology,* 43**,** 627-670.

School of Medicine. 2012. *A Guide for (non-EU) International Students* [Online]. NUI Galway Available: http://www.nuigalway.ie/medicine/documents/newinternational_student_guide_2012.pdf [Accessed 23/03 2015].

Schuwirth, L. W. & Van Der Vleuten, C. P. 2010. *How to design a useful test: the principles of assessment,* UK, Wiley-Blackwell.

Schwartz, S. 2004. *Fair admissions to higher education: recommendations for good practice* [Online]. UK: Admissions to Higher Education Steering Group. Available: http://dera.ioe.ac.uk/5284/1/finalreport.pdf [Accessed 10/08/2014].

Searle, J. & Mcharg, J. 2003. Selection for medical school: just pick the right students and the rest is easy! *Medical Education,* 37**,** 458-463.

Sharma, R., Gordon, M., Dharamsi, S. & Gibbs, T. 2015. Systematic reviews in medical education: A Practical Approach: AMEE Guide 94. *Medical Teacher,* 37**,** 108-124.

Shulruf, B. 2015. *Personal communication, email to Kelly M. Jan* 28th

Shulruf, B., Poole, P., Wang, G. Y., Rudland, J. & Wilkinson, T. 2012. How well do selection tools predict performance later in a medical programme? *Advances in Health Sciences Education,* 17**,** 615-626.

Siggins, L. 2010. Entry Test ´unfair´ to Irish Speakers. *The Irish Times*, Mar 2nd, p.A2.

Silverman, J., Kurtz, S. M., Draper, J., Van Dalen, J. & Platt, F. W. 2005. *Skills for communicating with patients,* Oxford, UK, Radcliffe Pub.

Simpson, P. L., Scicluna, H. A., Jones, P. D., Cole, A. M., O'Sullivan, A. J., Harris, P. G., Velan, G. & McNeil, H. P. 2014. Predictive validity of a new integrated selection process for medical school admission. *BMC Medical Education,* 14**,** 86.

Siu, E. & Reiter, H. I. 2009. Overview: What's worked and what hasn't as a guide towards predictive Admissions tool development. *Advances in Health Sciences Education,* 14**,** 759-775.

Smyth, D. 1946a. Some principles in the selection of medical students *British Medical Journal,* 2**,** 357-347.

Smyth, D. H. 1946b. Selection of Medical Students *British Medical Journal,* 2**,** 626.

Stafford, J., Bowman, R., Ewing, T., Hanna, J. & Lopez-De Fede, A. 1997. *Building Culture Bridges,* Bloomington, IN National Education Service.

Stagg, P. & Rosenthal, D. R. 2012. Why community members want to participate in the selection of students into medical school. *Rural and Remote Health,* 12, 1954.

References

State Examinations Commission. 2009. *The Leaving Certificate (Established) Programme* [Online]. State Exams Commission Available: http://www.examinations.ie/index.php?l=en&mc=ca&sc=sb [Accessed 08/05/ 2012].

Steele, K. 2011. Selecting Tomorrow's Doctors. *Ulster Medical Journal,* 80**,** 62-67.

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**,** 229-232.

Stowe, C. D., Castleberry, A. N., O'Brien, C. E., Flowers, S. K., Warmack, T. S. & Gardner, S. F. 2014. Development and implementation of the multiple mini-interview in pharmacy admissions. *Currents in Pharmacy Teaching and Learning,* 6**,** 849-855.

Streiner, D. L. & Norman, G. R. 2008. *Health Measurement Scales* Oxford, UK, Oxford University Press

Tamblyn R., Abrahamowicz M., Brailovsky C., Grand'maison P., Lescop J., Norcini J., Girard N. & Haggerty J. 1998. Association between licensing examination scores and resource use and quality of care in primary care practice. *Journal of the American Medical Association,* 280, 989-1000.

Tamblyn, R., Abrahamowicz, M., Dauphinee, W., Hanley, J., Norcini, J., Girard, N., Grand'maison, P. & Brailovsky, C. 2002. Association between licensure examination scores and practice in primary care. *Journal of the American Medical Association,* 288**,** 3019-3027.

Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., Smee, S., Blackmore, D., Winslade, N., Girard, N., Du Berger, R., Bartman, I., Buckeridge, D. & Hanley, J. 2007 Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *Journal of the American Medical Association* 298**,** 993-1002.

Tashakkori, A. & Teddlie, C. 2010. *Sage Handbook of Mixed Methods in Social and Behavioural Research* Thousand Oaks, California Sage Publications

Tavakol, M. & Sandars, J. 2014 Quantitative and qualitative methods in medical education research: AMEE Guide no 90: Part II. *Medical Teacher* 36**,** 838-849.

The Irish Times. 2006. Training doctors. *The Irish Times* August 6

The Sunday Times. 2013. Half of doctors fail the HPAT. *The Sunday Times*, 14th April, p.1.

Tiffin, P. A., Dowell, J. S. & McLachlan, J. C. 2012. Widening access to UK medical education for under-represented socioeconomic groups: Modelling the impact of the UKCAT in the 2009 cohort. *British Medical Journal (Online),* 344, e1805.

Tiffin, P., McLachlan, J., Webster, L. & Nicholson, S. 2014a. Comparison of the sensitivity of the UKCAT and A Levels to sociodemographic characteristics: A national study. *BMC Medical Education,* 14,7.

Tiffin, P. A., Illing, J., Kasim, A. S. & McLachlan, J. C. 2014b. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *British Medical Journal ,* 348**,** g2622.

Till, H., Myford, C. & Dowell, J. 2013. Improving student selection using multiple mini-interviews with multifaceted rasch modeling. *Academic Medicine,* 88**,** 216-223.

References

Tiller, D., O'Mara, D., Rothnie, I., Dunn, S., Lee, L. & Roberts, C. 2013. Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education,* 47**,** 801-810.

Tippins, N. & Adler, S. 2011. *Technology-enhanced assessment of talent,* San Francisco, CA Jossey-Boss.

Tonks, A. 2002. What's a good doctor and how do you make one? Summary of responses. *British Medical Journal,* 325**,** 711-715.

Truxillo, D. M., Bauer, T. N. & J., S. R. 2001. Mutliple dimensions of procedural justice: Longitudinal effects on selection system fairness and test taking self-efficacy. *International Journal of Selection & Assessment,* 9**,** 330-349.

Truxillo, D. M., Steiner, D. D. & Gilliland, S. W. 2004. The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment,* 12**,** 39-53.

Tufford, L. & Newman, P. 2012. Bracketing in Qualitative Research *Qualitative Social Work* 11**,** 16.

Turner, R. & Nicholson, S. 2011. Reasons selectors give for accepting and rejecting medical applicants before interview. *Medical Education,* 45**,** 298-307.

U.S. Department of Labor Employment and Training Administration 1999. Testing and assessment: an employer's guide to good practices. Washington, DC.

Uijtdehaage, S., Doyle, L. & Parker, N. 2011. Enhancing the reliability of the multiple mini-interview for selecting prospective health care leaders. *Academic Medicine,* 86**,** 1032-9.

UKCAT Consortium 2007a. UKCAT 2006/07 Annual Report. UK.

UKCAT Consortium 2007b. UKCAT 2007/08 Annual Report. UK.

UKCAT Consortium 2008. UKCAT 2008 Annual Report. UK.

UKCAT Consortium 2010. UKCAT Annual Report 2009/2010 UK.

UKCAT Consortium 2011. Annual report and financial statements for the year ended 31 March 2011. Nottingham UK

UKCAT Consortium 2012a. Annual Report and Financial Statements for the Year Ended 31 March 2012.

UKCAT Consortium. 2012b. *The UK Clinical Aptitude Test* [Online]. http://www.ukcat.ac.uk/. Available: http://www.ukcat.ac.uk/ [Accessed 17/02/ 2012].

UKCAT Consortium 2013. Annual report and financial statements for the year ended 31 March 2013. UK.

UKCAT Consortium 2014a. Annual Report and Financial Statements for the year ended 31 March 2014 UK.

UKCAT Consortium. 2014b. *The UK Clinical Aptitude Test* [Online]. UKCAT Consortium,. Available: http://www.ukcat.ac.uk/ [Accessed July 9th 2014].

Urlings-Strop, L. C., Stijnen, T., Themmen, A. P. N. & Splinter, T. a. W. 2009. Selection of medical students: A controlled experiment. *Medical Education,* 43**,** 175-183.

Urlings-Strop, L. C., Themmen, A. P. N., Stijnen, T. & Splinter, T. a. W. 2011. Selected medical students achieve better than lottery-admitted students during clerkships. *Medical Education,* 45**,** 1032-1040.

Vale, V. 2015. *Personal communication, email to Kelly M.* Feb 23rd

Van Den Bos, K. & Lind, E. A. 2001. The psychology of own versus others' treatment: Self-oriented and other-oriented effects on perceptions of

References

procedural justice. *Personality and Social Psychology Bulletin,* 27**,** 1324-1333.

Van Der Leeuw, R., Lombarts, K., Arah, O. & Heineman, M. 2012. A systematic review of the effects of residency training on patient outcomes. *BMC Medicine,* 10,65.

Van Der Vleuten, C. P. M. 1996. The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education,* 1**,** 41-67.

Veloski, J. J., Callahan, C. A., Xu, G., Hojat, M. & Nash, D. B. 2000. Prediction of students' performances on licensing examinations using age, race, sex, undergraduate GPAs, and MCAT scores. *Academic Medicine,* 75**,** S28-30.

Vermeulen, M., Tromp, F., Patterson, F., Kuyvenhoven, M. & Bottema, B. 2012. Candidates' reaction towards a new selection procedure for post graduate GP training *15th Ottawa Conference - Assessment of Competence in Medicine and the Healthcare professions* March 11th -13th Kuala Lumpur, Malaysia ASME

Violato, C. & Donnon, T. 2005. Does the Medical College Admission Test predict clinical reasoning skills? A longitudinal study employing the Medical Council of Canada clinical reasoning examination. *Academic Medicine,* 80**,** S14-S16.

Waheed, G., Mengal, M. A., Shah, S. S. & Sheikh, A. W. 2011. Faculty and students' perception about aptitude of professionalism in admission process of medical college. *Journal of Ayub Medical College,* 23**,** 166-9.

Walshe, J. 2009. Scrap unfair aptitude test for entry into medicine say doctors *Irish Independent August* 25th.

Westwood, M. A., Nunn, L. M., Redpath, C., Mills, P. & Crake, T. 2008. Applicants regard structured interviews as a fair method of selection: an audit of candidates. *Journal of the Royal Society of Medicine,* 101**,** 252-8.

White, J., Brownell, K., Lemay, J. F. & Lockyer, J. M. 2012. "What do they want me to say?" The hidden curriculum at work in the medical school selection process: a qualitative study. *BMC Medical Education,* 12**,** 17.

White, J. S., Lemay, J. F., Brownell, K. & Lockyer, J. 2011. "A chance to show yourself" - How do applicants approach medical school admission essays? *Medical Teacher,* 33**,** e541-e548.

Whitla, D., Orfield, G., Silen, W., Teperow, C., Howard, C. & Reede, J. 2003. Educational benefits of diversity in medical school: a survey of students. *Academic Medicine,* 78**,** 460-466.

Wilkie, J. S. 1946. Analysis of Results of a test given at University College, London, for the Selection of Women Medical Students for the Session 1945-6. *British Medical Journal,* 2**,** 2.

Wilkinson, D., Zhang, J., Byrne, G. J., Luke, H., Ozolins, I. Z., Parker, M. H. & Peterson, R. F. 2008. Medical school selection criteria and the prediction of academic performance. *Medical Journal of Australia,* 188, 349-354.

Wilkinson, D., Zhang, J. & Parker, M. 2011. Predictive validity of the undergraduate medicine and health sciences admission test for medical students' academic performance. *Medical Journal of Australia,* 194**,** 341-344.

Wilkinson, T. M. & Wilkinson, T. J. 2013. Preparation courses for a medical admissions test: Effectiveness contrasts with opinion. *Medical Education,* 47**,** 417-424.

Wilson, J. 2010. Recent IELTS materials. *Elt Journal,* 62**,** 15.

References

Wood, D. F. 1999. Medical school selection - Fair or unfair? *Medical Education,* 33**,** 399-401.

World Federation for Medical Educators (WFME). 2012. *Basic Medical Education WFME Global Standards for Quality Improvement: The 2012 Revision* [Online]. Available: http://www.wfme.org/standards/bme [Accessed June 26th 2014].

Wright, S. R. 2012. *A Mixed Methods Study of Medical School Admissions: Issues of Fairness and Student Performance.* Doctor of Philosophy PhD, University of Newcastle upon Tyne.

Wright, S. R. & Bradley, P. M. 2010. Has the UK Clinical Aptitude Test improved medical student selection? *Medical Education,* 44**,** 1069-76.

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.

Yates, J. & James, D. 2013. The UK clinical aptitude test and clinical course performance at Nottingham: A prospective cohort study. *BMC Medical Education,* 13,32.

Yen, W., Hovey, R., Hodwitz, K. & Zhang, S. 2011. An exploration of the relationship between emotional intelligence (EI) and the Multiple Mini-Interview (MMI). *Advances in Health Sciences Education,* 16**,** 59-67.

Young, J. W. & Kobrin, J. L. 2001. Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis. New York, USA. College Entrance Examination Board.

Zaidi, N. B., Swoboda, C., Wang, L. L. & Manuel, R. S. 2014. Variance in attributes assessed by the multiple mini-interview. *Medical Teacher,* 36**,** 794-798.

Zendejas, B., Brydges, R., Wang, A. & Cook, D. 2013. Patient outcomes in simulation-based medical education: a systematic review. *Journal of General Internal Medicine,* 28**,** 1078-89

Ziv, A., Rubin, O., Moshinsky, A., Gafni, N., Kotler, M., Dagan, Y., Lichtenberg, D., Mekori, Y. A. & Mittelman, M. 2008. MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Medical Education,* 42**,** 991-8.