The Multiple Mini-Interview (MMI) in medical student selection

by

Michael James Dodson
MBBS, B.Sc, B.Med.Sc, MPH, MHM, ACCAM

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Deakin University

December, 2012
I am the author of the thesis entitled

The Multiple Mini-Interview (MMI in Medical Student Selection

submitted for the degree of Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1968.

I certify that I am the student named below and that the information provided in the form is correct.

Full Name: Michael James Donsol

Signed: ____________________________

Date: 17/1/2013
DEAKIN UNIVERSITY
CANDIDATE DECLARATION

I certify that the thesis entitled

The Multiple Mini-interview (MMI) in medical student selection

submitted for the degree of

Doctor of Philosophy

is the result of my own work and that where reference is made to the work of others, due acknowledgment is given.

I also certify that any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: .................................................................
(Please Print)

Signed: .................................................................

Date: .................................................................
3 December 2012
Frontispiece

‘No greater responsibility devolves upon medical educators than that of determining who are qualified to study medicine and who are not’

Victor Johnson, First World Conference on Medical Education 1953

Petrus Koning with his Master
Watercolor by J. H. Prins, 1803
Photographed at the Museum Boerhaave, Leiden, NL On loan from the national museum Kroller-Muller Otterlo.

Caption card reads:
Petrus Koning began his career as an apprentice to the Utrecht professor of anatomy and obstetrics, Jan Bleuland. They are depicted together here amid the University's collection of wet and dry human preparations.
Acknowledgements

Completion of this thesis would not have been possible without the help and support of a number of people. In particular, I would like to express my thanks to:

- Professor Alister Ward and Professor Brendan Crotty, for their support and guidance during the course of my studies,
- Ms Lili Smilevski, for her assistance with the collection of data for the experimental work contained in this thesis, for proof-reading the thesis, and for coordinating the Deakin MMI,
- The Deakin medical students who participated in the study,
- The MMI interviewers for their participation in the study and for contributing to the selection process for entry into the Deakin Medical Course,
- My wife, Lisa, for her assistance in translating various historical documents, and for her ongoing support and tolerance,
- My daughter, Aria, for distracting me enough to prevent me being consumed by my thesis, but not so much as to hamper my progress,
- Ms Diep Nguyen for making me to notice the flowers instead of always looking up at the trees and;
- Professor Tony Goodwin for being a role model, mentor and friend, no matter how far I have flown from the nest.
Contents

Introduction ............................................................................................................... 7
1. The Selection of Medical Students: an Historical Survey ..................................... 11
  1.1 Early transmission of medical knowledge ...................................................... 11
  1.2 Selection of medical students in the East ........................................................ 12
  1.3 Selection of medical students in the West ...................................................... 41
  1.4 Issues raised and lessons learnt ................................ ..................................... 121
2. Reliability and validity of methods used to select applicants into medical school 142
  2.1 The aim of medical student selection............................................................142
  2.2 Reliability and validity .................................................................................143
  2.3 Reliability and validity of commonly used selection tools ............................149
  2.4 Selection methods: identifying and addressing limitations ............................182
  2.5 Models for the Selection of Medical Students ..............................................184
3. Attributes sought in candidates for medical school .............................................200
  3.1 Desirable qualities and professional values ...................................................201
  3.2 Classifying desirable qualities and professional values ..................................209
4. Introduction to Experimental work .....................................................................220
5. Reliability of the Deakin MMI ............................................................................230
  5.1 Introduction ..................................................................................................230
  5.2 Methods........................................................................................................231
  5.3 Results ..........................................................................................................234
  5.4 Discussion ....................................................................................................241
6. Scoring characteristics and Inter-rater reliability of interviewers for the MMI.....248
  6.1 Introduction ..................................................................................................248
  6.2 Methods........................................................................................................251
  6.3 Results..........................................................................................................255
  6.4 Discussion ....................................................................................................263
7. Validity of the Deakin MMI ...............................................................................272
  7.1 Introduction ..................................................................................................272
  7.2 Methods........................................................................................................276
  7.3 Results..........................................................................................................281
  7.4 Discussion ....................................................................................................288
8. Discussion & directions for further research .......................................................297
Introduction

The significance of the selection of candidates into medical training cannot be exaggerated. Medical practice is intimately linked with the lives of individuals and the welfare of the entire population (Kay 1944). Thus, it is essential that medical graduates are competent and able to be trusted with these responsibilities. Given that virtually all those admitted into medical school will graduate, and virtually all those that graduate will obtain a license to practise, selection into medical training represents the single greatest hurdle to becoming a doctor. Indeed, “the decision to admit individuals to medical school is, with few exceptions, tantamount to a decision to give them a license” (McGaghie 1990). Because of these factors, the selection of the student should be the prime concern of every medical faculty (Bauer 1956).

The selection of medical students has received varying degrees of academic and social interest over the past three millennia. However, it is surprising that a process that has such potential to influence the quality of health care and the status of the medical profession has not received more attention. The impact of the selection process on the lives of candidates must also be considered, as entry into medical training is a life-changing event for most candidates. In 1956, Bauer remarked that “the initial selection of the medical student is probably the single most important factor in the making of a competent, wise, compassionate physician. It is surprising to me that no critical evaluation has been attempted of how we can improve our ability to select good candidates for medical schools.” (Bauer 1956). Indeed, a concerted, systematic and evidence based approach to the selection of candidates for medical school has only manifest over the past 50 to 60 years.
Selection practises have significant consequences for the community and the medical profession. Because of the adverse sequelae that may result from the selection of candidates that fail to become good doctors, selection processes are often focussed on excluding undesirable candidates. However, “the concern is not just about letting through undesirables, but also of excluding potential successes” (Kay 1944). Robust and reliable selection methods are essential to ensure that candidates are assessed fairly and that their strengths and weaknesses are recognised. Selection methods must also be valid in that they must assess attributes that are consistently associated with success.

The purpose of selection is to choose those candidates most likely to become good doctors. The development of valid and reliable selection methods is challenging and depends upon an accurate definition of what constitutes a “good” doctor, the identification of key factors that contribute to the phenotype of a “good” doctor, and the development of selection instruments that produce results that are generalisable beyond the medical course.

This thesis provides an overview of the evolution of selection methods throughout history, the characteristics of a “good” doctor are considered and the ability of key selection instruments to predict success during and beyond medical school is discussed. The multiple mini-interview (MMI) is identified as a promising, commonly utilised, yet under-researched, instrument for assessing desirable non-cognitive attributes in candidates for medical schools around the world. Indeed, the MMI is used by Deakin Medical School as a key selection tool. A series of experiments are undertaken that assess the reliability and validity of the Deakin
MMI. Ways of optimising the reliability, validity and efficiency of the MMI by varying interview duration and interviewer type are then explored.
References


1. The Selection of Medical Students: an Historical Survey

1.1 Early transmission of medical knowledge

For much of recorded history, medical knowledge and skills have been passed from master to apprentice. The accumulation of “medical” knowledge is likely to have begun over 5000 years ago, in the late Mesolithic and early Neolithic periods (Lillie 1998; Merlin 1984). Cartwright recognizes three overlapping spheres of medical practices that developed during the Mesolithic and Neolithic periods (surgical techniques, medical plant and mineral use, and magico-medicine) and suggests that idiosyncratic use would have evolved into a culturally carried body of knowledge (Cartwright 2006). As the body of knowledge increased, it is likely that some individuals were better able to observe, interpret and retain medical knowledge (Schiffeler 1976). Thus, self-selection and natural aptitude were critical factors in determining those who would become early practitioners.

During the Late Neolithic period and Bronze Age, expert knowledge surrounding disease belonged to specific practitioners and developed alongside contemporary socio-cultural beliefs that centred around magic and the supernatural. These practitioners were called shamans in the west and chen-jen in the east (Cartwright 2006; Schiffeler 1976). They possessed expert knowledge and skills that permitted the diagnosis of disease and the prescription of a cure that was disease specific and culturally appropriate.
1.2 Selection of medical students in the East

In the East, tribalism metamorphosed into feudalism during the Shang dynasty (16th-11th century BC). The establishment of a state administration created a shifting socio-political landscape that, in conjunction with the ongoing infusion of new theories and techniques of folk medicine from merging social units, contributed to a rapidly evolving culture and the formation of distinct social classes (Schiffeler 1976). Two distinguishable factions of medical practitioners emerged: the sorcerer-physicians who practised religio-magical folk medicine among the peasants, and the priest-physicians who practised medicine among the nobility. The extension of the priest-physicians into the upper classes permitted them to develop their literary skills enabling them to document and transmit their observations and experiences. Acquisition of information by students of medicine began to require a degree of literary competence that was only attainable by the wealthy (Wang 2006, p22-25).

1.2.1 China

In China, during the Zhou dynasty (1065-256 BC), the medical practices of the priest-physicians became systematized at the level of the state. The Zhou did not endorse magical beliefs and an increasing reliance on experimental evidence allowed the form of medicine practiced by the priest-physicians began to break away from religion and evolve into an independent field (Hong 2004). It was during this period that medicine is first referred to as a ‘profession’ with the establishment of standards for evaluating and paying doctors for their services (Zhang & Cheng 2000).
The first medical schools were formed during the Zhou dynasty. Of note was the founding of the Imperial College of Medicine (Baer, Singer & Susser 2003, p326). Comprised of 30 high-ranking physician scholars, the college had access to an extensive library of medical and scientific knowledge and was responsible for educating junior colleagues (Magner 2007, p74). Training at the Imperial College of Medicine was initially available only to princes, sons of ministers and nobles’ children. However, mechanisms were soon introduced that permitted the entry of highly qualified students recommended by townships or lords (Wang 2006, p28-30). The ruling body conducted yearly examinations to assess the competence of those wishing to practice medicine and performance of these examinations had a profound influence on the eventual rank and salary of graduating physicians (Wang 2006, p74). Medicine strengthened as a profession and came to resemble traditional Chinese medicine. Within the profession, subspecialist physicians, surgeons, dieticians and veterinarians were recognized and described in the *Zhou li (Rituals of the Zhou Dynasty)* (Cai & Zhen 2003, p68; Felt 2007a). Outside the Imperial court, medical training occurred largely by apprenticeship, whilst a variety of ‘lower class healers’ were largely self-taught (Cai & Zhen 2003, p69; Magner 2007, p74).

The Eastern Zhou dynasty (771-256 BC), represented a period of intense philosophical cultivation, during which, the medical profession experienced ongoing refinement in tandem with the development of Confucianist thought. Confucianism, concerned primarily with the ethical and moral behaviour of men, found favour among the priest-physicians who came to be known as the *ju-yi* or Confucian physicians of Chinese society. The three cardinal guides (ruler guides subject, father guides son, husband guides wife), and five constant virtues
(benevolence, righteousness, propriety, wisdom and fidelity) had profound influence on medical education and the medical profession (Cai & Zhen 2003, p68). It promoted increasing responsibility between teacher and student and provided a guide for the types of qualities that were desirable in medical students and doctors. The infusion of Confucian philosophy into medical practice resulted in the development of an ethical and moral framework that outlined acceptable medical practices, responsibility and accountability.

Medical texts outlining the foundations of traditional Chinese medical practice, such as the *Huang di nei jing* (Yellow Emperors Canon of Internal Medicine), appeared in the eastern Zhou dynasty. The *Huang di nei jing* represents the first medical book in China to interweave medical experience within wider ideologies into a coherent system (Galambos 1996). It provides a comprehensive review of the dominant medical theories and practices by medical professionals of the eastern Zhou and Han eras (Hong 2004). The *Huang di nei jing* also provides advice for the selection of medical students. It highlights the importance of selecting the right apprentice and espoused three key principles regarding the selection and teaching of an appropriate apprentice. Firstly, the *Huang di nei jing* provided guidance regarding the attributes of candidates suited to medical training. Suitable candidates for medical training were naturally gifted, perceptive, intelligent, motivated, dedicated to study, virtuous and shared a common goal with the teacher (Wang 2006, p65 (*Huang di nei jing* Su Wen Ch 1, 58, 78, 80. Ling Shu Ch 73)). Suitable students were seen to possess the Confucian ideals of compassion and universal love, as only these students were capable of reinforcing the virtuosity of the medical profession. This is clearly outlined in the *Su Wen* of the *Huang di nei jing*:
The arrangements of yin and yang, exterior and interior, female and male, they store them in their bosom and they link the heart with the essence. If it is not this kind of person, do not teach him. If it is not this kind of truth, do not confer it. (Unschuld & Tesserrow 2011, p 94 (Huang di nei jing Su Wen Ch 4.29-1)).

The Tao is precious and is not to be passed on unless a student is sincere and compassionate toward human suffering. Only in this way can the great tradition remain pure and virtuous. (Ni 1995, p 110 (Huang di nei jing Su Wen Ch 4)).

Secondly, the Huang di nei jing argued that physicians had an obligation to teach suitable candidates and should not teach those who are not suitable. Recognising and teaching medicine to candidates with values that exemplified the Confucian ideals of compassion and sincerity was seen as both a duty and an essential component of a physician’s own development. This principle is clearly espoused in both Su Wen and Ling Shu:

To meet the right person but not teach him is called losing the way.
To transmit the doctrine to someone who is not the right person is, as to treat the heavenly treasure without respect (Unschuld & Tesserrow 2011, p242 (Huang di nei jing Su Wen, Ch 69.401-5)).
Teachers have a duty to teach those who display aptitude for medicine and failure to do so will undermine the future of medical education (Wang 2006, p66 (Huang di nei jing Su Wen Ch 69, 80, Ling Shu Ch 64, 73)).

Thirdly, as outlined in the Ling Shu, the Huang di nei jing states that teachers should adapt their teaching to the strengths or aptitudes of their apprentices:

For example, a student with great sight should be taught to diagnose the patients by “watching the body”; those with sensitive ears should be taught to diagnose patients by “listening to the pulse”; good communicators should be taught to be medical “explainers”; those with powerful fingers, should be taught to administer physical therapies (Wang 2006, p67 (Huang di nei jing Ling Shu Ch 73).

These principles illustrate the significance of selecting appropriate candidates into medical training so as to maintain the integrity and diversity of the medical profession.

The unification of China by the Han in 206 BC ended many centuries of warfare and permitted an unprecedented period of renaissance. Chinese medical thought underwent a process of comprehensive standardization and systematisation. The organization of the Imperial College of Medicine was further developed in the Qin and Han dynasties. As the infrastructure of the Imperial court grew in complexity, Imperial Physicians were granted aides and became associated with the Shao Yu (Chamberlain for Palace Revenues) (Cai & Zhen 2003, p68).
Outside the capital, the medical profession was fragmentary and medical training continued to follow the apprenticeship model (Wang, Chen & Hsieh 1999, p147). However, the development of a standardized, rational basis for the practice of medicine within a framework of Confucian values where the emphasis was on benevolence, caring about patients and self-cultivation in virtue brought with it synergistic coupling of expert knowledge and ethical loyalty (Zhang & Cheng 2000). Zhongjing Zhang (150-219 AD), an eminent physician of the time suggested that a detailed knowledge of medical theories and treatments was required for physicians to achieve their goal of loving the people (Zhang, Z 1963). As such, it was essential that students accepted into medical apprenticeship possessed the cognitive and non-cognitive characteristics required to acquire sufficient knowledge and to apply it appropriately. The strength of the Confucian culture was such that the selection of medical students, medical training and professional conduct were regulated by practitioners through self-reflection rather than by prescribed standards (Zhang & Cheng 2000).

Support for Confucianism fell at the end of the Han dynasty and Taoism and then Buddhism appear as major influences on social and medical philosophy until the gradual re-emergence of Confucianism during the Tang and Sung dynasties (Unschuld 1985, p155-166). During the Six dynasties period (220-581 AD) the influences of Taoism and Buddhist thought reintroduced supernatural and religious concepts to medical practice (Unschuld 1985, pp116, 120-131). Within the Imperial Court, the medical faculty remained well structured. The Imperial Physicians of the Imperial Medical Office were renamed Medical Scholars and a Director of Palace Medication was appointed. A rudimentary Medical College was
established in the capital in 442 AD in an effort to standardize the quality of physicians who served the Emperor (Hong 2004). The Medical College included two doctors and two professors in each of the specialties of medicine, massage and incantation, and 200 physicians in charge of medical work (Wang, Chen & Hsieh 1999, p148). During the Sui dynasty (582-617 AD), the scale of the Imperial medical faculty further expanded.

Outside the Imperial Court, medical training continued by apprenticeship, however the loss of a unified, culture-based set of professional standards resulted in increasing variation in the characteristics of students accepted into apprenticeships and the quality of medical training. This social status of medical practitioners fell and doctors were frequently described as ineffective, potentially harmful and no better than the “fiendish shamans” (Qian 1998, p398-399). Qian suggests that the public view of medical practitioners occurred because doctors were not worthy of their trust (Qian 1998, p399).

In the Tang dynasty (618-906 AD), the organization of the medical profession continued to develop within a sociocultural framework that was increasingly influenced by Buddhist thought and the re-emergence of Confucianism (Unschuld, 1985, p157). During the Tang dynasty, the scale of the medical education system was expanded into a systematized national program administered centrally that was, however, limited in its effectiveness (Wang, Chen & Hsieh 1999, p148).

Within the Imperial Court, the status of the medical faculty rose as the Tang emperors turned towards their physicians in their search for alchemical immortality.
(Felt 2007c). Although most doctors were male, the Tang palace supported at least ten women who were trained to handle herbal medicines. Although selection practices are unclear, some insight is gained from the Seiji yooryaku, which describes the recruitment of female doctors in Heian Japan (794-1183 AD) and quotes frequently from early Tang processes (Lee 2003):

Female doctors are chosen among official slaves. Thirty of those aged between fifteen and twenty-five and who show their talent in comprehension will be allocated separately.

[This] says: the Section of Inner Palace Medicine, naiyakushi will build separate quarters for them. They will be instructed on matters such as calming the fetus, helping in childbirth complications, and healing wounds, swelling, broken limbs as well as methods of acupuncture and moxibustion. These are taught to them through oral education (Inryoo 1964, p701).

Outside the Imperial palace, medical care was often unsatisfactory. In response, the Emperor Xuanzong ordered each province to install a yixue boshi (medical official) and to designate specified quotas of yisheng (medical students). Provinces with more than 100,000 households were to designate twenty yisheng, whilst provinces with fewer than 100,000 households were to designate twelve (Chao 2009, p132). Women were increasingly excluded from medical training and indeed from even collecting or touching certain medicines as it was felt that they were liable to render the medicines ineffective (Lee 2003, p21).
The suitability of candidates for medical training was ascertained through consideration of each candidate’s social status and the alignment of a candidate’s personal qualities with contemporary sociocultural values and the dominant model of medical morality. Medical texts of this period demonstrate a convergence of Confucian, Taoist and Buddhist concepts and the embedding of the principles of universal love, moderation and compassion within the medical ethos (Wu 1980). The merging of these cultural values into a code of medical morality is illustrated by Sun Si Miao in his seminal work *Bei ji qian jin yao fang* (*Emergency prescriptions worth a thousand pieces of gold*). The volume contains two key chapters *Da yi xi ye* (*Practices of an excellent physician*) and *Da yi jing cheng* (*Good faith of an excellent physician*) that outline the requirements for a physician in medical practice. Cai and Zhen distil these requirements into 4 broad principles:

1. *Be assiduous and love the profession*
2. *Do your best to save lives regardless of payment and rewards*
3. *Work in proper style*
4. *Adhere to science and fight against superstition* (Cai & Zhen 2003, p70).

A section of the *Bei ji qian jin yao fang* also describes the requisite qualities for students of medicine. Once again, Confucian, Buddhist and Taoist concepts are interwoven, requiring medical students to be culturally and intellectually prepared, to live moderate lives and to be virtuous:

> Whoever wants to be a doctor . . . must understand yin and yang and be able to discern life’s fortunes (read people’s faces and see their fates). They must
also understand the cracks in the tortoiseshells of Zhou Yi (Yi Jing) . . . without such knowledge they will be like a blind person in the dark; they will fall down easily. You must also engage in other reading. Why? If you do not read the Five Classics, you will not understand justice, humanity and virtue. If you do not read the Three Histories, you will not know the past and the present. If you do not read the exponents of the various schools of thought, you will not understand what is happening in front of your very eyes! If you do not read the Nei Jing you will not know the virtue of mercy, sorrow, happiness, giving. If you do not read Zhuang Zi and Lao Zi, you will not know how to conduct your daily life. As for the theory of the Five Phases, geography, astronomy . . . you also need to study these. If you can study and understand such knowledge, there is no hindrance on the road of medicine. You can become perfect (Zhang & Rose 2010, p3).

Growing corruption and disintegrating administrative control led to the collapse of the Tang dynasty in 907 AD. Fifty-three years passed before peace was restored to the nation by the first Song emperor. It was during the Song dynasty (961-1279 AD), that the medical faculty of the Imperial Court was restructured. The Imperial Medical Office was created in an effort to standardize medical training and ensure its effective transmission throughout the Empire. The apprenticeship model of medical training gradually gave way to a centrally-based system overseen by the Imperial Medical Office. Regular medical schools were organized in the capital and in other parts of China (Calman 2007, p20; Hong 2004). The social status of doctors rose as the medical profession received increasing recognition from successive Emperors and efforts were made to improve medical training and
regulatory oversight (Goldschmidt 2009, p28). The number of available training places was increased, entrance and qualifying examinations were standardized and official titles were bestowed on graduates (Goldschmidt 2009, pp48-52). The medical profession was seen as a respectable alternative to positions within the governmental bureaucracy. These changes increased the attractiveness of medical training and competition for training positions became fierce. As the number of candidates increased, quotas were imposed and the minimum age for study was increased from 15 to 30 (Calman 2007, p21; Goldschmidt 2009, p47). By the end of the Song Dynasty, all candidates for entry into medical school were to be above 30 years of age, of good medical knowledge, high moral character and esteemed by their friends (Calman 2007, p21). Candidates were required to provide details of their family, to provide a recommendation from an official serving in a medical position, and sit formal oral entrance examinations to demonstrate competence in medical diagnosis and treatment (Goldschmidt 2009, p47, 54). Upon admission, students faced progress tests and qualifying examinations. Successful graduates were given a license to practice whilst those who failed were ordered to change profession (Calman 2007, p21).

The Mongols controlled China during the Yuan dynasty (1264-1368 AD) and medical training during this period appears disorganized and fragmentary with little development within medical training (Felt 2007b). The return of China to the Chinese at the beginning of the Ming dynasty (1368-1643 AD) bolstered national pride and brought with it renewed interest in Chinese culture and Confucian philosophy. The academic knowledge of medical scholars in the Jin and Yuan dynasties combined with the philosophy of Neo-Confucianism in the Ming and
Qing dynasties fostered a revitalization of medical practice (Meng 2002). This culminated in the creation of the Imperial Academy of Medicine. Composed of 13 departments, it remained the dominant medical organization throughout the Ming dynasty (Cai & Zhen 2003, p68-69). Different positions were established for medical professionals including *yuan shi* (Director of the Academy), administrative assistant, imperial physician, medical secretary, medical official, physician and student of general medicine. Physicians and students of general medicine were selected from those who had passed their examinations with excellent results and qualified (Cai & Zhen 2003, p69).

The Imperial Academy of Medicine existed as an isolated centre of medical excellence. In the absence of a standardized national teaching program, or an academic body that could effectively set the norm for medical learning, the apprenticeship model remained the predominant model for medical education in China from the 15th to early 20th century (Leung 2003). The ability of potential students to comprehend an increasing number of medical introductory tests is likely to have been a major factor determining their suitability for medical training (Leung 2003). Leung describes an example of an aptitude test imposed upon an aspiring medical student, by a potential teacher and mentor, in Hangzhou in 1911. The test involved a series of questions on the content of four of the most common medical introductory texts of the Ming-Qing period (spanning from the fifteenth to the eighteenth centuries) that had been given to her three months earlier (Leung 2003).
Traditional Chinese medical practice entered a period of relative decline as medical texts from Europe reached China during the 18th and 19th centuries (Hong 2004). The first Western-style medical school opened in China in 1886 (Felt 2007b). The establishment of the Peoples Republic of China by the communist party in 1949 brought with it a revolution in medical education. This involved the development of a two-tiered medical education system that was based on the Soviet model of medical education. Medical training was offered via 3-4 year secondary-level training courses carried out in secondary medical schools and via higher-level, 5-6 year university courses. University medical courses included an intern year and graduates were eligible to practice medicine. In contrast, graduates of secondary medical school were required to undertake an intern year before being eligible to practice and graduates were expected to practice in rural areas (Fox 1984). The secondary medical schools greatly outnumbered the university medical schools and by 1965 there were 298 secondary medical schools and 92 medical universities (Reynolds & Tierney 2004).

Entry into medical training was dependent on academic achievement and performance on specific entrance examinations. Successful completion of middle school was required for entry into secondary-level training courses, while high school graduation was required for entry into upper-level training courses at university (Reynolds & Tierney 2004).

Entrance examinations for medical training were in place in China between 1949-1965 and reinstated in 1977 after the cultural revolution (Gao et al. 1999). During the cultural revolution selection processes based on academic performance were
abandoned. “Peer-chosen, worker-peasant-soldier” students were selected for entry into courses that were highly practical and grounded in political ideology (Reynolds & Tierney 2004).

The tiered system of medical education in China was phased out towards the end of the 20th century. Secondary medical schools ceased to exist as increasing resources were funnelled into university training to cater for increasing societal expectation for highly training doctors (Schwarz, Wojtczak & Zhou 2004).

Selection into medical training now occurs after Senior Middle School, but on rare occasions, highly performing high school students from selected schools may be permitted direct entry into medical training. Candidates require a strong background in science and most candidates must sit the National College Entrance Examination (NCEE), a standardised university entrance examination that covers mathematics, physics, chemistry, Chinese and English (Schwarz, Wojtczak & Zhou 2004). Candidates are required to submit their preferences for places of study and are primarily selected on the basis of their preferences and performance on the NCEE (Schwarz, Wojtczak & Zhou 2004). In some cases, entrance scores may be lowered for candidates who agree to work for a defined period in an area of need. Additional factors that may be considered during selection include letters of recommendation from high schools and selection interviews (Schwarz, Wojtczak & Zhou 2004). When instituted, interviews tend to focus on the linguistic and intellectual abilities of candidates (Schwarz, Wojtczak & Zhou 2004).
Many universities consider only candidates within their regional catchment area. Competition for positions in the leading universities is intensely competitive as these institutions are in high demand and consider candidates from all over China (Schwarz, Wojtczak & Zhou 2004).

1.2.2 Japan

Until the introduction of Western medicine in Japan in the mid-19th century, medical practice in Japan resembled traditional Chinese medicine and medical training followed the apprenticeship model (Izumi & Isozumi 2001). In 1824, the Narutaki Cram School was opened by Philipp von Siebold, a German ophthalmologist and likely represents the earliest school providing a Western style medical education (Izumi & Isozumi 2001). However, formal medical classes began in 1857 at the Western Nagasaki Magistrate’s Office, following the appointment of a Dutch naval surgeon, Pompe van Meedervoort, as surgeon of the Japanese Navy teaching group. The course was for military personnel and not a dedicated medical course, and one non-military student, the son of the founder of the first private hospital in Japan, was admitted into the course (Izumi & Isozumi 2001). As the number of applications from non-military students increased, the first Japanese medical school (Igaku Denshu Jo) was established in Nagasaki. The entry of students into the medical course was controlled by feudal clans who sent pre-selected members for medical training (Izumi & Isozumi 2001).

The Meiji reformation was associated with the adoption of the German model of medical training and the establishment of medical schools in Kagoshima and Tokyo in the 1870s. During the early years of the Tokyo Medical School, there was little restriction on entry. This resulted in students that were too numerous and
inadequately prepared for the study of medicine (Bowers 1965). In response, quotas were imposed and rigorous selection processes were implemented to select the most gifted and industrious students. Candidates were required to have completed 15 years of primary and secondary school educations and were selected based on performance on an examination designed to test their general ability and cultural achievements (Bowers 1965). Once accepted into the course it was “most unusual for any student to be unsuccessful” in graduating from the course (Bowers 1965).

During the following decades a university medical school was established in Kyoto and a number of smaller private or municipal medical schools opened. This gave rise to a two-tiered system of medical education administered by universities and technical colleges. Although the duration and content of each type of course was similar, a higher degree of previous academic achievement was required for entry to the university medical courses (Schofield 1917). Instead of the 15 years of primary and secondary school education required for entry into university, admission into medical courses at technical colleges was possible after 12 years of study. Motivation to enter university medical courses was strong as graduates of universities were awarded higher military ranks and greater administrative opportunities than graduates of technical colleges (Bowers 1965). By 1920, seven university medical schools were in operation in Japan and applications for training positions far outnumbered available places (Bowers 1965; Schofield 1917).

Medical education blossomed after the First World War, mainly due the expansion of medical schools associated with the technical colleges. Competition for places in medical courses at both university and technical colleges remained strong.
However, the defeat of Japan in World War II brought with it a dramatic restructuring of Japanese medical education. The Council on Medical Education was established and a single university standard of medical education was established (Bowers 1965). Medical schools at technical colleges were required to meet university standards or close. In many cases, technical colleges were able to attain the required standards via prefectural or private sponsorship and continued to educate medical students. A national university entrance examination was introduced in 1978 to standardise entry into Government run (National and Prefectural) tertiary institutions. The test became mandatory at all Government universities and has also been used increasingly by private universities. The test is largely multiple choice and addresses a variety of subject areas. Universities may specify those subjects required by applicants for entry into particular courses. At both Government and private universities, university-specific tests for entry into medicine continued to be used in conjunction with the standard national exam to select candidates into medical training. Licensing examinations were also introduced to ensure a minimum level of competence.

There are currently, 42 national, 8 prefectural and 29 private undergraduate medical schools in Japan. High school graduates wishing to enter medical training are required to sit two entrance exams. The National Centre for University Entrance Examination (NCUEE) is taken by all applicants for entry into all Japanese national and prefectural Universities, and most private Universities. Applicants for medical school are required to complete components of the test relating to Japanese literature, English, geography or history, mathematics and at least two of physics, chemistry and biology. Scores in excess of 90% on the examination are generally
required for entry into medicine. Universities also required candidates to sit their own university-specific entrance examinations. These examinations commonly assess English, mathematics, physics, chemistry and biology. Examinations for entrance into medical school are more difficult than those for other courses, English and mathematics (Bowers 1965). The level of difficulty is set by the administering institution and those at Tokyo and Kyoto are known to be the most difficult (Bowers 1965). Ronin schools are available to help prepare students for the University Entrance Exams.

Most medical schools also require applicants to complete an essay and most also conduct selection interviews. Other than the requisite for candidates for national and prefectural medical schools to sit the NCUEE, the selection processes employed by medical schools, and the way they are used to rank candidates for admission remain at the discretion of each school.

1.2.3 India

In India, a gradual shift from magico-medical practice towards a more systematized approach based upon an expanding bank of medical knowledge also occurred. Although medical institutions, akin to those in the West and Far East developed, the apprenticeship model represented the dominant model for medical training. Medical knowledge was passed from master to apprentice and recognized by formal degree (Puschmann 1966, p7). The selection of apprentices was the responsibility of established physicians who tended to favour those of higher social status (Puschmann 1966, p7). There was also an increasing emphasis on selecting
those students into medical training with particular personal qualities considered
noble or virtuous.

The Ayurveda, *Sushruta Samhita* provides guidance, advising that medical training
should be available to noblemen, priests and freemen. Ideally those suited to
medical training were those who were:

*Calm, of noble nature, not indulged in mean acts, with good-looking eyes,
mouth and nasal ridge; having thin, red and clear tongue; with no
abnormality in teeth and lips, not speaking with nasal utterance, having
restraint, without vanity, intelligent, endowed with reasoning and memory,
with broad mind, born in a family of physicians of having conduct like that,
having insistence for the truth, without any deformity or impairment of
senses, humble, un-haughty, having ability to understand essence of the ideas,
without anger and addictions, endowed with modesty, purity, good conduct,
affection, dexterity and sincerity, interested in study, devoted to
understanding of ideas and practical knowledge without any distraction,
having non greed or idleness, compassionate to all creatures, following all
the instructions of the teacher and being attached to [him] (Van Loon 2003,
p111).

The *Charaka Samhita* also recognized the reciprocal relationship between student
and teacher and advised youths considering medical study that a teacher should
have a:
A clear idea of the subject, should have seen the practical applications, be skilful, amicable, pure, having practical experience, well-equipped, possessing all the senses in normal condition, acquainted with human constitutions, well-versed in courses of actions, having his knowledge uncensored, free from conceit, envy, anger, forbearing, fatherly to disciples, having qualities of a good teacher and capable of infusing understanding. The teacher possessing such qualities inculcates physician’s qualities in his disciple in a short time like the seasonal cloud providing good crop in a suitable land (Van Loon 2003, p110).

Thus, it was clearly recognized that the production of a good physician depended upon the qualities and qualifications of both teacher and student and that the ‘utmost care should be exercised in the selection of both the teachers and the taught’ (Sen 1988, p14). The significance of the relationship between master and student in Ancient and Medieval India is described by Mookerji in the following way:

*It is not like the admission of a pupil to the register of a school on the payment of his prescribed fee. It is a spiritual initiation into a new life, for the pupil is now a twice born (doija), inasmuch as the teacher impregnates him with his spirit* (Mookerji 1989, pxxxvi).

Once selected into training students were expected to follow the professional code of their masters. Students were cautioned to know their limitations, to be humble, to collaborate and commit to ongoing learning, and to devote themselves to medical
practice and their patients (Van Loon 2003, p112-113). Training began at age 16, generally lasted 6 years, and involved theory and practical components. Upon completion the student was expected to seek permission of the king of the country for commencing his medical or surgical practice (Sen 1988).

Unlike the situation in the Far East where the evolution of Chinese medical thought and medical training occurred largely without influence from developments in the West, the subsequent development of Indian medicine and medical training was increasingly influenced by Greek, Roman and Arab medicine. Entry into medical training did, however, continue to be heavily influenced by cultural values (Puschmann 1966, p7-18). The consolidation of Muhammadan power at the turn of the first millennium AD, brought with it an influx of ideas, including models for medicine, science and education, from Arab, Christian and Jewish sources (Sheehan & Hussain 2002). Fusion of Western and Ayurvedic medical thought culminated in the development of the Unani Tibb (Sen 1988). Medical training continued to be largely tutor oriented with teachers able to select students based upon recognition of qualities considered desirable within Greek and Ayurvedic philosophy (Jalil 1978). As in the West, religion also played an important role in shaping the selection and training of medical students. Monasteries and universities developed by Buddhist monks predate those of Europe and, within them, senior monks were responsible for choosing and instructing students in a variety of disciplines that included medicine (Sen 1988).

At the time of the Mughal Empire (1526-1857 AD), Ali Nadeem Rezavi describes a medical training system characterized by few schools of medicine and dominated
by the apprenticeship model (Ali Nadeem Rezavi 2001). In addition, important
centres of medical education during the sixteenth and seventeenth centuries were
located in neighbouring countries such as Iran (Ali Nadeem Rezavi 2001, p41). The
number of medical practitioners rose considerably in India during the 16th and 17th
centuries due to Imperial support and an increasing demand for medical services
(Singh 2006). Medical training was readily available, although in the absence of
regulatory oversight and standardized admission, training and assessment
procedures, the practice of medicine was:

> Open to all Pretenders, here being no Bars of Authority, or formal
Graduation, Examination or Proof of their proficiency; but everyone
ventures, and everyone suffers, and those that are most skilled, have it by
tradition, or former Experience descending in their families (Fryer 1985,
p114).

However, measures were put into place in order to improve the quality of medical
care. Such measures included, employer specific pre-employment tests and the
‘sending out’ of doctors from famous medical schools such as Sirhind (Ali Nadeem
Rezavi 2001, p41). This appears to have been associated with a progressive rise in
the prominence of the medical profession and the development of different classes
main classes of medical practitioners; vaidyas and hakeems (local practitioners)
that received apprentice-based education, ambitious physicians and surgeons who
received additional university and hospital-based training, and general medical
The Mughal Empire weakened during the eighteenth century due to political instability, decentralisation of government authority and the colonisation of India by Imperial Britain. The political disintegration of the country led to the deterioration of institutionalised education, and the re-establishment of apprenticeship as the dominant form of medical training with selection of apprentices dependent upon the availability and judgement of established practitioners (Montgomery 1888). Literacy in Greek, Arabic, Sanskrit or Persian was required by students of medicine as part of the education process involved the study of Greek and Arabic texts or their translations (ibid). There was however little regulation of practitioners or their conduct, and medically trained practitioners existed alongside uneducated and/or religio-magical practitioners. This combination of factors resulted in a situation where the reputation of the medical profession fell, such that:

The science of medicine and its practitioners in India had reached the lowest depth. The physician was an out and out quack and the only surgeons were barbers or old women (Anonymous 2000, p28).

Western medicine was introduced into India from the 16th century by the Portuguese and the British (Pandya 1982). Attempts at teaching western medicine to the natives started around the middle of the 16th century (Gracias 1941). Its contribution towards strengthening Portuguese diplomatic relations with Indian courts (de Escola 1941, p221-292). Jesuit priests at the Collegio de Sao Paulo dos
Arcos or the Seminario da Santa Fe, founded in 1541 taught medicine along with theology, mathematics, astronomy and philology (Pereiera 1980). Medical education was available to natives who were selected and funded by regional communities in need of medical care (Pandya 1982).

A formal three year course was introduced in Goa by the Portugese in 1801, and had developed into a formal medical school, the Escola Medice-C'irurgica de Nova Goa, by 1842 (Pandya 1982). Meanwhile, the British had established a seat of medical learning in Calcutta that led to the formation of the Calcutta Medical School in 1835 (Supe & Burdick 2006). The Calcutta Medical School was established to train native youths aged between 14 and 20 irrespective of caste and creed in the principles of medical sciences with the mode adopted in Europe (Palit, Dutta & Corpus Research Institute 2005, p276).

Nevertheless access of native Indians to Western medical education was reduced by the formalisation of medical education in the 19th century. University-affiliated medical education became the norm in the 1850s, after the opening of the first three Indian universities in Chennai, Mumbai, and Calcutta (Supe & Burdick 2006). Opportunities for medical education in these institutions were made available to students who mostly belonged to the very privileged upper class of the society, and often only Europeans and converts to Christianity. Secondly, the Medical Council of India accepted the British norms of medical education in order to gain recognition of Indian medical degrees from the British Medical Council. This enabled some of the physicians, who were ‘the select among the select’, to go to Great Britain to get higher medical education (Banerji 1973). Women were initially excluded from medical education, however, during the second half of the 19th
century Medical Schools at Madras, Bengal and then Calcutta opened their doors to female medical students—provided they had passed standardised academic entrance requirements (Palit, Dutta & Corpus Research Institute 2005, p273).

At the time of independence from Britain (1947) there were 19 medical schools with an output of 1200 doctors (Richards 1985). Admission requirements were variable across the country although common criteria for admission included age of at least 17 and performance in an ‘intermediate examination’ taken two years after high school matriculation that assessed physics, chemistry, zoology and botany (Misra 1954). Some medical schools required candidates to sit additional premedical examinations in the same four key disciplines and panel interviews where candidates were examined physically and engaged in discussion on topics of “general interest” (Misra 1954). In his review of selection methods across India, Misra (1954) voices some concern about the dominance of scholastic achievement in determining entry into medicine.

The 1960’s were significant for the addition of an intern year to the standard 4.5 year medical course and marked the beginning of a dramatic increase in the number of medical schools in India to 86 by 1965, 114 by 1985 (13600 enrolments) and 258 by 2006 (27677 enrolments) (Richards 1985; Supe & Burdick 2006).

In 1985, the basic academic entry requirement for medicine was an aggregate of 50% in the pre-degree 10+2 examination (12th standard) in biology, physics, and chemistry. In addition, students had to be 17 by the 31st of December of the first year of the course. Some medical schools, mostly those run privately or by central
government, held their own entrance exams, and some, such as the Christian colleges, have additional religious requirements to fulfil. In Vellore, for example, 75% of places were reserved for students who have been sponsored by different churches in India. These students were then required to return to their own state and work in a mission hospital for two years before continuing with their chosen career (Richards 1985).

Since then, selection into medical training at state-owned universities has continued to be based largely upon academic performance. In an effort to standardize entrance requirements, the All India Premedical Test (PMT) was introduced during the 1990s and was established as the major tool for selection. The PMT consists of two parts, a preliminary examination consisting of 200 objective type questions (four options with single correct answer) from physics, chemistry and biology (botany & zoology), followed by a final examination of 120 objective type multiple choice questions (MCQs) for those who pass the preliminary examination. The PMT takes place once a year after high school graduation examinations. It is offered by each state to its residents. A national examination is also offered to allow students from one state to apply for admission in another. In order to be considered for selection, applicants must have passed in the subjects of physics, chemistry, biology and English individually and must have obtained a minimum of 50% of marks taken together in physics chemistry and biology (Medical Council of India 1997). Qualifying scores for admission are lower for individuals from socially disadvantaged groups so as to increase their representation (Medical Council of India 1997).
Currently, students are accepted to their preferred medical school based on their PMT ranking and availability of seats in the school. From 2013, the PMT will be replaced by another standardised test, the National Eligibility-cum-Entrance Test for MBBS Course (NEET) (Medical Council of India 2012). Like its predecessor, the test examines knowledge of physics, chemistry and biology and passes in all components, and a minimum of 50% of marks taken together in physics chemistry and biology, will be required for consideration for selection (Medical Council of India 2012).

Interviews are rarely used in the selection process, and other candidate traits are not significantly considered by most schools. Some points in the scoring system for admission may be awarded for community service, sports, and military service (Sood & Adkoli 2000).

Selection processes for medical courses in India have been criticised for their reliance upon marks obtained by applicants on the PMT, the validity of which has been questioned as has its focus on factual recall and the failure of the selection process to consider attitudes, and communication skills which form essential traits of any health professional (Sood & Adkoli 2000). In response, the Medical Council of India has recommended that merit in the standardised entrance tests should be combined with an aptitude test so as to form the criteria for selection tests (Medical Council of India 1996).

1.2.4 Singapore

In the late 19th century Singaporean candidates for medical education were selected into training at the off-shore Medical College in Madras by a three stage approach
that considered personal details, academic performance and aptitude. Candidates were required to be between 16 and 19 years of age, of suitable character and respectability, and of adequate fitness (Tambyah 2005). Academic merit was assessed by way of a standardized test that included 7 components: (i) an exercise in dictation and handwriting, (ii) a colloquial examination of either Hindustani or Malay, (iii) ancient and modern history, (iv), general geography, (v) arithmetic, vulgar and decimal fractions and proportion, (vi) algebra and (vii) the first book of Euclid (Tambyah 2005). Candidates found suitable in an examination to test their educational qualifications would enter a 1 or 2 year hospital training course at Singapore General Hospital (Lee 1978). Successful completion of the course resulted in acceptance into a 3 year medical course.

In the early 20th century, the first Singaporean medical school was founded. Selection requirements were largely academic and all students were required to have passed Junior Cambridge English composition, geography and Mathematics. Requirements for English composition were soon raised to the Senior Cambridge, and the list of compulsory subjects was adjusted to include English, mathematics, Latin and a modern language, which could include French, German, Malay, Chinese, Tamil or Hindustani (Lee 1978).

In 1923 the Cambridge Senior Local Examinations (the predecessors of the A level examinations) or either of the London matriculation or the Hong Kong matriculation were adopted as the measure of academic performance. Candidates were required to have completed English, mathematics, a language other than English, and one or more from history, geography, physical science, natural
science, advanced mathematics, or a variety of recognised languages (Tambyah 2005). An additional requirement for the completion of A level chemistry was introduced in the 1960s (Tambyah 2005).

Between 1960 and 1980, medical students, like all students at the University of Singapore, were required to produce a “certificate of political suitability”, presumably to exclude students of an extremist persuasion, prior to enrolment (Tambyah 2005).

Admission to medical training in Singapore is handled through a Joint Admissions Office, with applicants expressing preferences for particular courses (Harman 1994). Selection continues to be based primarily on academic achievement (performance in the Singapore-Cambridge Ordinary (O) and Advanced (A) Level examinations or their equivalent) (Harman 1994). Minimum academic A level requirements are a minimum of two good passes in two relevant subjects, plus a satisfactory grade in a general paper. However, academic requirements are driven up by intense competition.

There have been recent efforts to reduce the dependence on academic performance for selection. All candidates must now undergo a Vocational Assessment Scheme and interview during the selection process for medicine (Harman 1994). Interview panels contain broad representation from key stakeholder groups that include physicians, academics, representatives from the Ministry of Health, junior doctors, senior nurses and senior medical students (Tambyah 2005). In addition, the University of Singapore requires all candidates to take the Singapore University
Medial Entrance Test (SUMET) and complete an on-the-spot essay assignment (Tambyah 2005), The National University of Singapore requires candidates to attend two independent interviews and to complete an essay and portfolio review comprising two letters of reference, a personal statement and curriculum vitae (National University of Singapore 2009). Recently, the National University of Singapore has provided a route for entry that provides for candidates with exceptional ability in a “non-A-level” field of achievement (Tambyah 2005).

All candidates for the medical course are also required to provide proof that they have satisfactorily completed a course of immunisation against hepatitis B, or show that they are not infectious (Tambyah 2005).

1.3 Selection of medical students in the West

Medical training by apprenticeship was the norm in the ancient civilisations of Egypt, Greece and Rome and, despite the gradual rise of university teaching and ongoing attempts to standardize the selection and training of medical students during the latter part of the first millennium, remained an important method of medical instruction until relatively recently. Under the apprenticeship model, students acquired knowledge, practical skills and professional etiquette from a local practitioner (Beck 2004). Students were selected by teachers and learned medical practice through observation, literary study and by assisting their teacher (Pikoulis et al. 2008).
1.3.1 Ancient Egypt

Egyptian medicine was a unique mixture of magic, religion and scientific fact. Doctors, priests and magicians were all involved in healing and many doctors also held titles of priest or magician (Nunn 1994). The *Ebers papyrus* includes the following section that highlights the parallel roles of doctors, priests and magicians in health (Ebbel 1937, 854a: 99, 2-5):

*There are vessels in him to all his limbs. As to these: If any doctor, any wab priest of Sakhmet or any magician places his two hands or his fingers on the head, on the back of the head, on the hands, on the place of the heart, on the two arms or on each of the two legs, he measures [or examines] the heart because of its vessels to all his limbs. It [the heart] speaks from the vessels of all the limbs.*

Apprenticeship was the major route of medical teaching. Transmission of knowledge from father to son is believed to have been common. However, there are only two known examples of a son following his father as a doctor in Ancient Egypt (Nunn 1994). Papyri containing medical information were copied and stored in the *per ankh* (house of life), a recognized seat of knowledge. Literacy was thus a requirement as much of the required learning required the ability to read papyri and it was not uncommon for scribes to become physicians (Nunn 1994). Though most physicians were men, female physicians existed as well and industry and talent have been described as the only real selection criteria for those wishing to study medicine (Calman 2007, p25). Students were expected to commit to learning, to
demonstrate diligence in their studies, to put their study before personal pleasures, and to develop their communication skills (Puschmann 1966, p19).

1.3.2 Ancient Greece

Transmission of knowledge and skills through mentoring and apprenticeship was highly regarded in ancient Greece; indeed, it was the means by which Chiron and Asclepius passed medical knowledge on to their sons and apprentices (Pikoulis et al. 2008). In Grecian society, the transmission of medical knowledge through apprenticeship continued from father to son or assistant. Masters chose apprentices and study took place within an apprenticeship that was bound by an agreement or oath (Calman 2007, p31). The agreement created a father-son type relationship that brought with it benefits for both master and apprentice. In principle, the pursuit of medical knowledge was virtually unrestrained, and anyone could become students of medicine provided they could learn from a practising doctor (Pikoulis et al. 2008). This principle was captured by Plato in his Laws (Taylor 1934, BookIV: 720):

_All bear the name [of physicians], whether freemen, or slaves who gain their professional knowledge by watching their masters and obeying their directions in empiric fashion, not in the scientific way in which freemen learn their art and teach it to their pupil._

However, selection as an apprentice depended upon the judgement of an available and willing mentor. Characteristics sought in potential apprentices included intelligence, a firm grasp on reality and dedication to study (Pikoulis et al. 2008).
Given that physicians relied upon reputation for securing work, motivation was strong to select good apprentices and to produce good doctors. Indeed, successful and experienced physicians would take care in recruiting, selecting and training their apprentices (Pikoulis et al. 2008).

As influential physicians developed their own theories of medicine, devoted students and followers gathered around them (Drabkin 1957). Medical schools appeared in Greece as collections of like-minded physicians and their apprentices from approximately 500 BC (Pikoulis et al. 2008). Different schools were based on different philosophies of medical education and healing and practitioners of different schools are likely to have valued different traits in potential apprentices.

Hippocrates described the desirable characteristics sought by the School of Cos in the Canon (Chadwick 1978, p68):

> For a man to be truly suited to the practice of medicine, he must be possessed of a natural disposition for it, the necessary instruction, favourable circumstances, education, industry and time. The first prerequisite is a natural disposition, for a reluctant student renders every effort vain. But instruction in the science is easy when the student follows a natural bent, so long as care is taken from childhood to keep him in circumstances favourable to learning and his early education has been suitable. Prolonged industry on the part of the student is necessary if instruction firmly planted in the mind is to bring forth good and luxuriant food.
Desirable non-cognitive attributes included a natural disposition for the study of medicine, favourable circumstances, a minimum level of education, industry, and time (Chadwick 1978). Hippocrates also recognized the need for doctors to be intelligent, empathic and proficient communicators (Adams 1972). Candidates suited to the study of medicine were those with ‘a natural disposition, instruction, a favourable position for the study, early tuition, love of labour’ and ‘natural talent’ (Hippocrates 1910, Vol. 38, Part 1, of 8, Sect 2). Students must also have the ability to reflect on instruction and experiences and should be respectful to their teachers. Hippocrates saw these characteristics as pivotal in re-establishing medicine as the most noble of all the arts (Hippocrates 1910).

Despite these practices, the broadly held view of the medical profession in Ancient Greece was that doctors were relatively low class labourers or craftsmen, offering a service for pay (Chang 2008). Some considered medical students to be of the same calibre as prostitutes, as illustrated by Aeschines description of Timarchus (Adams 1919, pp40-2):

> First of all, as soon as he was past boyhood, he settled down in the Peiraeus at the establishment of Euthydicus the physician, pretending to be a student of medicine, but in fact offering himself for sale, as the event proved.

This situation is likely to have been perpetuated by the selection practices of some schools that attracted students indiscriminately with claims of quick apprenticeships and simple instruction, and compounded by a lack of regulation (Chang 2008; Drabkin 1957). Indeed, in Ancient Greece the personal motivations and
characteristics of apprentices can be observed by the division of apprentices into those studying medicine as a craft, as an academic discipline or as part of a broader education (Pikoulis et al. 2008).

1.3.3 The Roman Empire

As more candidates applied to physicians associated with medical schools, the number of apprentices to physicians or professors increased and at the medical school of Alexandria (300 BC) apprentices greatly outnumbered teachers (Pikoulis et al. 2008). The trend continued into the Roman era to such a degree that the number of apprentices accepted by physicians was observed by some to impact on physician performance (Ker 1947). In Rome, men, women and even slaves could qualify themselves medically as doctors, as all that was required was the permission of the local magistrate (Cilliers & Retlef 2006). A lack of standardized training schools, examinations or licensure requirements resulted in medical training and qualifications that were highly variable. This was compounded by an increasing number of fringe “healers” that offered alternative medico-spiritual therapies, and attempts to provide some medical teaching to all students (Cilliers & Retlef 2006). The quality of medical care, underlying ethics and accountability often came under public scrutiny (Calman 2007, p36). Indeed Cicero (106-43BC) considered doctors to be tradesmen, rather than gentlemen, and of relatively low standing (Cilliers & Retlef 2006).

Galen (AD 129-199) was a major influence on medical training and medical thought in Rome. After a thorough grounding in literature, mathematics and philosophy, Galen spent 12 years in apprenticeship and study in Pergamum,
Smyrna, Corinth and Alexandria before embarking on a career in medicine that would greatly influence medical practice. In a life devoted to advancement of medical knowledge and self-improvement, Galen integrated science, clinical practice, ethics and philosophy (Leiser 1983).

In Galen we see evidence of a combination of cognitive and non-cognitive attributes that remain desirable in candidates for medical school today. Galen considered that not all candidates were suitable for medical training and the training of one unsuited to the profession would not yield the desired results (Levey 1967, pp18-94). Indeed, he explicitly advised physicians against teaching the art of medicine to unworthy students (Levey 1967). Galen wrote extensively about characteristics required by medical students. According to Galen, medical students needed both innate ability and the willingness and motivation to commit to ongoing learning and he was concerned that students of medicine did not possess either (Singer 2002, p33):

*In my experience, other accomplishments follow if one is well endowed with will and ability; if either of these is lacking, it is quite impossible for a goal to be reached....... So are today’s doctors deficient on both counts? Do they lack both potential and sufficient eagerness in their preparation for the art? Or do they have one but lack the other?*

Galen was also concerned that many students did not have sufficient general education or academic ability for the study of medicine and he attributed this to a lowering of societal morals (Singer 2002, p31):
That no one should be born with sufficient mental powers to learn an art which is so beneficial to mankind seems absurd, since the world is essentially the same as it was in previous times; the seasons have not changed order, nor has the sun’s course altered, nor has any one of the stars- either a fixed star or a planet- admitted of change. It must be because of the bad upbringing of our times, and because of the higher value accorded to wealth as opposed to virtue, that we no longer get anyone of the quality of Pheidias among our sculptors, of Apelles among our painters, or of Hippocrates among our doctors.

Galen felt it was important that candidates for medical training were virtuous, putting public benefit before personal wealth, and had the mind of a philosopher (Magner 2007, p123).

During the time of the Roman Empire, the formal teaching of medicine became more systematized and standardized (Fulton 1953). Great teachers attracted increasing numbers of students, and guilds and colleges grew up around them (Calman 2007, p36-7). Universities and research centres formed and foci of medical training developed (Puschmann 1966, p97). The College of Physicians was formed by eminent physicians who were also commonly teachers and became responsible for electing new doctors as the need arose (Cilliers & Retlef 2006). As the medical profession grew, privileges were awarded to the more respected physicians (Leiser 1983). This represented an attempt to remove financial pressures from these great physicians, allowing them to focus on teaching and the selection of students with qualities they considered essential. In theory, this would provide a fertile ground for the development of good doctors. This approach may not have
been entirely successful, however, as reports from 6th century Rome, describe doctors as ineffective and potentially harmful, low class, untrustworthy persons in a disreputable profession (MacKinney 1955). This sentiment is clearly apparent in the following communication between Apollinaris Sidonius and his brother (AD 461-7) (Dalton 1915, XII):

So under Christ's guidance we are determined to fly the languor and heat of town with all our household, and incidentally escape the doctors also, who disagree across the bed, and by their ignorance and endless visits conscientiously kill off their patients.

Concern had also long been expressed from within the medical profession. Galen for example, compared doctors to robbers, differing only in their location of practice (Kuhn 1821-33). This fall into disrepute resulted in a shift from classical medical training to the incorporation of medical teaching into the education of the more highly respected clergymen (MacKinney 1955).

1.3.4 The Islamic Golden Age

In Islam, from the 7th to the 10th century, medical training was most commonly achieved through apprenticeship between father and son (or occasionally daughter) (Leiser 1983). Students could also educate themselves through medical texts or could train through participation in classes in hospitals or medical schools (Leiser 1983). Although medical education was readily accessible to most, the route was long and tedious. Motivation and the ability to comprehend medical texts were key factors required by self-learners. At the time, medical opinion appears to have been
divided over the best way to learn medicine, it is of note however that the most influential and respected physician of the era, Avicenna, was largely self-educated.

Acceptance into medical schools required a degree of competence in Islamic studies, philosophy, astronomy, chemistry and the arts (Al-Ghazal 2004). In some cases teachers excluded students on the basis of factors such as religion or race (Leiser 1983). The flourishing of medical ethics in the 9th and 10th centuries brought attention to the importance of choosing suitable students for medical studies. Al-Ruhawi a notable physician in 9th century Islam, felt that the choosing of an inappropriate student for medical training had caused the moral decline of the medical profession since the times of Hippocrates (Levey 1967). Indeed, Al-Ruhawí devoted a chapter of his seminal work, Practical Ethics of the Physician, to a description of the qualities required of students of medicine. The chapter, titled On the Medical Art That is Not Fit for Everyone Seeking It but for Those Who Are Suitable for It in Nature and in Morals, outlines four key principles for excluding candidates for medical training (Levey 1967, p78 (Adab al-tabib MS, fol. 88b-89a)):

One of them concerns the complexion of the body when it becomes immoderate. By this, I mean the change of one's morals and acts of the soul. The second is that one must grow accustomed to bad habits in the presence of other bad ones, and to be familiar with ignorant ones, to follow what they do when they become good. The third is when these two come together. This is much worse in corruption and much diminished in goodness. One may add to these that if the student learns the art of medicine, not for its nobility or its benefits to himself
and his body and to the bodies of children of his species, but for worldly purposes as to be wealthy and achieve power, etc., it is better that he not attain most of his desires nor enjoy the favor of what he had hoped.

He emphasized that those selected into medical training should be fit in body and soul and should be motivated by virtue, rather than power and wealth. Commitment and diligence to study were critical. Al-Ruhawi believed that the sons of physicians were ideal candidates as they had the advantage of constant exposure but he acknowledged that not all sons of physicians were suited to the study of medicine (Levey 1967, p79 (Adab al-tabib MS, fol. 90a)): 

As to the soul which is not fit to receive the medical art [i.e., a worthless son of a physician], the belief which decrees it is only a permissive one. In the opposite case, it is understood that he is fit to receive it. However when he is from those who cannot preserve what is in the oath, then he should not study the medical art.

Al-Ruhawi also recognized the important role of the medical profession in embracing and nurturing such desirable qualities. As such, a medical student with suitable qualities was like ‘a good wine that was only fit to be preserved in a vessel that would preserve the taste and purity of its colour, the goodness of its odour, and, in short, its other good qualities.’ (Levey 1967, p78 (Adab al-tabib MS, fol. 88b-89a))

There was an increasing focus on the total intellectual, moral and physical character of doctors and detailed descriptions of the desirable qualities of doctors were
provided by renowned physicians of the day. *Practical Ethics of the Physician* includes a detailed description of the appropriate moral and personal attributes expected of doctors. Physicians must be sensible, learned, pious, and must act without haste. A virtuous and ethical approach, integrity, reason, respect, patience and self-restraint, were essential as were the ability to communicate clearly and a commitment to ongoing education (Levey 1967). Careful attention to a healthy lifestyle and the maintenance of personal hygiene were also advocated.

Comments on desirable qualities possessed by candidates for medical training were not confined to the cognitive and non-cognitive domains. In his *Book of What is Best in Medicine*, Ibn Habal described the physical characteristics of a doctor (Ibn Hubal 1943-4, fi 1-tibb):

>[he] should be of attractive appearance and shape. He should be in good spirits, neither frowning nor scowling. People should look forward to seeing him, smile when speaking with him and when flocking to him. He should be of moderate stature, neither fat nor lean. If not, leanness is better. He should have a light rosy complexion if he is from a country where such a complexion is found. His forehead should be wide in proportion to his face and there should be a wide area between his eyebrows. His eyes should be bluish black and appear as if they are laughing or always looking at something pleasant. They should be of moderate size, neither protruding nor indented. If not, indentedness is better. A person with such eyes should seem mild and not overbearing. His tears should not be visible. Such a person should have facile cheeks, a small jaw, and sparse beard of moderate length which is neither
curved nor inclined to the shoulders. He should have long upper arms, large hands with delicate edges and little flesh on the palms, and long fingers. He should have smooth buttocks, moderately fleshy thighs, straight legs, and discernable arches in his feet. His feet should point exactly in the direction that he walks and he should have a light deliberate walk.

In the absence of a standardized medical program, there was an increasing need to regulate the medical profession and to verify the competence of doctors through examinations (Leiser 1983). Authors such as Al-Ruhawi advocated standardized examinations to assess competence prior to practising medicine (Levey 1967). Regulatory screening of medical practitioners commenced in the 9th or 10th century and only those with satisfactory qualifications were granted a licence to practice (Al-Ghazal 2004; Levey 1967). As the first millennium approached, physicians came to be recognized as educated persons with knowledge, wisdom and high moral standards and the status of the profession rose (Al-Ghazal 2004).

1.3.5 The Middle Ages

In Europe, medicine had become intricately associated with the church and the monastery after the fall of the Roman Empire and clerics, as the only educated class in the early Middle Ages, were responsible for creating healers of clerical candidates (Bloch 1988; Lagarde 1915, p66; Magner 2007, pp135-137). Monasteries were founded in the 6th century, cloister schools were established, and monks prepared for the practice and teaching of medicine and surgery. Magico-religious medicine dominated as science was considered hostile to theology and medical training was passed down to students in conjunction with religious lore
(Bloch 1988; Reisman 1935; Roth 1953). The entry of students into theological training initially followed an apprenticeship model and monastic priests were advised to take a child under his care and to teach him (Lagarde 1915, p519). Students of intellectual promise went on to a monastic or cathedral school where medical teaching was largely general and theoretical (MacKinney 1955). Education became more accessible as Grammar schools developed in association with larger parish and collegiate churches, however ‘promising’ youths and sons of wealth families often had access to direct one-on-one teaching from abbots, bishops or local clergy (Rowdon 1971).

Medical study often began early in life and a host of desirable personal, behavioural and physical qualities were considered essential prerequisites for entry into medical training (Galvao-Sobrinho 1996; MacKinney 1955). The Letter to Arsenius, composed during the latter part of the first millennium outlines these comprehensive phenotypic requirements (MacKinney 1952, p8):

First, [he] should test his personality to see that he is of a gracious and innately good character, apt and inclined to learn, sober and modest,... charming, conscientious, intelligent, vigilant and affable, in all details adept and skilful... amiable, humble and benevolent... not timid, turbulent or proud, scornful or lascivious, or garrulous, a publican, or a woman-lover ... not drunken or lewd, fraudulent, vulgar, criminal or disgraceful ... [He] should not have faults, but instead discretion, taciturnity, patience, tranquility, and refinement.
Following a change in church policy in the early 13th century, the responsibility for medical education and practice shifted from the clergy to academic institutions (Bloch 1988). Medicine became institutionalized within the university system of medieval Europe during the 12th and 13th centuries (Magee 2004). The development of universities in Europe was sporadic and students often travelled great distances to train under renowned physicians (Magner 2007, p141). Notable medical schools of this period were founded in Bologna, Padua, Salerno, Naples and Montpelier. The medical school of Salerno (founded in the 10th century) is of particular note as its development heralded the beginning of the modern history of medical education. In relation to the medical school at Salerno, Walsh describes a liberal admission scheme with clearly articulated academic prerequisites, supported by a robust regulatory process (Walsh 1911, p15):

*Students came from all over, from Africa and Asia, as well as Europe, and when abuses of medical practice began to creep in, a series of laws were made creating a standard of medical education and regulating the practice of medicine….. Finally a law was passed requiring three years of preliminary work in logic and philosophy before medicine might be taken up, and then four years at medicine, with a subsequent year of practice with a physician before a license to practise for one's self was issued.*

Similar selection practices were adopted by other medical schools, however many schools also considered the gender, race and religion of candidates (Roth 1953). Women and Jews in particular were frequently unable to gain entry into formal medical training, particularly outside of Italy (Magner 2007, p149; Roth 1953).
Entry into medical training was available to laypersons and innate personal qualities were considered less important than a commitment to abide by ethical norms (Bloch 1988; Galvao-Sobrinho 1996). Students characteristically entered medical training at universities after mastering basic grammar, rhetoric, logic arithmetic, geometry, astronomy and music (Walsh 1911, p79). Age does not appear to have been a major admission criterion and it was common for students to commence training at university at 12 or 13 years of age (Walsh 1907). Some medical schools did however require graduates to be of sufficient age prior to commencing medical practice. At Salerno, for example, candidates were required to be at least 21 years of age and have studied medicine for at least 7 years in order to sit licensing examinations (Lagarde 1915, p35).

Formalisation of medical training within the university system and standardisation of admission requirements was associated with a standardisation of curriculum and training. However, as only a small percentage of all medical practitioners of this period had formal university training the standard of medical practice remained highly variable (Magner 2007, p144). Walsh (Walsh 1911, 87) reports that;

> Almost anyone who wished could set up as a physician, and those who were least fitted were often best able to secure a large number of patients by their cleverness, their knowledge of men, and their smooth tongues.

The response was a tightening of licensing and regulatory practices in southern Europe from the 13th century. Bishops issued decrees forbidding all from medical practice except graduates of Montpellier (Walsh 1911). Furthermore, from 1272 all
those practising medicine at Montpellier were required to be examined and licensed (Roth 1953). Similar laws were passed in the Pyrenees in 1289 and by the end of the Middle Ages licensing for medical practitioners had become firmly established (Roth 1953). By this time licensing examinations commonly covered not only medical knowledge but also involved consideration of the candidate’s premedical training and academic achievements (Roth 1953). This may have represented one way of dealing with a lack of regulation regarding the required academic criteria for candidate selection into medical training by apprenticeship.

Candidates who were unable to enter formal medical training obtained their medical knowledge and skills through books and apprenticeship. Selection of an apprentice is likely to have depended upon the judgement of each practitioner. The potential personal benefits gained by the practitioner appear to have been particularly motivating factors for accepting an apprentice. This is evident in, Roth’s report of a contract, drawn up between a practitioner and her apprentice in 1326, that required the apprentice to give all fees received for his medical services during the apprenticeship to his teacher (Roth 1953).

1.3.6 The Renaissance

The Renaissance was associated with the proliferation and development of medical schools, medical guilds, faculties and colleges in Europe (Calman 2007, p90). Race, religion, economic status and knowledge of Latin were key factors influencing selection of students into university medical schools (Shatzky 1950). There was increasing recognition of candidates past academic achievements and, in Paris, candidates for medical training were expected to have followed lectures in
the faculty for several years prior to entry into the bachelor of medicine course (Calman 2007, p113). In Vienna, candidates’ character and social position were considered, however this occurred prior to graduation rather than at selection into the medical course (Calman 2007, p114). An early manual for graduating medical students provided further insight into the qualities expected of a medical student, stipulating that s/he should be (Durling 1970, p23):

of good character and good memory, well formed, well behaved, daring in diseases where nothing is to be feared, circumspect in dangerous cases, let him flee severe diseases, be gracious to the sick, peaceable with his colleagues, cautious in prognosis, chaste, sober, pious, compassionate, not grasping or extortionate.

Throughout Europe, the ability to pay course costs was a major factor affecting entry into medical training and the less fortunate candidates were forced to obtain funding from wealthy sponsors in return for their services after graduation (Shatzky 1950). In some cases, later year entry into medical courses was possible provided successful completion of studies in natural philosophy and medical theory at respectable institutions were demonstrated (Pelling & Webster 1979).

1.3.7 The United Kingdom

The first medical course in Britain opened at the University of Aberdeen in 1495 (Carter 1994). However, most practicing physicians in Britain in the 14th to 17th centuries studied medicine in Europe and then applied to have their foreign degrees incorporated at Oxford, Cambridge or both (Allen 1946, p116). The development
of medical education in Britain lagged behind that occurring in Europe and few medical degrees were given at either Oxford or Cambridge before the 17th century (Allen 1946, p116).

Entry into medical training at Oxford at this time required completion of bachelors and masters degrees in arts (Magee 2004). At Cambridge, an arts degree was not mandatory for admission, but was held by most enrolled medical students (Allen 1946, p122). A Decree of the Heads of Colleges in 1684 made it obligatory for candidates for the degree of Bachelor of Medicine to study for 6 years and to reside in a university college for ‘the greater part of nine several Terms at least’ (Allen 1946, p122). Medical students of this era appear to have been far more adventurous socially than intellectually, and to be motivated by personal comfort rather than altruism.

Licensure by royal patronage was also possible and medical degrees were uncommonly granted by the Royal family as a sign of royal favour. In 1531, an act of Parliament made it possible for the Archbishop of Canterbury to give degrees in medicine (Allen 1946, p131). Although the Church of England was permitted to grant medical licenses, practitioners licensed in this way were not held in high regard and it was common for those granted a degree from the church to then formally study medicine at university (Allen 1946, p131). The licensing power of the Church was soon restricted by the privileges of the Royal College of Physicians.
The Royal College of Physicians (1518) and the Royal College of Barber Surgeons (1540) licensed practitioners in London from the 16th century (Calman 2007, p110). Licensure candidates were required to pass an examination in Latin and the classics, from which only medical graduates from Oxford or Cambridge were exempt (Magee 2004).

Despite the gradual institutionalisation of medical education across Europe and Britain, and the accompanying rise in status of university educated doctors, non-university educated practitioners far outnumbered those with recognized degrees (Allen 1946). In London, for example, most licensed physicians held university degrees, yet most medical care in London was provided by unlicensed barber surgeons, apothecaries or midwives (Calman 2007, p120). These practitioners were generally of limited education, found their way into medical practice through self-selection or apprenticeship to an established healer (Pelling & Webster 1979). Apprenticeships were secured by way of a formal legal agreement that bound apprentice to teacher for five to seven years. During this period the expectation was that the apprentice ‘his master well and faithfully shall serve; his secrets shall keep; taverns he shall not haunt; at dice, cards, tables, bowls or any other unlawful game he shall not play’ (Warren 1951, p305). Apprentices were not able to marry or commit fornication while apprenticed to their teacher (Lane 2002, p59). They also had to possess sufficient funds to pay for the teacher’s services, needed to be fluent in Latin, and to display self-motivation for learning (Warren 1951). The apprenticeship system persisted because it provided sufficient control of entry into medical practice and competent practitioners despite its expense, relative inflexibility and lack of a governing system (Lane 2002).
It was not until the rise of medical education in Scotland towards the end of the 17th century that formal medical training began to be seen as a realistic option to apprenticeship. Three medical chairs were created at Edinburgh University in 1685 and by 1750 medical faculties had opened in Edinburgh, Glasgow, Aberdeen and St Andrews (Kett 1964). Although, university based medical training became increasingly common in Scotland during the 18th and 19th centuries, apprenticeship outside the university system remained the predominant mode of medical training in most parts of the United Kingdom until the 19th century (Hull & Geyer-Kordesch 2003). Attempts were made to provide means for apprenticeship-trained practitioners to gain university qualifications prior to registration as a medical practitioner (Magee 2004). This resulted in an increase in the number of “qualified” physicians, however, in the absence of rigorous licensing practices, this led to a situation where qualifications did not reflect competency or quality of medical care (Kett 1964).

In England, it was not until the 19th century that efforts were made to increase the number of medical schools and systematise medical education within the university system. The number of medical courses increased significantly as provincial medical schools were established across the country. Standards for selection in these medical courses were not high and the provincial schools had to advertise locally and even award prizes in order to attract applicants (Reinarzp 2006). Consideration does not seem to have been given to personal and professional qualities as students are commonly described as undisciplined (Ormerod 1945, p38).
Apprenticeship training fell out of favour following the passing of the 1858 Medical Act (Hull & Geyer-Kordesch 2003). Steps were taken to formalise and standardise medical training across the United Kingdom. From 1857, in Scotland, potential medical students were required to pass a preliminary examination that assessed English composition, Latin, mathematics and natural philosophy. This procedure was adopted in England following a review of medical education and examinations by the General Medical Council (General Medical Council 1881). In the report, the General Medical Council called for a formalisation of the process of medical student enrolment and training. Entry into medical training was based on academic merit and, from the 1st of January 1882, all candidates were required to pass a preliminary examination in English, history and modern geography prior to registration as a medical student (Anonymous 1881b). Although a minimum age for entry into training was not specified, the report recommended that medical training should be at least four years and that candidates should not be awarded a license to practice before the age of 21 (Anonymous 1881a).

Thus, the formalisation of medical training within tertiary institutions was associated with specific and standardised admission requirements. It was also associated with a dramatic increase in costs. Both factors are likely to have contributed to a remarkable transformation in the nature of medical students during the 19th century. At the beginning of the 19th century, most British medical students were apprenticed outside the university system and were known for their unruly behaviour and their lack of focus on study (Calman 2007, p173). The motivations for medical training were largely financial. However, by the end of the century, the majority of medical students obtained their training within a university system,
were studious and respectful, and were motivated by their teachers (Calman 2007, p174).

Towards the middle of the 20th century, concerns arose about the efficacy of current selection tools (Himmelweit & Summerfield 1950). These concerns were related to the inability of standard selection methods to discriminate borderline cases and to permit unsuitable candidates to enter training (Himmelweit & Summerfield 1950). It was also felt that the strong focus on academic achievement created an overly competitive environment that placed strain on candidates (Himmelweit & Summerfield 1950). In 1944 the Goodenough committee advised that only universities, and institutions that conform to university standards, were appropriate to conduct medical training programs (Inter-departmental Committee on Medical Schools 1944). The committee recognised the importance of a robust selection process for medical school entry and recommended that:

- candidates must have received a good general education
- selection should not be based wholly on examination results
- all candidates should be interviewed, preferably by a small committee
- great attention should be paid to reports obtained from head masters and head mistresses
- selection methods should be supplemented by means of identifying those found unsuitable in the early years of the course
- there is room for the development of new methods of selection
In an attempt to address discrimination on the basis of gender, the report decreed that unsuitability for a medical career should be the only barrier to admission to a medical school (Anonymous 1944).

The first three of the committee’s recommendations were adopted rapidly by British medical schools and, by the middle of the 20th century, medical schools selection processes in Britain commonly considered final high school examination results, personal interview and reports from secondary school headmasters/ headmistresses (Smyth 1946a). Interest in selection procedures for medical schools increased steadily due to an increase in numbers of candidates and the successful implementation of selection procedures by the War Office Selection Branch (Brinton 1954). A need to invest in research in student selection was recognised by the Medical Research Council which provided funding for several large scale cohort studies of students (Himmelweit & Summerfield 1950).

The number of applicants for medical schools continued to rise after the Second World War (Brinton 1954). By the middle of the 20th century, selection procedures in Britain commonly involved three stages: consideration of the application and confidential reports, personal interview and observations and examination results recorded during a pre-medical university year (Brinton 1954; Whitby 1956). Consideration of the application and associated reports involved a study of the candidate’s previous academic record, and of statements and testimonials from referees (Whitby 1956). The academic record provided valuable information relating to academic ability and achievement (Brinton 1954). Referee reports addressed six questions designed to enable an assessment of character and moral
qualities such as leadership, independence and creative power, popularity, communication skills, capacity for inspiring confidence and ability to cooperate with others (Whitby 1956). The aim of the interview was to assess the personality and motivations of each candidate (Brinton 1954; Whitby 1956). Interviews were occasionally conducted by the dean of the medical school, but were more commonly administered by admissions committees comprised of a panel of interviewers (Whitby 1956). Brinton (1954) described wide variation in composition of interview panels between universities. The length and form of the interview was also highly variable between medical schools (Himmelweit & Summerfield 1950; Whitby 1956). In some schools interviews were to all candidates, whereas in others they were reserved for only borderline candidates (Himmelweit & Summerfield 1950). The experience of one candidate suggests that, at least in some cases, interviews screened for social status rather than personal qualities (Anonymous 1962). Observation during the first year of medical school was a selection tool employed by only some institutions. It was perceived by some as a method of selecting out those who were not coping with the course and was considered a means of ensuring that enrolled students are worthy of maintaining a place in the course (Whitby 1956).

Some medical schools also required candidates to sit written entrance examinations that tested candidates’ breadth or depth of knowledge (Himmelweit & Summerfield 1950). However, such entrance examinations were not routinely employed as they were taxing on university resources and were considered to provide little extra information above that available from other sources (Brinton 1954). Standardised entry examinations were also developed to assess the aptitude of potential medical
students and included questions designed to test ingenuity, reasoning, scientific knowledge, arithmetic and powers of description (Smyth 1946b). An entrance test described in detail by Wilkie (Wilkie 1946), comprised 10 short answer questions divided into three parts designed to assess scientific aptitude and intelligence, general knowledge and instruction, and general intelligence respectively. The usefulness of such studies was a source of debate, some advocated for an expansion of their use, some questioned the validity and relevance of such tests, and many were strongly opposed to their use in student selection (Himmelweit & Summerfield 1951). Such aptitude tests were adopted by only a minority of British medical schools (Himmelweit & Summerfield 1950; Whitby 1956). As a result, psychological testing and intelligence testing did not become widespread tools in selection in the United Kingdom. Rather it was suggested that aptitude tests should be reserved for assisting with allocation within a profession rather than selecting for the profession (Whitby 1956).

Over the next twenty years, the number of candidates applying for entry into medical schools continued to increase and, by the end of the 1970’s the number of candidates applying for entry into medical schools outnumbered available places by at least 3 to 1 (Anderson, Hughes & Wakeford 1980). A centralised national admissions process was introduced to improve the efficiency of student selection. Past academic achievement and personal interview remained the cornerstone of selection practices in the United Kingdom during the 1960s and 1970s. To be considered for entry to any medical school at the end of the 1970’s candidates required three A levels that must include chemistry or physical science, and O level passes in physics, mathematics, and a biological subject (Thurman 1979). However,
the selection process was described as “secretive and the objectives, policies, and selection practices of different medical schools differed markedly (McManus & Lockwood 1979). Some schools only considered university applicants who had applied solely for admission into medical courses. Some schools selected primarily on academic achievement, which served to drive up the academic requirements of those accepted into the medical course. This was particularly true of medical schools in Scotland and Wales. Other schools supplemented academic criteria with other processes such as interviews. In some cases, age, life experience and locality were given consideration (McManus & Lockwood 1979).

Interviews were an established part of the selection process in 22 of the 31 medical schools present in the United Kingdom in 1979 (Fulton 1979). The importance of the interview did, however, vary between schools. Some afforded little weight to interviews whilst others relied upon interview to select from a pool of candidates that fulfilled minimum academic requirements. Interview formats varied greatly between schools and interviews were criticized as being unfair and misdirected (Toynbee 1978). Indeed, the entire selection process in the United Kingdom in the late 1970’s was considered by some to be inappropriate and ineffective (Jacobson 1978).

Anderson and colleagues (1980) proposed a comprehensive code of practice for medical student selection that required medical schools to clearly describe their selection procedures, routes of entry and to state any quotas for types of entrants. Medical schools were expected to provide explicit details regarding the way academic and non-academic attributes were assessed and to list any academic
standards and personal attributes that may be required by candidates, or considered particularly desirable in candidates during the selection process. They were also expected to outline the decision making processes involved in selection and to provide information regarding the influence of factors such as candidates’ order of preferences on chance of selection. Ideally, this would enable candidates to make an informed choice when applying to medical school.

At St Mary’s in 1980, the selection process involved assessment of cognitive and non-cognitive attributes using a sequence of selection tools (McManus & Richards 1984). In order to deal with an oversupply of candidates, those that fulfilled minimum academic admission criteria were short-listed using a standardised admissions form. Short-listed candidates were required to complete the state-trait anxiety inventory, the Eysenck personality questionnaire, and a survey that asked about personal interests, and ethical, political, and social attitudes (McManus & Richards 1984). McManus and Richards (McManus & Richards 1984) examined the relationship of twenty four demographic, educational, and applicational variables to chance of selection. A level grades were the most important factor determining selection of candidates. O levels and a medical parent were also found to increase chance of selection. However, chance of selection was not influenced by schooling, gender, social class, career preferences or performance on the state-trait anxiety inventory and Eysenck personality questionnaire.

Attention was drawn to the lack of research in the UK on selection for entry to medical school and the lack of knowledge about the success (or otherwise) of selection procedures (Anderson, Hughes & Wakeford 1980). As late as 1992,
selection procedures in the UK were described as amateurish and haphazard (Lowry 1992). Students could enter medical school at 18 years of age with 'A' levels obtained at secondary schools (Parsell & Bligh 1995). Support grew for the mandatory inclusion of selection tools that assessed non-cognitive qualities (Crisp 1984; Linke, Chalmers & Ashton 1981; Parkhouse 1979). Some authors suggested that psychometric testing may be more useful selection tool than academic performance (Lowry 1992). Interviews continued to be employed by many schools, however selection processes varied widely between medical schools and little effort was made to ensure the quality and consistency of interviews (Lowry 1992). In 1993, the report of the General Medical Council (GMC) Education Committee “Tomorrow's doctors” prompted a review of assessment and selection methods (Fowell & Bligh 1998).

“Tomorrows doctors” brought about a renewed interest in methods of selection for medical school and an increasing recognition of the importance of communication, teamwork and other non-cognitive attributes (Calman 2007). A review of medical student selection processes by Parry and colleagues in 2006 (Parry et al. 2006) found that medical schools commonly selected students based on academic ability coupled with a “well rounded” personality demonstrated by motivation for medicine, extracurricular interests, and experience of team working and leadership skills. A minimum standard of academic performance was essential. All schools required A levels that included biology and chemistry and grades of at least ABB. Most required grades of AAA or AAB. Two UK medical schools also required candidates to take the biomedical admissions test (BMAT), a composite test of
three sections; Aptitude and Skills, Scientific Knowledge and a Writing task designed to assess written communication skills (Emery & Bell 2009).

Processes used to assess non-cognitive qualities varied substantially. Some schools used information presented in the candidate’s personal statement and referee’s report while others ignored this because of concerns over bias. A few schools sought additional information from supplementary questionnaires filled in by the candidates. One university utilized a personal qualities assessment tool to assess a range of personality attitudinal traits considered important for the study and practice of medicine.

Interviews were routinely employed by most schools to supplement selection. Most schools interviewed only candidates short-listed according to academic performance or a wider range of non-academic criteria. When employed, interviews were observed to vary in terms of length, panel composition, structure, content, and scoring methods. Most medical schools conducted 15-20 minute interviews assessed by panels comprising academic staff, community members and senior medical students. At some schools interview questions were predefined by the school, whilst at others they were interviewer led. In all cases, questions were directed towards predefined criteria of particular interest.

Parry (Parry et al. 2006) advocated for a more consistent national approach to selection and suggested that a number of assessment processes could be established nationally. Efforts to improve the measurability, reliability and consistency of tools used to measure non-academic factors culminated in the development of the UK
Clinical Aptitude Test (UKCAT) in 2006 (UKCAT Board 2008). The test consisted of four sections (verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis) and was designed to assess candidates’ aptitude for medicine, judgement and decision-making ability. The scores on these four subtests contribute equally to the total UKCAT score. The test awards overall and section-specific scores to candidates that are then available to medical schools. A fifth section, non-cognitive analysis, was trialled in 2007 but not recommended for use for selection purposes until its psychometric properties have been evaluated (Lynch et al. 2009).

Most UK medical schools currently select students on the basis of academic achievement, aptitude for medical study and desirable personal qualities. High school grades (or GPA for graduate entry courses) continue to be the gold standard for measuring the academic achievement of candidates. Minimum academic achievements remain high and continue to be driven higher by competition between candidates. For example, to be considered for entry into the medical school at Edinburgh University, British high school graduates require 3 A level A grades to be eligible for selection, however most candidates selected into the course have obtained at least 6 A grades (University of Edinburgh 2010a).

Aptitude tests are used to facilitate selection at most UK medical schools (Cassidy 2008). The UKCAT has been adopted by 26 of the 32 UK medical schools and now represents an important component of their selection processes. Four medical schools; Oxford, Cambridge, Royal Free and University College and Imperial College use the BMAT, whilst candidates applying for entry into graduate medical programs at Keel, Nottingham, Swansea, Limerick and St George’s University of
London medical schools are required to sit the GAMSAT (Cassidy 2008). However, the way aptitude test scores are integrated into selection processes varies between schools.

Interviews, letters of reference and/or personal statements are used to select for desirable personal qualities such as empathy, ethical reasoning, communication and interpersonal skills, motivation for a career in medicine, social and cultural awareness, leadership, responsibility, teamwork and organisational skills. Interviews are conducted at almost all UK medical schools, however, letters of reference and personal statements are also often considered and, in some schools, such as the University of Edinburgh medical school, replace interviews (University of Edinburgh 2010b).

Provided minimum academic requirements are met, academic and non-academic criteria both commonly contribute to final rankings for admission into the course. However, concerns still remain regarding the lack of uniformity in the approaches of medical schools to the choice, application and weighting of selection methods (Cassidy 2008).

A further development in selection for medical school in the UK has been the introduction of mandatory criminal and health checks with a requirement for vaccinations and a satisfactory immune status prior to enrolment into medical school. Medical school entry is not available to individuals who are positive for the hepatitis B surface antigen, regardless of their e antigen status, but does not
discriminate against those with human immunodeficiency virus (HIV) and hepatitis C virus (HCV) (Tambyah 2005)

1.3.8 The United States of America

In 18th century America, an apprenticeship with a practicing physician was the usual way of preparing for entry into the medical profession (Miller & Weiss 2008). During the 18th century, medical schools were established within universities and by entrepreneurs, and medical education in the United States continued to be administered primarily through apprenticeship in association with these (Beck 2004). However, this did not necessarily improve the quality of medical practice since medical school graduates were not necessarily more competent than those trained by apprenticeship outside the university system (Humphrey 1975).

Nevertheless, the presumed benefits of a structured program supplemented by formal coursework was recognised by some state licensing boards which exempted medical school graduates from licensing exams (Kett 1968). Lax licensing laws, a surplus of students and monetary incentives led to a flourishing of medical schools (Calman 2007). Indeed, between 1810 and 1910, some 457 proprietary schools of medicine appeared in the United States, many placing only a monetary price on admission of students to medical training (Severinghaus 1954). Indeed, medical schools routinely ignored candidate’s qualifications in an effort to boost enrolments (Miller & Weiss 2008). Medical students of this period were described as the “one son in the family thought too weak to labour on the farm, too indolent to do any bodily exercise, too stupid for the bar, and too immoral for the pulpit” (Severinghaus 1954). These characteristics, in combination with wide discrepancies
in quality of teaching, resulted in physicians that varied tremendously in their medical knowledge, therapeutic philosophies, and aptitudes for attending the sick (Hudson 1972).

Concerns that many medical schools represented “diploma mills” were raised by the American Medical Association and echoed in the press (Calman 2007). The deterioration of the medical profession caused great concern. In response, some medical schools made efforts to raise admission and graduation standards (Miller & Weiss 2008). In 1893, Johns Hopkins University medical school set an unprecedented admission standard of accepting only those having completed a four-year undergraduate degree (Miller & Weiss 2008).

Despite attempts by state licensing boards and the American Association of Medical Colleges to raise standards on a broad scale, less reputable medical schools persisted (Miller & Weiss 2008). This prompted a nationwide survey to assess and appraise all medical schools in the United States in several key areas: entrance requirements, qualities of the medical training program and faculty staff, and availability of laboratory and clinical facilities. The survey began in 1908 and was conducted by Abraham Flexner over a period of 18 months. He concluded that standards of medical practice in the United States were more variable than any other country in the world and advocated for the closure of 80% of American medical schools (Flexner 1910). The Flexner Report was monumental in the attention that it drew to medical education and the state of the American medical profession at the beginning of the 20th century. The report provided strong impetus for the implementation of increasingly strict admission standards and curriculum
requirements by state licensing boards, and led to the closure of almost a quarter of American medical schools (Hiatt & Stockton 2003). The formation of a Federation of State Medical Boards in 1912 contributed greatly to the standardization of medical training and medical practice across America (Beck 2004).

By 1915, ninety six medical schools remained active, of which eighty five required one or two years of college prior to admission (Rothstein 1987). By the end of the 1930’s, two years of college were a mandatory requisite for all schools (Rothstein 1987). The closure of many medical schools also served to greatly reduce the number of available training positions for medical students and admission standards were driven up further by the educational level of competing candidates (Rothstein 1987). Although not a requirement of most medical schools, many students accepted into courses had completed bachelors degrees (Burke 1982).

As the middle of the 20th century approached, previous academic achievement remained a key component of the selection process in American medical schools, Secondary school records and college transcripts were assessed as markers academic achievement and represented the first hurdle into medical training (Bloomgarden 1957).

However, other methods were developed to extend selection beyond the academic domain. Some medical schools developed checklists of qualities or characteristics against which candidates were assessed (Bloomgarden 1957). Data sourced from the college residences of candidates provided information regarding a candidate’s “general qualifications” for the study of medicine (Severinghaus 1954). Candidates
were required to submit information relating to extracurricular activities, their reasons for applying for entry into the medical course along with letters from character references (Bloomgarden 1957). Reference letters usually required referees to comment on personality, character, appearance and intellectual capacity (Kay 1944). In some cases, referees were clearly directed towards qualities considered desirable, such as honesty, integrity, judgement, reliability and industry (Kay 1944).

Personal interviews were employed by most medical schools and were valued for their ability to provide data about candidates not available from other sources (Severinghaus 1954). Such data related to personality, communication skills, grooming and physical appearance (Bloomgarden 1957). Given the resources required for interviews, these were often restricted to subsets of candidates according to criteria determined by selection committees (Bloomgarden 1957; Kay 1944). Interviews commonly targeted factors such as reasons for wanting to study medicine, financial situation, appearance, personality and attitudes (Kay 1944). Interviews were largely unstructured and non-standardised and the properties of interviews varied widely between medical schools. Interview duration ranged from under 15 minutes to 90 minutes (Kay 1944). Interviewer panels varied in number from one to seven and in some schools, all candidates were interviewed by the same interviewer or panel of interviewers, whilst in others interviewers varied between candidates (Kay 1944).

Aptitude testing appeared in its infancy in 1929 (Smyth 1946b) and became an increasingly complex tool for assessing cognitive and non-cognitive characteristics
not readily observable from a candidate’s academic record. Intelligence testing was employed by some medical schools. An IQ of at least 125 was considered desirable by most medical educators (Bloomgarden 1957), as it was felt that those with IQ below this would struggle to cope during medical school and with medical practice after graduation (Brosin 1948). Rorschach testing was strongly advocated by some authors as “the best method known to us for selection purposes” (Brosin 1948), but dismissed by others to be ‘unjustified in the selection of medical students’ (Eron 1954)

The Medical College Admissions Test (MCAT) was developed by the Association of American Medical Colleges and, by 1942, was administered to almost every college student intending to apply for entry into medical school (Kay 1944; Severinghaus 1954). The test was developed to assess whether candidates possessed sufficient academic aptitude, background knowledge and reasoning ability for successful performance as a medical student (Fruen 1983). The test focussed on general and scientific knowledge, analyses and problem solving and candidates were provided with overall and specific scores for four subscales of interest: verbal ability, quantitative ability, science and understanding modern society (Erdmann et al. 1971). Medical school candidates generally performed well in the MCAT, however only those ranked above the 50th percentile were likely to gain admission (Bloomgarden 1957).

In 1942, a survey was sent to all 65 medical schools in the United States requesting information relating to selection procedures (Kay 1944). Replies received from fifty
three (82%) of the schools provide a snapshot of pre-world war two selection processes in America (see Table 1.1).

**Table 1.1 Criteria used for selection into American medical schools**

<table>
<thead>
<tr>
<th>Selection tool</th>
<th>% medical schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum 3 years of college</td>
<td>100</td>
</tr>
<tr>
<td>Minimum &gt; 3 years college</td>
<td>8</td>
</tr>
<tr>
<td>Aptitude testing (MCAT)</td>
<td>98</td>
</tr>
<tr>
<td>Statement of reasons for studying medicine</td>
<td>62</td>
</tr>
<tr>
<td>Handwriting (on application form)</td>
<td>2</td>
</tr>
<tr>
<td>Writing skills (on application form)</td>
<td>8</td>
</tr>
<tr>
<td>Letters of recommendation</td>
<td>96</td>
</tr>
<tr>
<td>All candidates interviewed</td>
<td>77</td>
</tr>
<tr>
<td>Only superior candidates interviewed</td>
<td>11</td>
</tr>
<tr>
<td>Only borderline candidates interviewed</td>
<td>11</td>
</tr>
</tbody>
</table>

Common requirements were for candidates to have completed at least 3 years of college, sat the MCAT, provided letters of recommendation and attended an interview. Although the way these tools were used to determine rankings for admission varied between schools, academic record generally received the most weight followed by letters of recommendation, MCAT and interview performance (Kay 1944). Interview practices, format and duration were arbitrary and so variable that their potential ability to measure any useful qualities was questioned (Kay 1944).

By the middle of the 20th century, it was claimed that “for no professional group are school applicants more carefully screened than those who apply to enter medical school” (Severinghaus 1954). An increasing abundance of information was gathered from candidates, however the effectiveness of the methods employed was
questioned. In 1952, the New York Board of Regents heavily criticised the admission practices and methods of personality appraisal of 9 local medical schools, stating that the methods employed “neither shed light on their nature or breed confidence in the methods of their measurements” (Bloomgarden 1957). Despite preliminary data regarding the usefulness of tests such as the MCAT and IQ test in predicting performance in the early years of medical school (eg. (Brosin 1948; Chesnay et al. 1936)) the usefulness and validity of methods used remained uncertain (Bauer 1956). Indeed, only 50% of matriculants graduated, failure rates in licensing exams were high and concerns remained regarding the suitability of the intellectual or personal qualifications of candidates for the study or practice of medicine (Kay 1944; Rothstein 1987).

During the 1960’s, selection processes continued to focus on academic achievement and aptitude. However, concerns continued to be raised that these measures selected for good medical students rather than good doctors. This prompted a revision of the MCAT that included expansion of the section devoted to modern society into a general section that more broadly assessed non-science information (Erdmann et al. 1971). In the 1970’s efforts were made to consider personal, non-cognitive qualities of applicants and to select those most suited to medical practice, and those with the greatest potential to meet society’s needs for improved health care (Fruen 1983). Efforts to improve the validity of the MCAT involved further revisions of the test in 1978 and 1985. The first of these revisions involved a restructuring of test content into 6 subscales: biology, chemistry, physics, science problems, skills analysis: reading and skills analysis: quantitative (Fruen 1983).
The second revision included the addition of a 45 minute essay question (Jackson & Graham 1986).

Fruen (Fruen 1983) described 5 general selection factors considered by American medical schools in the early 1980’s:

- **Academic preparation**: most medical schools considered only those applicants with a bachelors degree that included biology, physics and chemistry. Some institutions accepted candidates with as few as 2 years of undergraduate study and limited places were available for direct entry from high school into a 6 year combined bachelor and medicine degree.

- **Academic achievement**: great emphasis continued to be placed on traditional indicators of academic performance such as GPA and MCAT scores.

- **Demographic and biographic factors**: some medical schools considered the site of residence, as well as racial, economic and community characteristics of candidates. Many schools provided mechanisms to facilitate the entry of local applicants into the medical course.

- **Personal qualities**: increasing recognition was given to qualities that were considered desirable in medical students and medical practitioners such as motivation, integrity, diligence and interpersonal skills. These qualities were assessed via letters of recommendation, personal statements, selection interviews and occasionally via personality testing.

- **Non-academic achievements**: extracurricular accomplishments were sometimes considered in terms of assessing a candidate’s wider personal qualities, and also in light of any effects that such activities may have had on previous academic achievements.
American medical schools ranked GPA the most important selection tool, followed by selection interviews, MCAT and reference letters (Puryear & Lewis 1981). Interviews were dropped from the selection process at Brown University in 1982 (Smith 1991), but remained a routine process at all other American medical schools (Association of American Medical Colleges 1982). Candidates were short-listed for interview on the basis of academic, demographic and personal data available from candidates’ application forms, academic results, MCAT results and any letters of recommendation or personal statements (Fruen 1983). Short-listed candidates were then interviewed by selection committee members, academic staff, medical students or alumni in one or more interviews (Fruen 1983).

The publication of a report by the Special Advisory Panel on Technical Standards for Medical School Admissions drew attention to the need for medical school students to have basic skills and abilities required for work as a doctor. To be eligible for admission, the panel recommended that candidates had the capacity to observe and communicate with patients and have sufficient motor skills to diagnose and treat patients. Candidates were also required to possess the intellectual, conceptual, integrative and quantitative abilities to conceptualize and problem-solve common issues, and must possess the emotional health required for full utilisation of those abilities (Association of American Medical Colleges 1979). The recommendations of the Panel were rapidly adopted by American medical schools (Association of American Medical Colleges 1979).

Over the next decade, attempts were made to align selection processes with the changing requirements of medical practice. Rapid technical and biomedical
advances made it increasingly important that medical students and doctors were able to gather and assess data, to apply the basic concepts of medicine to solve scientific and clinical problems, and to communicate effectively (Mitchell, Haynes & Koenig 1994). In response to these needs, the MCAT was updated in 1991 to provide a more contemporary assessment of candidates' aptitude for medicine. The new test assessed the candidates’ mastery of basic concepts in biology, chemistry and physics, problem solving ability, analytic thinking and writing ability (Mitchell, Haynes & Koenig 1994). The test included 4 sections: verbal reasoning, physical sciences, biological sciences and writing sample. The first three sections each offered a maximum score of 15, whilst the writing sample section was scored on a letter scale with a maximum score of T. This resulted in a maximum composite score of 45T.

Candidates were also increasingly required to demonstrate desirable non-cognitive qualities, their commitment to the constant updating of their knowledge and skills and their ability to communicate effectively with patients, colleagues and the public. Interviews were considered an optimal means of assessing these qualitative factors. The most common non-cognitive characteristics assessed were motivation for entry, interpersonal skills and character, maturity, extracurricular activities, communications skills, empathy, social awareness and self awareness (Johnson & Edwards 1991). Medical schools also commonly used the interview process to provide applicants with information about the medical course. Cognitive qualities such as knowledge, judgement and problem solving ability were less commonly assessed (Johnson & Edwards 1991).
In 1991, interviews represented a key component at all but 3 American medical schools (Edwards, Johnson & Molidor 1990). At 60% of schools, at least 2 interviews were conducted per applicant (Johnson & Edwards 1991). When conducted, interviews contributed to the final ranking for selection in most schools and received an average weighting of 35% (Johnson & Edwards 1991). A national survey conducted by the MCAT Validity Studies Advisory Group in 1993 revealed that selection interviews, MCAT, GPA, letters of evaluation from academic referees and knowledge and commitment to healthcare were considered the most important elements of selection (Johnson & Edwards 1991; Mitchell, Haynes & Koenig 1994).

The majority of medical schools conducted one-on-one interviews. In 75% of medical schools, interviews were standardised for all applicants, however in most cases interview questions were not regulated by the admissions committee (Edwards, Johnson & Molidor 1990; Johnson & Edwards 1991). Interviewers were most commonly community members, academics, medical students, alumni, admissions staff and residents (Johnson & Edwards 1991). In a majority of schools, the assessment of characteristics of interest was achieved using Likert scales, adjective scales or behaviourally-based scales. However, final ratings commonly consisted of only a single summary narrative comment or a recommendation to accept or reject the candidate (Johnson & Edwards 1991).

Most medical schools provided some form of interviewer training. This often included questioning techniques but rarely addressed issues such as interviewer bias and interview structure (Johnson & Edwards 1991). Only a fifth of schools
evaluated the effectiveness of their interviews and even fewer assessed the reliability of their interviewers (Johnson & Edwards 1991).

Inconsistencies between medical schools in the way different selection instruments were used to assist in the admissions process and an ongoing under-emphasis of personal characteristics in the admissions process prompted a review by the AAMC. In 2001, Dr Jordan Cohen, President of the AAMC, called for medical schools to consider using an Medical College Admission Test (MCAT) threshold to eliminate high-risk applicants from consideration and then to use personal non-academic characteristics to determine selection for admission (Albanese, Farrell & Dotti 2005b; Kreiter 2007).

Currently in the United States, requirements for admission to medical school vary from school to school and include minimum academic levels (indicated by undergraduate grade point averages), performance in the medical college admissions test (MCAT), and interview to assess one or more of a range of non-academic characteristics (Parry et al. 2006). Medical Schools often use a regression approach to screen applicants. This approach combines admissions data scores into a weighted equation to produce a total score which is then used to generate a predicted USMLE Step 1 score that is used to rank applicants for admission (Albanese, Farrell & Dotti 2005a). An example is the equation utilized by the University of Iowa medical school is

*Predicted USMLE Step 1 = 91.62 + (2.27*Total MCAT) + (17.30*Science GPA)*

(Kreiter 2007).
At some medical schools, other tools such as referee reports, personal statements and biographical facts receive variable weighting in the selection process. To be eligible for selection into Harvard medical school, candidates are required to have completed at least one year of college biology, physics, mathematics and expository writing, and at least 2 years of college chemistry. At least 15 hours of literature, languages, the arts, the humanities and the social sciences at the college level are also recommended. (Harvard University 2010c) Selection for admission involves consideration of academic records, MCAT performance, an applicant essay, letters of evaluation, evidence of extracurricular activities, summer occupations and experience in the health field, and performance during a one-on-one selection interview (Harvard University 2010b). No minimum GPA or MCAT scores are stipulated, however academic excellence is required for candidates to be competitive for selection. In 2005, the average GPA of candidates selected into the course was 3.76, and average MCAT scores were 11.01, 12.09 and 12.03 for the verbal, physical sciences and biological subscales respectively (Harvard University 2010a).

For all American medical schools, high levels of academic achievement and performance on the MCAT continue to be critical factors in determining whether candidates are accepted into medical training. The chance of admission increases with increasing GP and overall MCAT score, and admission is unlikely for those with GPA below 3.0 or an MCAT score under thirty. (see Table 1.2). However, consideration of non-cognitive qualities now plays an increasingly important role in the selection process. Depending upon the nature of a medical school’s selection tools and procedures, suitability for the medical profession based on non-cognitive
factors may result in exclusion from admission, regardless of academic rank. In other cases, candidates that display highly desirable non-cognitive qualities may be selected.

Table 1.2 Table 24: MCAT and GPA Grid for Applicants and Acceptees to U.S. Medical Schools, 2009-2011 (aggregated) adapted from (AAMC 2012)

<table>
<thead>
<tr>
<th>Total MCAT Scores</th>
<th>5-14</th>
<th>15-17</th>
<th>18-20</th>
<th>21-23</th>
<th>24-26</th>
<th>27-29</th>
<th>30-35</th>
<th>36-38</th>
<th>39-45</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.80-4.00</td>
<td>Acceptance rate %</td>
<td>1.4</td>
<td>4.4</td>
<td>16.7</td>
<td>26.0</td>
<td>41.7</td>
<td>66.0</td>
<td>82.1</td>
<td>89.8</td>
<td>91.5</td>
</tr>
<tr>
<td>3.60-3.79</td>
<td>Acceptance rate %</td>
<td>0.6</td>
<td>3.4</td>
<td>10.5</td>
<td>18.4</td>
<td>28.0</td>
<td>50.6</td>
<td>71.7</td>
<td>79.9</td>
<td>84.5</td>
</tr>
<tr>
<td>3.40-3.69</td>
<td>Acceptance rate %</td>
<td>1.4</td>
<td>3.0</td>
<td>9.6</td>
<td>16.7</td>
<td>23.0</td>
<td>36.5</td>
<td>58.7</td>
<td>66.8</td>
<td>73.0</td>
</tr>
<tr>
<td>3.20-3.39</td>
<td>Acceptance rate %</td>
<td>0.6</td>
<td>1.0</td>
<td>7.8</td>
<td>12.7</td>
<td>17.9</td>
<td>25.9</td>
<td>38.9</td>
<td>51.9</td>
<td>61.6</td>
</tr>
<tr>
<td>3.00-3.19</td>
<td>Acceptance rate %</td>
<td>0.5</td>
<td>2.9</td>
<td>6.2</td>
<td>11.3</td>
<td>16.3</td>
<td>24.3</td>
<td>30.0</td>
<td>42.2</td>
<td>44.4</td>
</tr>
<tr>
<td>2.80-2.99</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>1.0</td>
<td>4.2</td>
<td>10.7</td>
<td>15.8</td>
<td>15.6</td>
<td>24.8</td>
<td>32.5</td>
<td>33.3</td>
</tr>
<tr>
<td>2.60-2.79</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>1.7</td>
<td>4.3</td>
<td>6.9</td>
<td>10.7</td>
<td>14.7</td>
<td>22.1</td>
<td>25.0</td>
<td>21.1</td>
</tr>
<tr>
<td>2.40-2.59</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>0.0</td>
<td>2.2</td>
<td>3.8</td>
<td>6.1</td>
<td>10.1</td>
<td>19.0</td>
<td>22.8</td>
<td>13.3</td>
</tr>
<tr>
<td>2.20-2.39</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>0.0</td>
<td>2.2</td>
<td>4.1</td>
<td>11.4</td>
<td>10.7</td>
<td>3.3</td>
<td>15.8</td>
<td>25.0</td>
</tr>
<tr>
<td>2.00-2.19</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>2.3</td>
<td>2.3</td>
<td>3.8</td>
<td>5.3</td>
<td>0.0</td>
<td>7.7</td>
<td>50.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1.47-1.99</td>
<td>Acceptance rate %</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>12.5</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>All</td>
<td>Acceptance rate %</td>
<td>0.4</td>
<td>2.2</td>
<td>8.1</td>
<td>15.5</td>
<td>24.8</td>
<td>42.8</td>
<td>61.4</td>
<td>73.5</td>
<td>81.2</td>
</tr>
</tbody>
</table>

1.3.9 Canada

At the beginning of the 18th century, fewer than 100 medical practitioners were recognized in Canada. No formal medical schools existed at this time and aspiring medical students were forced to learn by apprenticeship or travel abroad for
Medical practice was of variable quality and this led to the creation of licensing boards composed of designated physicians in Canada after 1790 (Kett 1967). Unlike their American counterparts, licensing boards in Canada provided an effective means of regulation throughout the 19th century (Kett 1967). In the absence of a robust Canadian medical education system, licensing boards accepted medical practitioners with degrees from institutions in Britain, Europe and the United States, and those licensed by British licensing faculties and colleges (Kett 1967).

Formal medical education lagged behind practitioner regulation. The early proprietary medical schools in Canada were renowned for their low entry standards and ruthless competition for potential students. The first university medical school opened at the Montreal Medical Institution (McGill University) in 1823 (MacDermot 1955). The school was opened in an attempt to raise professional standards of medical practice in Canada (MacDermot 1955). Dedicated attempts were made to strengthen and centralize medical education over subsequent decades with the founding of medical schools at the University of Montreal (1843), University of Toronto (1944), University of Ottawa (1945), Universite Laval (1953), Queens University (1954), Dalhousie University (1968), University of Western Ontario (1881) and University of Manitoba (1883) (MacDermot 1955). As in the United States, proprietary medical schools were active in Canada, however these were not able to grant medical degrees (Kett 1967). Effective regulation of practitioners, restrictions on proprietary schools and focused development of university teaching helped prevent the deterioration in medical standards that occurred in the United States during the 19th century.
At the middle of the 19th century, medical training at university medical schools in Canada involved a mix of apprenticeship, school-based lectures, and clinical experience in a hospital (Gidney & Millar 1994). Limited data is available regarding criteria governing selection of students into the early university medical courses. Flexner reported a relatively low matriculation standard at the University of Toronto which required admission candidates to hold junior matriculation rather than two or three years of pre-medical studies in a college (Flexner 1910). However, no Canadian medical schools were found inadequate by Flexner and none were forced to close or merge. Indeed, all Canadian medical schools reviewed by Flexner achieved an A grade within 20 years suggesting that all had implemented requirements for all candidates to have completed at least a high school diploma and two years of college or university study primarily devoted to basic science.

The major hurdle for licenced medical practice in the second half of the 19th century appears to have been obtaining a licence to practice following graduation from medical school. In 1865, the General Council of Medical Education and Registration of Upper Canada was formed and mandated that eligibility for licencing required four years of professional study, at a recognized medical school, at least a year of hospital experience and at least six months of study with a qualified practitioner (Kett 1967). In 1869, the College of Physicians and Surgeons of Ontario was incorporated and empowered to examine all candidate practitioners including university graduates (Kett 1967). These licensing bodies had the power to determine standards of medical education and to make these standards the foundation of the licence to practice (Gidney & Millar 1994). However, this also
served to hinder the attempts of universities to raise standards of education above those of the licensing bodies. In 1878 the University of Toronto medical school introduced a higher matriculation standard, annual graded examinations, clinical examinations in the senior years, more subjects, and more clinical work (Gidney & Millar 1994). Rather than producing better doctors, this resulted in a dramatic drop in student numbers as medical school applicants were drawn to less demanding courses (Harris 1976).

During the 20th century, medical education and selection methods in Canadian medical schools developed in parallel with their American counterparts. At the middle of the century, all twelve Canadian medical courses were graduate entry. Medical schools aimed to select entering medical students of integrity and intelligence with a solid foundation of basic medical science (Weaver 1954). Admission processes considered the demographic details (place of birth, legal residence and age), academic performance and personal attributes of candidates. Demographic details were collected on the admissions application form and most universities would allow only those candidates aged between 19 and 30 to progress through the selection process (Weaver 1954).

Admission candidates were required to have completed secondary school and at least three years of study in a college or university, with preference given to those with a bachelors degree (Weaver 1954). Most medical schools in Canada would only consider candidates with a premedical average of at least 65% (Weaver 1954). In cases where the academic performance of candidates were close to the threshold for admission, academic records were carefully scrutinised by faculty staff to select
for individuals with positive trends. Personal attributes were assessed by means of letters of reference and selection interviews. These methods selected for candidates demonstrating leadership, creativity, agreeable personality, reliability, and good interpersonal and communication skills. Other factors that may have been considered during interview include the health and physical status of candidates, their reading habits, their motivations for becoming physicians, and their career aspirations. (Weaver 1954).

Concerns regarding the ability of available selection procedures to select for students who would perform well within and beyond the medical course, combined with the subjectivity of interviews have provided the motivation for the ongoing refinement of selection processes. Since its inception in 1969, McMaster University has been at the forefront of developments in selection. In the late 1970’s, McMaster University was observed to have the “most comprehensive and exhaustive selection procedure” (Fulton 1979). The selection processes were praised for their rigor and dedication to selecting candidates with the academic competence and the capacity for self-directed learning, self-assessment and problem solving required by the McMaster medical course and by graduates (Fulton 1979).

As with other medical course selection procedures, the selection process at McMaster medical course considered candidates’ academic qualifications and reports from three confidential referees. However, candidates were also required to submit an autobiographical sketch, and a letter stating their goals and personal motivations. Candidates participated in a 45-minute interview and were required to participate in an observed group tutorial with other applicants. Performance in the
tutorial was assessed by a member of the university faculty of medicine, a medical student and a community member not affiliated with the school (Fulton 1979).

In 2003, steps were taken to formalise the cognitive and non-cognitive requirements in the form of a policy document produced by the Council of Ontario Faculties of Medicine that outlined the essential skills and abilities required for the study of medicine (Council of Ontario Faculties of Medicine 2003). In many ways, the document paralleled the earlier recommendations of the Association of American Medical Colleges (Association of American Medical Colleges 1979). As a result, candidates for admission to medical schools in Ontario are required now to demonstrate key abilities that candidates require to be able to observe, communicate, and perform routine physical and cognitive tasks faced by doctors. The document also requires candidates to demonstrate compassion, integrity, altruism and motivation and to have the emotional stability to cope with the medical course and to develop mature and sensitive relationships with patients and caregivers. The document highlights the need for medical graduates to diagnose and manage health problems and provide comprehensive and compassionate care to their patients. The document has since been adopted by a number of other medical schools in Canada and the United States.

Selection procedures at the 17 graduate Canadian medical schools continue to align with those of American medical schools and involve the same three major processes; fulfilment of minimum academic levels (indicated by undergraduate grade point averages), performance in the medical college admissions test (MCAT), and interview. Some medical schools require applicants to submit an
autobiographical statement or written essay and some consider written referee reports. Final rankings for admission are most commonly calculated by a weighted average of all factors considered.

All Canadian medical courses require completion of at least 2 years of an undergraduate degree, most require 3 years and some require a 4 year baccalaureate (Association of the Medical Faculties in Canada 2008). Many schools require or recommend candidates to have completed a science-based undergraduate courses. Although academic achievement is based on grade point average (GPA) (converted to a score out of 4.0), there is no national standard for this calculation. In some medical schools, GPA is calculated over the entire course, at others it is calculated over the two most recent years of study, and in other cases it is calculated using scores obtained in particular subjects of interest. GPA requirements are commonly between 2.8 and 3.6 (University of British Columbia 2009; University of Toronto 2009). In addition, some schools favour local applicants. At the University of Calgary, local and non-local applicants require GPAs of at least 3.2 and 3.6 respectively over the best two years of study (University of Calgary 2009). At Dalhousie University, the requirements are 3.3 and 3.7 for local and non-local candidates respectively (Dalhousie University 2009). Scores required for entry are driven up further by intense competition for places. At the University of Calgary, the average (2 year) GPA of candidates selected into the medical course in 2010 and 2011 was over 3.8 (University of Calgary 2009).

Most medical schools require candidates to take the MCAT, however a few schools such as those at McMaster University, the University of Ottawa and Laurentian
University do not require applicants to take the test (ACMC 2005; Association of the Medical Faculties in Canada 2008). MCAT scores generally remain valid for up to 5 years, but some schools accept scores obtained up to 18 years prior to admission (Association of the Medical Faculties in Canada 2008). When included as a prerequisite, some medical schools place no threshold on MCAT performance, while others clearly stipulate MCAT requirements. At the University of Toronto for example, candidates require scores of at least nine in each section of the MCAT (University of Toronto 2009). At Dalhousie University, local applicants require an overall MCAT score of 24 with a minimum score of 8 in each of its three sections. Non-locals require an overall MCAT score of 30 with scores of at least 10 in each section (Dalhousie University 2009). At McGill University, candidates require an overall score of at least 30 (ACMC 2005; Association of the Medical Faculties in Canada 2008).

Candidates are generally short-listed for interview on the basis of academic merit. This most commonly involves a weighted average of GPA and MCAT. However in some schools, such as the University of Calgary, factors such as letters of recommendation and personal folios are also considered (University of Calgary 2009). Interview formats vary, however the MMI is routinely employed by many universities, including McMaster University, the University of Calgary and Dalhousie University. During MMI, candidates are typically confronted with 8-10 stations of 6-8 minutes duration and assess personal attributes considered desirable in medical students. McMaster University also involves candidates in a simulated problem-based learning tutorial to assess problem exploration ability and communication skills (Moruzi & Norman 2002).
1.3.10 Australia

Prior to the appearance of the Australian medical schools in the second half of the 19th century, all Australian doctors were educated and obtained their qualifications overseas (Barwon Health 2009). In 1862, the University of Melbourne medical school launched a five year undergraduate medical course. The course was available only to male students and selection into the course was on based on academic achievement. To gain admission into the course, candidates were required to pass secondary school matriculation examinations and must have achieved a high standard in any 6 of the 9 subjects taken (Allen 1914). From 1862, the medical school also assessed candidates for *ad eundem gradum* Melbourne medical degrees. To be eligible for examination, candidates were required to be medical graduates of any mainland European or British medical school (Russell 1977).

The number of medical students at the University of Melbourne increased from just four in 1862 to 180 in 1882 (Russell 1977). The decision to accept female students into the course from 1887, female students further increased numbers and by 1911, the intake of first year medical students had risen to 105. The number of medical students at the University of Melbourne increased further as the length of the course was increased to 5 years and 2 semesters in 1922 and then to a full six years in 1931 (Russell 1977). By 1937, the number of students taking the medical course prompted the consideration of additional selection criteria to limit entry into the course (Russell 1977). Aspiring medical students were soon required to enter and complete a common premedical year, before applying for entry into the medical course. Quotas for entry into first year were introduced in 1946 and increasing
consideration was given to ways of selecting the best and most suitable candidates for entry into the course (Russell 1977).

By the middle of the 20th century, medical schools had opened in Sydney (University of Sydney), Adelaide (University of Adelaide), Brisbane (University of Queensland) and Perth (University of Western Australia), Academic prerequisites for entry were established during this period, For example at the University of Adelaide candidates were required to have had passed either Elementary Physics or Elementary Biology before entering the medical course (Adelaide 2012). However the University of Melbourne was the only Australian university that limited entry into medical school via a comprehensive selection process (Sunderland 1954). In order to remove variability associated with final high school examination results, potential medical students were accepted into a common “premedical” year during which they studied biology, chemistry, physics and scientific method. The selection process for medical school primarily considered candidates’ end of year examination results, in combination with one or more of the following: survey of confidential reports, interviews, and intelligence and aptitude testing (Sunderland 1954).

Between 1959 and 1963, Melbourne and Sydney each opened a second medical school (at Monash University and the University of New South Wales respectively) and the first medical appeared in Tasmania (University of Tasmania) (Brooks, Doherty & Donald 2001). The opening of a medical course in Newcastle (Newcastle University) in 1973, and the establishment of Flinders University (Adelaide) in 1974, were significant events for their focus on problem-based
learning and increasing consideration of routine interviews as part of the selection process (Brooks, Doherty & Donald 2001). During the 1980’s Newcastle University also introduced a requirement that candidates take the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and performance on the test was considered as part of the selection process in an effort to assess aptitude for medicine.

By 1990 there were 10 undergraduate medical schools in Australia. All were located in state capital cities except the medical school at Newcastle University. Selection processes for all considered prior academic performance (based on final high school examination score), and some also considered performance at interviews and/or the UMAT.

The late 20th and early 21st century represented a period of expansion and change for Australian medical schools. The first graduate entry programs opened at Flinders University (1996), Sydney University (1997) and the University of Queensland (1997) (Geffen 1991). Admission was dependent upon acceptable performance on the Graduate Australian Medical School Admissions Test (GAMSAT) or Medical College Admission Test (MCAT) and an appraisal of candidates’ personal qualities through structured interview (Brooks, Doherty & Donald 2001). Soon after, Melbourne University and the University of Western Australia offered both undergraduate and graduate courses. This was followed by the opening of new graduate medical schools at the Australian National University, Griffith University, the University of Wollongong, Notre Dame University and Deakin University and new undergraduate medical courses at James Cook
University, Bond University and the University of Western Sydney. In 2011, the first postgraduate medical program in Australia opened at the University of Melbourne. These developments were accompanied by significant revision of curricula, assessment and selection processes in universities across Australia. Major changes included the restructuring of medical curricula from a discipline based approach to a theme-based approach and a standardisation of selection processes. In both cases, increasing importance has been placed upon non-cognitive attributes and skills.

Currently, medical training in Australia is available to high school or university graduates. Some medical schools run courses for undergraduates (school leavers) or university graduates, whilst others (such Monash University and the University of Western Australia) run separate courses for undergraduates and graduates. In addition, two universities (Flinders University and the University of Queensland) run graduate medical courses but provide for undergraduates that are required to complete a 2 year undergraduate basic science course before progressing into the medical course. The duration of a medical degree varies although undergraduate courses are typically five or six years in length whilst graduate entry courses last four years (Medical Deans 2008). Table 1.3 (extracted from Medical Deans 2008) shows the type and duration of medical course offered at each Australian medical school.
Applications for admission into undergraduate Australian medical schools are lodged through state and territory based agencies that service applications for entry into a range of tertiary courses in that state or territory. These include the Victorian Tertiary Admissions Centre (VTAC) in Victoria, the Universities Admissions Centre (UAC) in New South Wales and the Australian Capital Territory, and the Tertiary Institutions Service Centre (TISC) in Western Australia, Applications for admission into graduate Australian medical schools are handled by a central on-line application and matching system, the Graduate Entry Medical School Admissions System (GEMSAS). Applicants are required to put forward their six top preferences for study and are considered for admission at their preferred place of study. Applicants who are not offered a place at their preferred medical school may have their applications forwarded to their second (and possibly lower) preference school, where they may be considered for selection. In both cases, applications are forwarded to the nominated medical schools for their consideration.
For both undergraduate and graduate courses, selection for medical training is generally based on three primary factors; academic achievement, performance on standardised admissions tests, and selection interviews. Required standards of achievement are high and are driven up further by fierce competition for places. For undergraduate courses, the assessment of academic achievement is based on performance in standardised senior high school years and the Undergraduate Medicine and Health Sciences Admission Test (UMAT). For graduate courses, academic performance is determined by grade point average obtained during an undergraduate degree and the GAMSAT. UMAT/GAMSAT scores must generally be achieved within two years of application.

**Academic performance**

State wide performance rankings have emerged as the preferred measure of high school academic performance. Although each state or territory developed its own description of rank; University Admissions Index (UAI) in the Australian Capital Territory and New South Wales/NSW, Tertiary Entrance Rank (TER) in South Australia, the Northern Territory, Western Australia and Tasmania, Equivalent National Tertiary Entrance Rank (ENTER) in Victoria, and Overall Position (OP) in Queensland), the Australian Tertiary Admission Rank (ATAR) was adopted in 2010 as a standardized measure of senior high school performance by all states and territories except Queensland. The ATAR describes the overall position of a student in relation to the student body for that year across the state. Representing an aggregate of scaled marks in English and other units undertaken, it is the primary academic criterion for admission into undergraduate medical courses in Australia. Australian Tertiary Admissions Ranks of between 95 and 99.5% are generally required in order to be considered for selection into undergraduate medical courses.
Australian GPA is calculated from a maximum of 7.0. In most cases, a weighted GPA is calculated from the most recent undergraduate degree, although the number of years considered in the calculation, and weightings awarded per year vary between institutions. There is a requirement for the undergraduate degree (or an acceptable postgraduate degree) to have been awarded no greater than ten years prior to the expected year of commencement of the medical course. Some medical schools reward students for postgraduate studies whilst others omit postgraduate grades from GPA calculations.

**Admissions Tests**

The UMAT is an aptitude test developed on a yearly basis by ACER to assist with the selection of students into undergraduate medicine and health science degree programs (Australian Council for Educational Research 2009). The test is designed to assess a variety of skills considered important to the study and later practice of medicine. It is divided into three sections that assess critical thinking and problem solving, understanding people and abstract non-verbal reasoning. The UMAT does not require any knowledge or skills in mathematics or sciences or any other area of the curriculum at secondary school level and is designed to complement, rather than replace academic results (Australian Council for Educational Research 2009). Only one undergraduate medical course (James Cook University) does not require applicants to complete the UMAT.
The GAMSAT was developed as a standardised selection instrument by ACER in collaboration with the graduate-entry medical schools to provide a means of ranking applicants according to a number of desirable cognitive qualities. The test comprises of three sections designed to assess proficiency in written English, and reasoning within social and scientific contexts and has been designed to ensure that individuals accepted into medical training possess requisite background knowledge and skills in written communication and problem solving. The assumed level of knowledge for the section on reasoning in the biological and physical sciences corresponds to the first year of university studies in biology and chemistry and final year high-school in physics.

**Interviews**

All but one medical school in Australia currently include an interview as part of the selection process (interviews were omitted from the selection process at the University of Queensland in 2008 in a decision that received some criticism (Harding & Wilson 2008)). Interviews may involve a single interview with a selection panel or, more commonly, a series of short, focussed interviews with independent interviewers (multiple mini-interviews). Candidates are generally short-listed for interview on the basis of academic merit and the number interviewed is usually approximately 1.5 times the number of available places (by mutual agreement between schools). In order to reduce bias, interviews are conducted by trained interviewers who are unaware of candidates’ performance on other criteria. The contribution of interviews to the final ranking of candidates for admission varies between schools.
To be eligible for admission to any medical course in Australia, applicants must also satisfactorily pass mandatory checks relating to criminal history and suitability to work with children and must have obtained requisite vaccinations.

A number of medical schools have introduced mechanisms that facilitate the entry of individuals with rural backgrounds and indigenous heritage into medical training. These mechanisms have been introduced to address equity issues and to help address a current shortage of medical practitioners in rural and remote areas, including indigenous communities. Another initiative to increase numbers of rural doctors includes the introduction of rural bonded scholarships that require students to commit to predefined periods of rural practice by the Australian government.

Most Australian medical schools accept full fee paying international students. Indeed the number of international students studying medicine in Australia has continued to rise over the past 20 years. Selection procedures for international candidates are not considered here in any great detail but generally involve assessment of academic achievement, and satisfactory performance on standardised admission tests and selection interviews. Academic requirements for overseas applicants are generally equivalent to those expected from Australian candidates.

The past 10 years has seen an increasing commitment to the evaluation and quality improvement of selection methods. Routine processes are in place at all medical schools to train interviewers and assess the reliability and performance of selection instruments. At the University of Sydney, a silent, trained observer may be present
to provide written comment to the Admissions Committee regarding the process and conduct of the interview (University of Sydney 2009b).

The following sections outline the factors considered for entry in Australian medical courses. These factors are summarised in Table 1.4.

Table 1.4 Primary Entrance Criteria for Australian medical courses

<table>
<thead>
<tr>
<th>Medical Course</th>
<th>History of academic performance</th>
<th>Standardised entrance test</th>
<th>Interview</th>
<th>Other factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Adelaide</td>
<td>Yr 12 results*</td>
<td>UMAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td>University of Tasmania</td>
<td>Yr 12 results+</td>
<td>UMAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td>University of Newcastle</td>
<td>Yr 12 results*</td>
<td>UMAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td>University of New South Wales</td>
<td>Yr 12 results*</td>
<td>UMAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td>University of Western Sydney</td>
<td>Yr 12 results*</td>
<td>UMAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>James Cook University</td>
<td>Yr 12 results*</td>
<td></td>
<td>Panel</td>
<td>Application form</td>
</tr>
<tr>
<td>Bond University</td>
<td>Yr 12 results*</td>
<td>UMAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>University of Sydney</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>Australian National University</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>Griffith University</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td>Wollongong University</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>MMI</td>
<td>Portfolio</td>
</tr>
<tr>
<td>University of Notre Dame (Fremantle)</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>Panel</td>
<td>Personal biography</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RUI</td>
</tr>
<tr>
<td>University of Notre Dame (Sydney)</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>Panel</td>
<td>Personal biography</td>
</tr>
<tr>
<td>Deakin University</td>
<td>GPA</td>
<td>GAMSAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>University of Western Australia</td>
<td>Yr 12 results</td>
<td>UMAT</td>
<td>Panel</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>Yr 12 results</td>
<td>UMAT</td>
<td>MMI</td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Monash University+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Queensland</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flinders University</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of Melbourne</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* include accessory route for entry of non-school leavers into the course. In these cases GPA or other measure of university performance replaces Yr 12 results.

** Yr 12 results and UMAT for entry into undergraduate course (Clayton), GPA and GAMSAT for graduate entry (Gippsland)

### Undergraduate courses

**University of Adelaide**

The medical program at the University of Adelaide is a six year undergraduate course. Candidates are ranked for selection on the basis of 3 components: TER, UMAT and a structured oral assessment. The TER must be over 90 and UMAT requirements are high. Transfer into the undergraduate course from other tertiary courses is possible and candidates applying through this route are selected for entry on the basis of GPA (>5.0 credit average), UMAT and structured oral assessment (University of Adelaide 2009).

**University of Tasmania**

Selection into the 5 year undergraduate medical course at the University of Tasmania is based on both academic achievement and performance in the UMAT. Domestic candidates must have a Tasmanian minimum TER 95 and have obtained a satisfactory pass in Year 12 Chemistry and English. In addition, a sound background in Mathematics is considered desirable (University of Tasmania 2009).
An alternate entry pathway also exists for up to 10 Australian applicants each year based on performance in a number of local biomedical and science courses. To be eligible to apply for the available places, candidates must have completed a full-time year of study in the year prior to the proposed Medicine enrolment and achieved at least a Distinction average (70%). Candidates must also have satisfied the year 12 pre-requisites for Medicine and achieved a combined UMAT score at least equal to the lowest UMAT score in the cohort selected in other pathways (University of Tasmania 2009).

University of Newcastle
The 5 year undergraduate medical program at the University of Newcastle and University of New England selects applicants who have reached a high academic standard and have personal qualities important to practise medicine. Entry requirements for the joint program are based on UMAT results, UAI and interview (Newcastle 2009). GPA replaces UAI for non-school leavers and the Rural/Remote Admission Scheme (RRAS) provides an important means of facilitating the entry of those with a significant rural background into medical training.

University of New South Wales
Students are selected for entry into the 5 year UNSW undergraduate medicine program on the basis of past academic achievement, results of the UMAT and performance at an interview. Each receives equal weighting in determining final rankings for admission. The minimum ATAR for eligibility is 95.00, however the median ATAR required for entry is always greater than 99.60. There are no
prerequisites for entry into UNSW Medicine (University of New South Wales 2009).

Transfer from other tertiary courses is possible. Applicants who have completed at least a year of undergraduate tertiary studies are assessed on the basis of both high school and undergraduate results combined in the ratio of 1:1. Those who have completed less than the equivalent of three-quarters of one year full time of tertiary studies prior to the year of intended entry are assessed only on the basis of their high school qualifications. In all cases, candidates are required to attain an academic rank of at least 95 to be eligible for consideration for entry into medical training. In addition, up to 15 places are available for UNSW Bachelor of Medical Science (BMedSc) to enter year 4 of the Medicine program with advanced standing. For this cohort, rank for entry is determined by the weighted (1:1:1) average mark for the first two years of BMedSc, the UMAT score and interview results (University of New South Wales 2009).

University of Western Sydney

The medical course at the University of Western Sydney is primarily a 5 year undergraduate course, however transfer from other tertiary courses is possible. To be eligible for selection, candidates are required to meet minimum thresholds for academic achievement, as measured by ATAR (or equivalent) or GPA, and performance on UMAT and interview. Interviews are offered to all satisfying academic and UMAT thresholds and are of MMI format. For those applying for entry in 2010, candidates required an ATAR of 95 or a GPA of 5.5. UMAT thresholds are determined by competition between candidates. Final rankings for
admission are based on UMAT score and interview performance (University of Western Sydney 2009).

**James Cook University**

Candidates are selected for entry into the 6 year undergraduate medical program at James Cook University on the basis of academic performance in year 12 studies (or tertiary studies for non-school leavers), a written application form and performance at interview (James Cook University 2009). Academic requirements are driven up by fierce competition and, in 2009, non-school leavers were required to achieve a GPA of 5.75 to be considered for an interview. The interview is designed to assess the suitability of candidates’ attributes to a career in medicine. Candidates are selected for interview on the basis of their application and academic results (James Cook University 2009).

**Bond University**

Bond University medical school accepts students into a 4 year and eight month undergraduate medical program. Candidates are selected on the basis of academic achievement and demonstration of personal attributes such as communication, leadership, ethical decision-making skills, motivation, critical thinking, problem-solving and non-verbal reasoning. Candidates are required to have successfully completed Year 12 chemistry, maths and English. Year 12 results and performance on the UMAT are used to shortlist applicants for interviews which are of the MMI format (Bond University 2009).
Graduate Medical Schools

University of Sydney

Selection for entry into the 4 year graduate medical course at the University of Sydney is based on performance (GPA) in an undergraduate degree, the Graduate Australian Medical School Admissions Test (GAMSAT) and a selection interview (University of Sydney 2009a).

Candidates must have an undergraduate degree, or be in the final year of an undergraduate degree at an approved Australian or overseas university. There are no prerequisite courses or subjects. In addition, no restriction is placed on the duration of study required to achieve the degree, although those who have undertaken an appropriate two-year bachelor degree in an accelerated program (e.g. three semesters per year) may only apply after completion of the full degree (University of Sydney 2009a). The Bachelors degree (or completion of a postgraduate degree at the diploma level or above) must have been awarded within 10 years of application.

A weighted GPA (1:1:1) is calculated for the last 3 years of the most recent undergraduate study. Candidates require a GPA of at least 5.5 (7.0 scale) at the time of application (equivalent to a credit level at the University of Sydney or 65% average) in order to be considered for admission (University of Sydney 2009a).

Postgraduate coursework qualifications are not used in the calculation of a GPA, however, applicants who have completed a PhD or a Masters degree by research from a recognised institution within 10 years, and whose GAMSAT results have
met the standard requirements are considered eligible for interview regardless of their undergraduate GPA. Applicants with PhD by research receive a bonus of 3 percentage points to their total final score (University of Sydney 2009a).

The GAMSAT is a compulsory component of the application process and candidates require a minimum score of 50 in each section of the GAMSAT to be considered for entry, however short-listing of candidates for interview is based on overall GAMSAT score (in conjunction with GPA). The minimum GAMSAT scores for admission in 2005, 2006, 2007, 2008, 2009 and 2010 were 58, 59, 61, 61, 59 and 63 respectively (University of Sydney 2009a). The number of candidates short-listed for interview is between 1.5 and 2 times the total number of places available (University of Sydney 2009b).

Interviews are of the MMI format and are designed to assess a number of qualities considered desirable by the medical school such as good communication skills, an empathic and sensitive approach, effective decision-making ability, teamwork, an holistic approach and motivational factors (University of Sydney 2009b). The MMI consists of 9 different scenario-based stations of 7 minutes duration conducted by independent interviewers. Two minutes are provided between stations for interviewers to complete marking sheets and for candidates to prepare for the next station. All interviewers are trained volunteers classified as academics, senior students and graduates from the Medical Program or Bachelor of Dentistry program or community members (University of Sydney 2009b).
**Australian National University**

Grade point average, UMAT score and performance at interview are used to select candidates into the four year graduate course at ANU. Candidates are short-listed for on the basis of combined weighted GPA (score out of 42) and overall GAMSAT score. In 2010, minimum weighted GPA was 34 and minimum overall GAMSAT was 55. Interviews are used to assess a number of non-cognitive attributes considered desirable of medical students and doctors (Harris & Owen 2007). These include communication skills, problem solving, resilience and maturity, enthusiasm for medicine, ethics and awareness of common issues in medicine. Interviews involve an eight station MMI and serve only to exclude those considered unfit for entry into the medical course (Australian National University 2009). Interviewers are drawn from 4 key stakeholder groups; community members, recent medical graduates, university academic staff and health professionals (Harris & Owen 2007).

**Griffith University**

Entry into the 4 year graduate medical course at Griffith University is determined competitively on the basis of GPA obtained during an undergraduate degree, GAMSAT results and performance at interview (Griffith University 2009a). Applicants who meet the minimum GPA requirement will be selected for interview based on their overall GAMSAT score. Scores required to receive an interview are determined on a yearly basis. (Griffith University 2009b). Candidates must have attained a GPA of at least 5.00 to be eligible for entry and final rankings for admission are determined by a 1:1 weighting of performance on the GAMSAT and interview (Griffith University 2009a). The GPA is not included in the calculation of
the final score used to rank applications (Griffith University 2009c). Interviews are conducted by a panel of three trained interviewers representing three key stakeholder groups (community members, university academics and clinicians) and are designed to assess suitability based on personal qualities identified as desirable in the medical student and future practitioner. These include communication skills and pro-social attitude, learning style, decision-making ability, personal motivations, management and self-evaluation skills. Each interview lasts 45 minutes and candidates are scored independently by all three interviewers against prescribed criteria on a five-point, criterion-defined rating scale (Griffith University 2009b).

**University of Wollongong**

Admission to the 4 year University of Wollongong medical course is based on weighted GPA of a completed Bachelor's degree, GAMSAT score, admissions portfolio and selection interviews (University of Wollongong 2009). To be considered for admission, candidates must be in their final year or have a completed bachelor degree from an approved Australian or overseas university. In either case, weighted Grade Point Average (GPA) must be greater than 5.0 on a 7 point scale or 2.8 on a 4 point scale at the time of application. Candidates must also have achieved a minimum score of 50 for each of the three sections on the GAMSAT with an overall minimum score of 50 (University of Wollongong 2009).

All applicants are required to prepare an admission portfolio. The admissions portfolio requires candidates to consider qualities considered desirable in medical students and graduates: leadership, teamwork, service, ethics, diversity of
experience and achievement or performance in an area of human endeavour (University of Wollongong 2009). The portfolio constitutes a significant proportion of the criteria required for interview selection and is considered the single most important component of the admissions process (University of Wollongong 2009).

Scores from applicants GPA, GAMSAT and portfolio (including rurality score) are used in the calculation short-list candidates for interview. Interviews are offered to approximately 1.3 to 1.7 times as many applicants as there are places available and take the form of a scenario based, multi station process. Ranking for admission offers is determined by combined interview and portfolio scores. The candidates GPA and GAMSAT results do not contribute to the final ranking of applicants (University of Wollongong 2009).

Rural scholarships are offered. A subquota of three places is available to candidates of Aboriginal or Torres Strait Islander descent that reach acceptable levels of performance in GAMSAT, interview and GPA (University of Wollongong 2009).

Candidates need to have completed and hold a current approved First Aid Certificate (minimum 14hr training content) prior to commencement of study and must comply with the NSW Health Department Circular Occupational Screening and Vaccination Against Infectious Diseases. This requires candidates to obtain proof of immunity status or be vaccinated against diphtheria, tetanus, pertussis, measles, mumps, rubella, chicken pox, hepatitis B, influenza and tuberculosis. Candidates must also complete a criminal record check and a ‘Prohibited Employment Declaration’ form (University of Wollongong 2009).
**Notre Dame University Fremantle**

To be eligible for admission into the 4 year graduate medical course at Notre Dame Fremantle candidates require a GPA of at least 5.0 and a minimum score of 50 in each of the three sections of the GAMSAT. GPA is based on the final 3 years of the most recent Bachelors degree and a bonus of up to 0.5 is available to Masters and PhD graduates (Notre Dame University (Fremantle) 2009). Candidates in the final year of a bachelor degree may be eligible for provisional selection but are only accepted into the course if final GPA meets the standard required for admission (Notre Dame University (Fremantle) 2009).

Candidates are short-listed for interview on the basis of GPA, GAMSAT score and a measure of rural background termed the Rural University Index (RUI). The use of the RUI serves to select for candidates most likely to help address the shortage of doctors in rural areas. Other factors that may be considered include “outstanding personal qualities reflected in a history of community or other service” or “other exceptional abilities likely to be relevant to the practice of medicine” (Notre Dame University (Fremantle) 2009).

Short-listed candidates are required to demonstrate commitment to a career in medicine and personal qualities consistent with the aims and philosophies of the medical schools in a 40 minute panel interview. Each interview panel is composed of three trained interviewers comprising a medical practitioner, an academic and a community member (Notre Dame University (Fremantle) 2009).
Final rankings for admission are determined by a composite score derived from interview score, overall GAMSAT score, GPA and RUI (Notre Dame University (Fremantle) 2009).

Notre Dame University Sydney

In addition to GPA, GAMSAT score and interview score, selection processes at Notre Dame Sydney require candidates to submit the Notre Dame Supplementary Information Form. The form outlines the applicant’s motivation to pursue a career in medicine and to train at Notre Dame and provides for bonus scores that may be allocated to applicants with a rural background or higher research degree (Notre Dame University (Sydney) 2009).

Candidates are required to have completed a Bachelors degree in any field of tertiary study. The GPA is calculated from the final three full-time equivalent years of the most recently completed bachelor or postgraduate degree. No minimum GPA or GAMSAT score is prescribed but selection for interview is unlikely for those with GPA below 5.0 and/or overall GAMSAT score below 50 as interviews are granted on the basis of a composite score derived from GPA, GAMSAT score and the Supplementary Information Form (Notre Dame University (Sydney) 2009).

Offers are made on the basis of a ranked list of applicants with the ranking derived from a final score consisting of Supplementary Information Form score, GPA, GAMSAT and Interview with each contributing approximately 25% to the final score (Notre Dame University (Sydney) 2009).
Students are required to have completed an Australian registered senior first aid course, have evidence of full immunisation, obtain State and Federal Government Police Clearance, and a working with children clearance prior to enrolment (Notre Dame University (Sydney) 2009).

**Deakin University**

The 4 year graduate medical course at Deakin University accepted its first cohort of medical students in 2008. Selection is based on three key factors; GPA, GAMSAT score and interview performance. To be considered for admission, candidates require a weighted GPA of at least 5.0 and a GAMSAT score of 50 overall (and at least 50 in each of the three sections) (Deakin University 2009). The weighted GPA is based on the final three years of the most recent undergraduate degree. The final three years are weighted sequentially form the final year to the third last year in the ratio of 3:2:1 such that the final year is weighted most heavily. This course (or at least 1 year of postgraduate study at the Graduate diploma level or above) must have been completed within the last 10 years.

Candidates are short-listed for interview on the basis of combined, equally weighted GPA and GAMSAT scores. Candidates with prior clinical experience, demonstrated financial hardship and those from rural areas receive additional weightings of 2% to 4% (Deakin University 2009). Selection interviews follow the MMI format and initially involved 10 discrete 8 minute interviews, each conducted by an independent, trained interviewer, although interviews were subsequently reduced to 5 minutes duration on the basis of the results of this study. Each mini-interview is scenario based and designed to assess a characteristic considered
desirable in rural medical students and medical practitioners. These are communication skills, commitment to rural and regional practice, evidence-based practice, self-directed learning, teamwork, motivation, social justice, professionalism and effective resource use (Deakin University 2009).

Final rankings for admission are determined by an equally weighted combined score derived from GPA, GAMSAT and interview. Offers are made in order of ranking until all places are filled.

**Medical schools with undergraduate and graduate intakes**

**University of Western Australia**

To be considered for entry into the 5 year undergraduate medical course at the University of Western Australia, candidates must meet minimum thresholds in academic achievement (TER>96) UMAT score and interview performance. Candidates are invited to participate in structured interviews that assess a range of desirable personal qualities on the basis of academic merit. Final ranking of candidates depends on weighted (2:2:1), combined score of TER, UMAT and interview. Tertiary transfer into the medical course is possible and is assessed similarly, although academic achievement is measured via GPA rather than TER. The minimum threshold for academic achievement is GPA of at least 5.5 (University of Western Australia 2009b).

The University of Western Australia also accepts graduate students into a 4 year (plus bridging course) medical program. Candidates are short-listed for interview based on past academic performance as measured by GPA and performance on the
GAMSAT and are accepted into the course on the basis of performance in all three selection criteria (University of Western Australia 2009a).

Monash University

The 5 year undergraduate medicine course at Monash University is available only to current Year 12 students or applicants who have completed VCE no more than two years previously and have not undertaken any subsequent tertiary studies. Selection into the course is based on performance in UMAT, an interview and ENTER or equivalent Year 12 results. Applicants are required to have Year 12 English and Chemistry as pre-requisite subjects (Monash University 2009b).

Ancillary pathways into medical training are also available for those from rural or indigenous backgrounds through the Dean’s Rural List and Dean’s Indigenous List for applicants (Monash University 2009a). These mechanisms provide for additional interviews to be offered to eligible applicants not initially offered an interview. To be eligible for entry via the DRL, applicants must have resided for at least 5 years in Australian rural areas.

The 4 year graduate medicine course at the Gippsland campus of Monash University is only available to people that have previously successfully completed undergraduate degree. Selection into the course is determined by GPA, GAMSAT results and interview performance. To be considered for admission, candidates required a weighted GPA of at least 5.0. The GPA is weighted over the last three years of the candidate’s undergraduate course in the ratio of 3:2:1 so that greatest weight is given to the final year of study. Candidates with PhD or Masters by
Research qualifications receive additional bonuses of 0.2 and 0.1 respectively. Candidates also require a score of at least 50 in each of the three GAMSAT sections and must attain a threshold overall score that is determined annually (Monash University 2009c).

Candidates are short-listed for interview on the basis of GPA and GAMSAT results. The interview process utilises the MMI format and involves a series (5 to 10) of short interviews (approximately 10 to 15 minutes each) totalling less than 2 hours. Final ranking for admission is based on GAMSAT and interview performance (Monash University 2009c).

All successful applicants require a health check and have up-to-date and appropriate immunisation status. A satisfactory police check is required for each student (Monash University 2009c).

**University of Queensland**

Entry into the 4 year graduate medical program at the University of Queensland requires an undergraduate or higher level degree with a GPA of 5.0 and scores of at least 50 in each of the 3 sections of the GAMSAT (University of Queensland 2009a). All candidates must fulfil these requirements to be considered for entry into the course, including students accepted into dedicated positions for those who have lived for a significant period in rural northern Australia.

Preference is not given to any particular degree or field of prior study. However, the key degree, or any subsequent postgraduate study must have been completed
within 10 years of the expected MBBS program commencement date. In the latter case the GPA obtained in postgraduate research studies may be used instead of, or combined with, the key degree GPA. PhD or Masters (by research) degrees receive a converted GPA of 7.00 (University of Queensland 2009b).

Rankings for admission are based on GAMSAT scores. The cut-off score for admission in 2010 was 63 overall and 67 for major offer round for the CSP places. The GPA is not used to rank applicants, but is used as a tiebreaker (University of Queensland 2009b).

Final year secondary school students can apply for entry into the medical program at UQ via the Year 12 MBBS Provisional Sub Quota scheme. Under this scheme, students are accepted into an undergraduate degree at UQ before starting the MBBS graduate program. Entry requires a QLD OP 1, or the equivalent rank of 99* from other high school studies, and a minimum scaled score of at least 50 in each section of the UMAT (University of Queensland 2009c).

**Flinders University**

The 4 year graduate medical course at Flinders University was expanded in 2010 to provide both graduate and undergraduate avenues into medical training. Candidates for graduate entry must negotiate three selection elements. For graduate entry, candidates must have completed (or be in the final year of completion) an undergraduate course of at least three years duration, and must also have completed the GAMSAT and a structured selection interview. Candidates are short-listed for
interview on the basis of GAMSAT score. Final ranking for admission is based on a combined score from GPA, GAMSAT and interview (Flinders University 2009).

Candidates for the undergraduate stream apply for entry into a two year Bachelor of Medical Science (or Bachelor of Health Science) degree with subsequent automatic entry into the 4 year Bachelor of Medicine and Bachelor of Surgery course. Selection involves two major elements: results from the GAMSAT and TER of which TER is the most heavily weighted (Flinders University 2009).

**Postgraduate**

**University of Melbourne**

The first 4 year postgraduate medical course in Australia opened at the University of Melbourne in 2011. Candidates applying for entry into the Doctor of Medicine program are required to complete an approved three or more year bachelor degree-including prerequisite second year subjects of anatomy, physiology & biochemistry, as well as the GAMSAT and a multiple mini-interview (MMI) (University of Melbourne 2009).

A minimum weighted GPA is set prior to the application period and is usually between 5.5 and 6.0. The weighted GPA is derived from grades achieved in the last three years of the applicant's most recent undergraduate degree, with the years progressively weighted in the ratio of 1:2:3. Subsequent postgraduate qualifications are not included in the GPA calculation. All sections in the GAMSAT are weighted equally with no minimum GAMSAT score (University of Melbourne 2009).
Applicants are short-listed for the MMI on the basis of combined ranking of their GPA and GAMSAT results. Interviews are offered to 125% of the final number of students to be selected and will assess candidates’ motivation for undertaking medical training, their communication skills and ability to develop and maintain rapport with patients in a clinical situation, and their aptitude for collaboration and decision making (University of Melbourne 2009).

1.4 Issues raised and lessons learnt

The apprenticeship model of learning remained the dominant model of medical training from the very beginnings of medical thought until the Renaissance period in the West and even later in the East. Over this period, a number of attempts were made to institutionalize medical training, and it was not uncommon for medical training to be available concurrently through apprenticeship or more formal institution-based medical training. The coexistence of the two forms of medical training is identifiable in both Western and Eastern cultures.

Under the apprenticeship model, despite a lack of standardized admission requirements and processes, certain qualities have been commonly sought in candidates seeking entry into medical training. The qualities sought generally reflected those personal attributes considered “good” within the contemporary social and moral context and were broadly consistent across Greek, Islamic, Christian, Buddhist, Taoist and Confucian ideology. Indeed, the methods of selection and qualities sought in candidates for medical training have typically shown more similarity across these cultures than the way medicine has been conceptualized and practised. Despite the common goal of selecting those with
desirable qualities into medical training, the internalisation of these values into common medical practice in all cultures was limited by a lack of standards for selection and training, and deficient regulation of the profession.

The institutionalisation of medical training within the university systems of India and Europe during the Middle Ages, and attempts to regulate medical education from the Chinese Imperial court, involved a shift from personal selection of medical students to a more standardized system based approach. Personal attributes were considered less important than a commitment to conform to societal and moral norms. Also notable was the trend away from selection based on qualitative personal qualities factors towards the use of quantitative variables such as desirable socio-economic backgrounds, demonstrated academic prowess and financial capacity. This trend was closely associated with the move towards university based training and is likely to be partially a result of the usefulness of marks in providing a relatively objective measure of an applicant’s ability to retain information, grasp with complex issues and be trusted with the care of human life.

It was not until the first half of the 20th century that attention was drawn back to non-cognitive attributes and the development of tools designed to select for candidates with desirable qualities and aptitude for a career in medicine. Since then much work has been devoted to the assessment of selection tools that attempt to assess aptitude for medical practice and key non-cognitive qualities such as aptitude tests and selection interviews.
This historical survey of selection methods has identified several issues for consideration:

- How important is it to select appropriate individuals into the study of medicine and what is the impact of selection processes upon the medical profession and society more broadly?
- What does history teach us about the importance of cognitive and non-cognitive qualities in the selection of medical students?
- What are the relative benefits and risks associated with selection into medical training by individuals (eg for apprenticeship) versus selection by standardised institutional processes?

The history of medical student selection is chequered and intricately associated with evolving socio-cultural expectations, regulations and knowledge. Whilst there has been an ongoing, deep-seated social expectation that doctors should be both competent and virtuous, the medical profession has not always met these expectations in the eyes of contemporary society. Of particular note are those periods in Galen’s time and in nineteenth century America, when both the competence and virtuosity of physicians were questioned.

Common to both these periods was a lapse in the rigor of the selection process leading to a relative abundance in medical trainees that were introduced into the profession to serve the needs of their teachers rather than the qualities of the applicants or the needs of their patients. In both cases, this led to a situation where medical training was highly variable and often suboptimal. Importantly, these were also times when insufficient regulatory hurdles were in place at the level of
registration or professional practice to prevent the promulgation of substandard or unscrupulous doctors into society.

At other times in history, imperfections in the selection process were sufficiently countered by the virtuous ideals of the gatekeepers to training, and/or the presence of regulatory requirements prior to practice. For example, in Ancient Greece and Islam, the choice of an appropriate apprentice by a physician was made likely by his own virtuosity, in 19th century Britain, the failure to consider non-cognitive qualities in applicants for medical school was largely overcome by an increasing requirement to remonstrate competence in a standardised fashion prior to graduation and requirements to commit to professional codes of conduct.

Because of the interplay between social expectation, cultural shifts and evolving bureaucratic processes, it is difficult to determine the relative importance of cognitive and non-cognitive qualities in the selection of medical students. It is possible however to say that both types of qualities are desirable in doctors and are also expected of them by society. This means that there is an intrinsic need for selection processes to consider both cognitive and non-cognitive qualities, and also for regulatory processes to be in place to ensure that standards of professional practice do not fall once individuals have entered the profession.

Self-selection selection of apprentices predominates historically as a selection method. The success or failure of this approach has been largely dependent upon the integrity of the medical profession and the ability of the profession to self regulate. As a selection method it is, by nature likely to be plagued by levels of
variability and risk that would be unacceptable by today’s standards. The gradual shift towards a more systematised and objective approach has been necessary and has resulted in the development of a far more transparent and accountable process. This type of approach took root in the UK during the late 19th century where the acceptance of academic grades as a marker of cognitive ability, in the absence of objective measures for non-cognitive qualities, is likely to have contributed to the dominance of previous academic performance in the selection of applicants into medical training in the UK and Australia during the first half of the 20th century. The need to consider and assess additional personal qualities and motivations in an objective way then led to development of standardised aptitude tests and interview processes.

The historical events described in this chapter provide evidence that changes in the approach to medical student selection can have far-reaching social consequences and remain testament to the importance in striving for selection processes that ensure that appropriate individuals are selected into medical training. Once admitted into medical training, few individuals fail to become doctors. Because of this, selection represents the major hurdle motivated individuals must negotiate in order to become doctors. As such, the importance of the selection process cannot be understated. The following sections provide a critical appraisal of the reliability and validity of a variety of tools used to select applicants into medical training and an overview of the key non-cognitive qualities sought in applicants.


Anonymous 1881b, 'The student of today, the practitioner of tomorrow', *British Medical Journal*, vol. 2, no. 1080, pp. 461-8.


Anonymous 2000, *Chapter 1. Education in Ancient and Medieval India*,


Associaten of Ameican Medical Colleges, 2012, *MCAT and GPA Grid for Applicants and Acceptees to U.S. Medical Schools, 2009-2011 (aggregated)* accessed 15 September 2012,


Australian National University 2009, *ANU medical school; admission*,


Barwon Health 2009, *Australian medical pioneers index*,


Bloch, H 1988, 'The impact of social and political developments on the teaching of medicine', *Journal of the National Medical Association*, vol. 80, no. 8, pp. 916-8.


Cilliers, L & Retlef, F 2006, 'Medical practice in Graeco-Roman antiquity', *Curationis*, vol. 29, no. 2, pp. 34-40.


Ebbel, B 1937, 'The Ebers Papyrus', in Levin and Munksguard, Copenhagen.


Fulton, T 1979, 'The address to the students', paper presented to Opening of the Winter Session Royal Victoria Hospital, London, 4 October.


Gidney, R & Millar, W 1994, 'The reorientation of medical education in late nineteenth-century Ontario: The proprietary medical schools and the founding of the Faculty of Medicine at the University of Toronto', *Journal of the History of Medicine and Allied Sciences*, vol. 49, pp. 52-78.


Harris, S & Owen, C 2007, 'Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School', *Medical Education*, vol. 41, pp. 234-41.


Ibn Hubal 1943-4, Kitab al- mukh tarat, Hyderabad.

Inryoo, I 1964, Seiji yooryaku, surviving chapters collected vol. 28, Yoshikawa kabunkan, Tokyo.

Inter-departmental Committee on Medical Schools 1944, Report of the Inter-departmental Committee on Medical Schools (Goodenough Report). HMSO, London.


Jackson, E & Graham, C 1986, 'Helping students write the MCAT essay', The Advisor, vol. 5, no. 3, p. 3.


Kreiter, C 2007, 'A Commentary on the use of cut-scores to increase the emphasis of non-cognitive variables in medical school admissions', *Advances in Health Sciences Education*, vol. 12, pp. 315-9.


Lee, Y 1978, *The Medical History of Early Singapore.*, vol. publication no.14:7-19., SEAMIC, Southeast Asian Medical Information Center, Tokyo, Japan:


Leung, A 2003, 'Medical popularization and instruction in Ming-Qing China', *Late Imperial China*, vol. 24, no. 1, pp. 130-52.


Medical Council of India 1997, Regulations on Graduate Medical Education.


Ni, M 1995, The Yellow Emperor's Classic of Medicine (the Neijing Suwen), Shambhala Publications, Boston.

Notre Dame University (Fremantle) 2009. Undergraduate medicine: entrance requirements,


Ormerod, H 1945, *The early history of the Liverpool Medical School from 1834 to 1877*, Liverpool University Press, Liverpool.


Richards, T 1985, 'Medical education in India - in poor health', *British Medical Journal*, vol. 290, no. 13 April, pp. 1132-5.

Roth, C 1953, 'The qualification of Jewish physicians in the Middle Ages', *Speculum*, vol. 28, no. 4, pp. 834-43.


Thurman, J 1979, 'Some notes on the admission of graduate applicants to the Medical Schools.', Medical Education, vol. 18, pp. 4-9.


Zhang, Z 1963, A treatise on febrile and miscellaneous diseases, People's Health Press, Beijing.
2. Reliability and validity of methods used to select applicants into medical school

2.1 The aim of medical student selection

Entry into medical school represents the primary hurdle to becoming a doctor as progression rates in medicine are very high. Medical educators have therefore become aware of the importance of developing and utilising rigorous and effective selection processes. Selection tools must be fair, transparent, evidence-based and legally defensible (Benbassat & Baumal 2007). There has also been an increasing expectation that they provide useful information about the ability of candidates to perform within and beyond the medical course. As a consequence, there has been an increasing tendency to expand selection processes beyond the cognitive domain and to increase the involvement of key stakeholder groups in selecting appropriate candidates for medical training.

Selection has two primary purposes: to select the best candidates available at the time, and to select candidates that will perform well (Smyth 1946). It is important to select candidates that can cope with the demands of the medical course and those that are capable of gaining the skills and knowledge that will enable them to perform well on assessments administered during the course. However, it is more important to select those candidates that will perform best beyond medical school, that is, those that will make good doctors (Lowry 1992). Thus, admission tests for medical school should measure the cognitive abilities and personal qualities that enable an individual to do well in school as well as the qualities of a good physician.
143

(Frazer 2005). The information used to select candidates should then reliably
distinguish and adequately predict those who will succeed in the science, clinical,
and other components of medical education (Moruzi & Norman 2002).

2.2 Reliability and validity

In order for a test to be useful for selection it must be both reliable and valid so as
to provide an accurate measure of a pertinent trait or quality (Fruen 1983). Reliability
and validity are inextricably linked. Indeed, reliability provides an
estimate of the upper limit on validity (Conway, Jako & Goodman 1995). Reliability
provides an assessment of how well an instrument measures a particular
quality. However, validity is also affected by variations in relationships between
predictor and criterion constructs (Conway, Jako & Goodman 1995).

2.2.1 Reliability

Nunnally defined reliability as “the extent to which measurements as repeatable and
that any random influence which tends to make measurements different from
occasion to occasion is a source of measurement error” (Nunnally 1967). Tests are
reliable if they provide an accurate measure of a candidate’s performance in a
domain of interest. Reliability is most commonly reported as a value between zero
and one with higher values indicating higher reliability. Several types of reliability
are recognised. Three which are of particular relevance to instruments used for
selection are internal consistency, test-retest reliability and inter-rater reliability
(Conway, Jako & Goodman 1995).
Internal consistency represents a measure of the reliability of items intended to measure the same quality or characteristic, and is based upon the variability of candidate scoring across these items. The most commonly used measure of internal consistency is Cronbach’s alpha (Cronbach 1951), a statistic calculated from the pairwise correlations between items according to the following formula:

$$\alpha = \frac{K}{K-1} \left( 1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2} \right)$$

where $K$ is the number of components (K-items or testlets), $\sigma_X^2$ the variance of the observed total test scores, and $\sigma_{Y_i}^2$ the variance of component $i$ for the current sample of persons.

Cronbach’s alpha is often used to provide a measure of the “adequacy” of the internal consistency of a test. Although no real metric exists for judging the adequacy of the statistic values above 0.7 are generally considered adequate (Cortina 1993).

Test-retest reliability describes the likelihood that a candidate will achieve identical scores when faced with the same test on separate occasions. This assumes that there is no change in the characteristic being measured during the period between the two occasions. Test-retest reliability is measured using a correlation coefficient that compares performance on each occasion.
Inter-rater reliability refers to the consistency of scores awarded by different assessors to the same items. For dichotomous or categorical variables, inter-rater reliability may be measured in terms of the percentage of agreement between assessors or via a kappa statistic that considers percentage agreement but also takes into account the amount of agreement that could be expected to occur due to chance (Cohen 1960; Fleiss 1971). For continuous variables, the correlation between assessors’ scores provides a measure of inter-rater reliability. This may be expressed by a number of measures, the usefulness of which is dependent upon the nature of the test and number of assessors. The Pearson (Pearson & Filon 1898) and Spearman (Spearman 1904) coefficients provide rank-based comparisons of rater scores along an ordered scale. However, the coefficients do not consider differences in absolute scores and are not practical for assessing the reliability of large numbers of raters. The Intra-Class Correlation coefficient (ICC) (Everitt 1996) provides a reliability measure based upon correlations between assessors’ scores and differences in scores and is especially useful in providing an estimate of inter-rater reliability when small numbers of assessors are involved and when data sets are complete. Krippendorff’s alpha (Krippendorff 2007) is a particularly useful measure of inter-rater reliability as it has sufficient flexibility to account for any number of coders, incomplete (missing) data, to any number of values available for coding a variable. It also accommodates binary, nominal, ordinal, interval, ratio, polar, and circular metrics and adjusts itself to small sample sizes.

Standardised entrance examinations, such as the MCAT, represent the most reliable selection instruments followed by GPA, whilst the reliability of measures of non-cognitive attributes is often considerably lower (Fruen 1983). Standardised
entrance examinations are composed primarily of numerous multiple choice type questions. This serves to improve test reliability and to reduce examiner variability (Smyth 1946). The high reliability of standardised entrance examinations is a key advantage of these tests over other selection instruments (Fruen 1983). Such tests utilise large numbers of multiple choice questions that are administered and marked consistently across all candidates. Previous academic performance, as measured by GPA for example, is also a relatively reliable measure as it represents a composite score derived from performance over several years and across multiple domains and assessments. Differences in course requirements, assessment formats and marking between universities does however, reduce the reliability of the GPA (Fruen 1983). The reliability of interview-based tests is often more difficult to assess. Despite less, standardisation and objectivity, reliabilities reported for panel interviews are often higher than those from separate interviews. This is because reliability in a panel interview depends upon degree of agreement over interpretation and evaluation of applicant performance (inter-rater reliability), whilst reliability in separate interview formats may reflect both inter-rater reliability and consistency in applicant performance across interviews (internal consistency) (Conway, Jako & Goodman 1995).

Assessments that measure non-cognitive attributes are prone to be less rigorous, less structured, and less objective (McGaghie 2002). All these factors decrease reliability. The reliability of selection instruments used to measure non-cognitive qualities may be increased by maintaining a structured format, using standardised questions, response evaluations and methods for combining scores, by training assessors and by ensuring those qualities measured are relevant to the desired
outcome (Conway, Jako & Goodman 1995; Huffcutt & Arthur 1994). Achieving these outcomes is a particular challenge for tasks that assess non-cognitive attributes as assessment of the quality cannot be readily captured by checklists. Nevertheless, if a consistent, valid approach is taken to the conduct of such tasks, acceptable reliability can be achieved through use of global rating scales for the assessment of non-cognitive attributes tasks (Hodges et al. 1999).

2.2.2 Validity

Validity refers to the degree of agreement between a test score or measure and the quality it is believed to measure. This implies that performance on tests that are valid will predict performance in real-life situations that incorporate the characteristics or qualities measured by the test.

Validity may be further considered in terms of construct validity (the extent to which a test measures the construct or concept it is intended to measure) and criterion validity (how well performance on a test reflects performance on other tasks that measure or require the same qualities or characteristics that the test intended to measure). Criterion validity may be further considered in terms of concurrent and predictive validity. Concurrent validity describes how well the test results agree with other current measures of the same characteristic, while predictive validity describes how well the test results agree with future measures of performance in the domain of interest.

Selection tests for medical school are valid if they assess cognitive and non-cognitive attributes that are truly related to performance as a medical student and
doctor. The ability of tests that assess cognitive attributes to predict medical school examination success has been studied extensively, however, less is known about their ability to predict performance beyond medical school. Still less is known about the ability of tests for non-cognitive traits to predict performance during and after the medical course. The value of current selection instruments in predicting performance during and beyond medical school are discussed below.

Correlation coefficients may be calculated to determine how well the results of a test concur with other present or future tests of the same characteristic. However, the validity of selection tools for admission into medical school is difficult to assess as:

- one selection instrument may assess a number of variables
- variables measured may be difficult to define or subjective
- test validity is influenced by test reliability
- tests may utilise surrogate markers (questions or scales considered appropriate to assess or reflect a given domain) to estimate real-life performance within a domain
- there is an assumption that performance on a selection instrument is widely generalisable, however results achieved may be influenced by the nature of the test (method specificity)
- concurrent validity is difficult to assess because different selection instruments measure different variables
- predictive validity is difficult to assess because qualities measured are not static, but are subject to change. Personal attributes may be influenced by
medical education and experience. Sociocultural perceptions of qualities may also vary over time.

In addition, correlation coefficients used to assess the predictive validity of selection instruments are likely to be considered underestimates of $r$ because of a restriction of range produced by admission of only the highest performers into medical school (Donnon 2007; Emery & Bell 2009).

### 2.3 Reliability and validity of commonly used selection tools

In broad terms, the tools most commonly employed in the selection of medical students may be categorised in the following way:

- Demographic data
- Personal statements, Biographical data and Referee reports
- Academic Record
- Standardised Admission Tests
- Psychological Testing
- Problem-Based Learning Tasks
- Interviews

Medical schools commonly utilise records of past academic achievement, standardised entrance examinations, and selection interviews to admit suitable candidates into the medical course. The first two of these components are used for their ability to provide information about the cognitive abilities of candidates. Interviews provide valuable information about candidates’ non-cognitive qualities.
that may be supplemented by additional factors such as letters of reference, personal statements and essays.

The following sections discuss the reliability and validity of these tools.

2.3.1 Demographical data, biographical data and References

2.3.1.1 Age and gender

An Australian study that explored the relationship between the intra-course performance of medical students and a variety of demographic variables including age found that older students were less likely to be ‘not satisfactory’ than younger students (Kay-Lambkin, Pearson & Rolfe 2002).

Lumb and Vail (Lumb & Vail 2004) found male gender predicted relatively poor performance in the first three years of medical school. This was consistent with the findings of a systematic review of the literature undertaken by Ferguson and colleagues (Ferguson, James & Madeley 2002), who observed that women consistently performed better than men in their medical training, were more likely to attain an honours degree, and more likely to perform better in clinical assessments than men. However, these observations were based upon few studies and the review by Ferguson and colleagues identified several studies that reported no gender difference or better male performance during and after the medical course (Ferguson, James & Madeley 2002). The meaningfulness of any observed gender differences is also unclear as the observed differences are generally small and provide limited insight regarding performance as a doctor.
2.3.1.2 Ethnicity

Data relating ethnicity to performance is limited and somewhat dated. In an Australian cohort of students completing the Bachelor of Medicine/Bachelor of Surgery course between 1986 and 1990, being Caucasian was found to be a highly significant predictor of obtaining honours (James & Chivers 2001). Conversely, Aboriginal and Torres Strait Islander medical students and overseas medical students were found to have had difficulty in the first year of medical school (Kay-Lambkin, Pearson & Rolfe 2002). Aboriginal and Torres Strait Islanders were 3.1 times more likely to be ‘not satisfactory’, whilst overseas students were 1.5 times more likely to be ‘not satisfactory’ than other students. Differences in educational experience prior to medical school, course experiences, learning styles and personal factors were considered to be contributing factors.

Liddell and Koritsas (Liddell & Koritsas 2004) observed that there were differences in performance between medical students of Indigenous and ethnic (India, Asia and the Middle East) backgrounds, compared to Caucasian medical students; which was also evident in the final year of the medical course. Performance measured through final year OSCE scores, General Practice scores and overall score of the final two years of the course were all poorer for Australian medical students of Aboriginal and Torres Strait Islander, or ethnic backgrounds. Membership of a minority ethnic group has also been found to be associated with examination underperformance and problems during medical training in both the United Kingdom and the United States (Ferguson, James & Madeley 2002; Yates & James 2006).

Lower literacy skills in Indigenous and ethnic students may at least partially explain these observations. Cooper found an association between English as a
second language and lower reading grades (Cooper 2003). In a study of minority students, McGlinn and Jackson found that the reading subtest score was the only component of a medical college admissions test that predicted medical course performance (McGlinn & Jackson 1989).

Social disadvantage may be another contributing factor. Albanese and colleagues (2003) posited that certain admission and selection tools may be confronting for underrepresented and disadvantaged background students and that it may be stress associated with combating issues of disadvantage that prevents success in medical school rather than academic ability (Albanese et al. 2003).

2.3.1.3 Personal statements, Bibliographic data and Referee reports

Personal statements have been found by some authors to correlate weakly with performance (Ferguson 2003). This has not been a consistent finding, however leading other authors to describe them as inadequately researched and extremely variable in their usefulness in predicting performance in or beyond the medical course (Albanese et al. 2003; Pilotto et al. 2007). Albanese and colleagues further suggest that the usefulness of such statements in providing meaningful information is likely to be limited by the possibility that applicants may recruit help in preparing such statements and the lack of a standardised format (Albanese et al. 2003).

Albanese and colleagues also dismissed the usefulness of referee reports in providing information about personal characteristics of medical school applicants as the applicants themselves chose the writers (Albanese et al. 2003). There have been cases of fraudulent letters of support written by the student or others, and there
are challenges associated with non-standard formats so that reliable comparison and assessment is difficult (Pilotto et al. 2007).

Although data specific to the medical profession is limited, biographical data has been found to be a moderate predictor of job success (average validity 0.20-0.40) and reference checks less so (average validity 0.16-0.17) (Hunter & Hunter 1984). The interests of individuals may predict the particular occupation, or occupational family, that a person will choose (Holland 1985; Savickas & Spokane 1999). However, interests are poor predictors of performance once one has entered an occupation (Schmidt & Hunter 1998). Estimates of validity correlation coefficients for candidate interest approximate 0.10 (Hunter & Hunter 1984).

2.3.1 Academic Record and Grade Point Average (GPA)

Academic record has been an integral component of the selection process over the past century. In the middle of the 20th century, Bloomgarden considered grades to be the best criterion for selection tests (Bloomgarden 1957). Subsequent research suggests that past academic performance has significant utility in predicting future performance during the medical course and on licensing examinations. This observation is likely to result from a number of factors, both cognitive and non-cognitive, that are required for academic success. Such factors may include, intelligence, motivation, commitment, and other important personality and attitudinal characteristics for lifelong learning (McManus et al. 2005).

Secondary school records represent the major source of previous academic achievement for selection committees of undergraduate medical courses. A clear
relationship between secondary school performance and performance at medical school has been consistently observed (McManus et al. 2005). When compared to other selection instruments, high schools grade point average and global examination results were found to be the best predictors of academic success early in undergraduate medical courses with moderate to strong correlations between high school grade point average and performance during the first year of an undergraduate medical course.\( (r=0.41-0.61) \) (Touron 1987). High school grades for the individual disciplines of science and humanities have also been found to correlate moderately \( (0.3-0.4) \) with future performance as a medical student (Gough 1979; Roessler et al. 1978).

Grade Point Average (GPA) has been regularly used as a selection tool for graduate medical degrees. Although potentially affected by non-student factors, such as variability in grades assigned by different institutions and different courses, GPA has consistently been observed to correlate with performance during the medical course. Coca and colleagues found GPA to be a better predictor of medical school performance than the MCAT (Coca, Sakakenny & Johnson 1976). A recent systematic review and meta-analysis identified a wide degree of variation in reported correlation coefficients between GPA and performance during the medical course (Ferguson, James & Madeley 2002). The meta-analysis considered 753 correlation coefficients for undergraduate performance (total sample size of 21 905 participants) and 32 correlation coefficient for graduate training performance (total sample size of 2487 participants). In this study, Ferguson and colleagues observed a relatively strong average effect size of 0.30 (95% confidence interval 0.27 to 0.33, \( P<0.00001 \)) for undergraduate performance and a weaker more variable effect of
0.14 (95% confidence interval 0.05 to 0.23, P<0.05) for postgraduate performance. Once corrected for unreliability in both the predictor (previous academic ability) and outcome (medical training success) variables and for restriction of range, the coefficients increased to 0.48 (0.40 to 0.51) and 0.24 (0.08 to 0.37) respectively, indicating that up to 23% of variance in undergraduate medical school performance, and up to 6% of variance in graduate medical school performance can be explained by previous academic performance (Ferguson, James & Madeley 2002).

These figures are consistent with the findings of more recent studies in Australia and the USA. At the University of Queensland, Wilkinson and colleagues observed GPA to be more strongly correlated with academic performance than aptitude test and interview results ($r=0.47$) (Wilkinson et al. 2008). The association between GPA appeared strongest for the early years of the medical course ($r_{(Year \ 1)}=0.45$, $r_{(Year \ 4)}=0.36$) (Wilkinson et al. 2008). Correlation coefficients of between 0.43 and 0.51 have been reported for GPA and performance across 6 Australian graduate medical schools (Coates 2008). In the USA, undergraduate GPA correlated with performance both early ($r=0.54$) and later ($r=0.46$) across 14 American medical courses (Julian 2005). White and colleagues found GPA to be predictive of both academic performance ($r=0.26-0.44$) and clinical performance ($r=0.17$) (White, Dey & Fantone 2009). These findings suggest that the predictive value of grades decreases as students move from early pre-clinical experiences to the clinical aspects of their training (Ferguson, James & Madeley 2002; Fruen 1983; Kreiter & Kreiter 2007).
Undergraduate GPA appears to have substantial utility in predicting performance on licensing examinations. White and colleagues found GPA to be strongly predictive of performance on US licensing examinations ($r=0.26-0.41$) (White, Dey & Fantone 2009). Other studies have found undergraduate GPA to be a useful predictor of performance on US licensing examinations, but less useful than MCAT score (Ogunyemi & Taylor-Harris 2005; Veloski et al. 2000). However, GPAs were found to have the more utility than MCAT in predicting overall performance on the Canadian licensing examinations (Moruzi & Norman 2002). Kulatunga and Norman (Kulatunga-Moruzi & Norman 2002) reported a correlation coefficient of 0.31 for GPA and performance on Part 1 of the Canadian Licensing examination.

GPA has also been found to predict clinical reasoning performance on the first ($r=0.29$) part of the Canadian Licensing exam clinical reasoning ($r=0.19$) as well as problem solving ($r=0.25$) and data acquisition ($r=0.21$) on the second part of the Canadian licensing exam (Kulatunga-Moruzi & Norman 2002; Violato & Donnon 2005). However, the ability of GPA to predict performance on more complex practical tasks and real-life clinical encounters is unknown. Although some have reported a correlation between undergraduate GPA and performance during residency and beyond (Kreiter & Kreiter 2007), most authors report a lack of association between pre-medical GPA and clinical clerkship performance (Basco et al. 2002; Grey et al. 2001; Hughes 2002; Loughmiller et al. 1973).

Whether previous academic performance is associated with particular non-cognitive qualities is largely unknown but appears unlikely. Undergraduate GPA does not appear to predict subsequent communications skills performance (Moruzi & Norman 2002). In addition, no correlation has been observed between GPA and
subsequent scores on the Jefferson Empathy Scale (Hojat et al. 2002). Willoughby and colleagues did observe a correlation between preclinical GPA and conscientiousness, but not attitude, peer relations, patient rapport, maturity or integrity (Willoughby, Gammon & Jonas 1979).

Overall, previous academic performance appears to be a moderate predictor of performance as a medical student with a correlation coefficient in the order of 0.4. As such, previous academic performance is likely to account for about 20% of the variability of performance within the medical course (Parsell & Bligh 1995). Previous academic performance is also a useful predictor of performance on licensing examinations, however it appears to have limited predictive ability for performance in clinical settings.

2.3.3 Standardised entrance examinations

Standardised entrance examinations have been an integral part of the selection process in the United States of America since the introduction of the MCAT in the 1940’s. Since then, standardised entrance examinations have gained support internationally and now represent a core component of the selection process for many medical schools in Canada, Australia, the United Kingdom, Japan, Singapore and Germany. In most cases, nations have chosen to develop their own standardised tests, however Canadian medical schools have adopted the MCAT. Most standardized entrance examinations are aptitude tests that assess a mixture of fluid intelligence (logic and critical reasoning) and crystallised intelligence (consisting of general culturally acquired knowledge) (McManus et al. 2005). Standardised tests
have been specifically designed with the aim of determining aptitude for medicine and predicting performance (Fruen 1983; Mitchell, Haynes & Koenig 1994).

2.3.3.1 The Medical College Admissions Test (MCAT)

The MCAT represents the most researched and most revised standardised entrance test used for the selection of medical students. The MCAT gained widespread acceptance after performance on the original version of the test was shown to provide some indication of performance during the medical course. Poor performance on the test was associated with an increased chance of failure on medical school examinations. Chesnay and colleagues showed that elimination of candidates below the 10th centile on the MCAT would remove 26% of the failures, 16% of the mediocre students, 7% of the fair students, 3% of the good students and only 1% of the excellent students (Chesnay et al. 1936). Subsequent studies supported the notion that those performing poorly on the MCAT were likely to struggle during a medical course (Blackwell 1984; Jackson & Dawson-Saunders 1987; Jones & Vanyur 1984).

Through several revisions of the test, increasing data have become available that overall performance on the MCAT correlated moderately \( r=0.2-0.6 \) with subsequent performance on basic medical science and written clinical examinations during medical school (Friedman & Bakewell 1980; Mitchell 1990; Touron 1987). The science-related subscales within the test functioned best as predictors of medical school performance (Friedman & Bakewell 1980; Fruen 1983). Correlations between overall MCAT scores and performance on licensing examinations appeared even stronger than those observed for performance in the
medical course (r=0.35-0.70) (Mitchell 1990). When predicting licensing examination performance, MCAT data was shown to account for more criterion variability than GPA (Mitchell 1990). However, despite substantial research efforts, the broader predictive ability of the test has been difficult to clarify (Donnon 2007). This is likely to be at least partly due to difficulties measuring behavioural attributes (Mitchell 1990).

Attempts to evaluate the predictive ability of the MCAT and its subscales intensified following the most recent revision of the test in 1991. Preliminary data suggested that the predictive validity of the revised MCAT for performance during the preclinical years of medical schools was at least as good as that of previous forms of the test (r= 0.38-0.78) (Julian 2005; Mitchell, Haynes & Koenig 1994) (Julian 2005). Subsequent studies that assessed predictive validity for longer term academic performance produced broadly consistent results. Julian reported a correlation coefficient (corrected for restriction of range) of 0.46 between MCAT and clerkship performance, further suggesting that MCAT score may predict performance beyond the medical course (Julian 2005). In 2007, Donnon conducted a meta-analysis of studies published between 1991 and 2005 that examined the ability of the MCAT to predict future examination performance (Donnon 2007). The meta-analysis considered 12 studies that compared MCAT results to performance on medical school basic medical science examinations and 4 studies that compared MCAT scores to performance on medical school clinical examinations. The Pearson correlation coefficient for overall MCAT score and performance during preclinical years of medical school was 0.39, whilst coefficients for subscale MCAT scores ranged between -0.13 (writing sample
subscale) and 0.32 (biological sciences subtest) (see Table 2.1 below). The uncorrected Pearson correlation coefficient for overall MCAT score and performance during the clinical years of medical school was 0.34, whilst subscale coefficients were between 0.07 (physical sciences and written subscales) and 0.14 (verbal reasoning). Nineteen studies that compared MCAT scores to performance on one or more steps of the USMLE were included in the meta-analysis. Uncorrected Pearson correlation coefficients for overall MCAT score and performance on Step 1, 2 and 3 of the USMLE were 0.60, 0.38 and 0.43 respectively. Subscale correlation coefficients are show in Table 2.1 below.

<table>
<thead>
<tr>
<th></th>
<th>Performance on basic sciences examinations at medical school (n=12)</th>
<th>Performance on clinical examinations at medical school (n=4)</th>
<th>Performance on USMLE Step 1 (n=16)</th>
<th>Performance on USMLE Step 2 (n=10)</th>
<th>Performance on USMLE Step 3 (n=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall MCAT score</td>
<td>0.39 (0.21-0.54)</td>
<td>0.34 (0.29-0.39)</td>
<td>0.60 (0.50-0.67)</td>
<td>0.38 (0.26-0.49)</td>
<td>0.43 (0.32-0.54)</td>
</tr>
<tr>
<td>Biological sciences</td>
<td>0.32 (0.21-0.42)</td>
<td>0.12 (0.00-0.23)</td>
<td>0.48 (0.41-0.54)</td>
<td>0.30 (0.20-0.39)</td>
<td>0.11 (0.03-0.19)</td>
</tr>
<tr>
<td>Physical sciences</td>
<td>0.23 (0.09-0.36)</td>
<td>0.06 (-0.05-0.18)</td>
<td>0.47 (0.43-0.51)</td>
<td>0.25 (0.03-0.46)</td>
<td>-</td>
</tr>
<tr>
<td>Verbal reasoning</td>
<td>0.19 (0.12-0.25)</td>
<td>0.14 (0.02-0.25)</td>
<td>0.27 (0.19-0.35)</td>
<td>0.27 (0.22-0.32)</td>
<td>0.27 (0.20-0.34)</td>
</tr>
<tr>
<td>Written</td>
<td>-0.13 (-0.30-0.05)</td>
<td>0.07 (-0.05-0.19)</td>
<td>0.08 (0.02-0.14)</td>
<td>0.05 (-0.02 to 0.12)</td>
<td>-</td>
</tr>
</tbody>
</table>

Adjustment for restriction of range of the MCAT and its subtests produced the following correlation coefficients.
Table 2.2 Correlation between MCAT and USMLE performance (corrected for restriction of range)

<table>
<thead>
<tr>
<th></th>
<th>Performance on basic science examinations at medical school (n=12)</th>
<th>Performance on clinical examinations at medical school (n=4)</th>
<th>Performance on USMLE Step 1 (n=16)</th>
<th>Performance on USMLE Step 2 (n=10)</th>
<th>Performance on USMLE Step 3 (n=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall MCAT score</td>
<td>0.43</td>
<td>0.39</td>
<td>0.66</td>
<td>0.43</td>
<td>0.48</td>
</tr>
<tr>
<td>Biological sciences</td>
<td>0.40</td>
<td>0.15</td>
<td>0.58</td>
<td>0.38</td>
<td>0.14</td>
</tr>
<tr>
<td>Physical sciences</td>
<td>0.26</td>
<td>0.07</td>
<td>0.52</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Verbal reasoning</td>
<td>0.24</td>
<td>0.18</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Written</td>
<td>Not calculated</td>
<td>Not calculated</td>
<td>Not calculated</td>
<td>Not calculated</td>
<td>Not calculated</td>
</tr>
</tbody>
</table>

Donnon concluded that:

- The MCAT total has an adjusted medium predictive validity coefficient effect size for basic science and clinical performance.
- Performance on the biological science subscale of the MCAT was a medium predictor of preclinical performance.
- Correlations between all subscales and clinical performance were small.
- The MCAT total has a large predictive validity coefficient effect size for USMLE Step 1, and medium validity coefficients for USMLE Step 2 and Step 3.
- The writing sample has little predictive validity for both medical school performance and the medical board licensing examination.
- Domain and method specificity may account for differences in the ability of MCAT performance to predict medical school and licensing examination performance. Higher predictive ability for performance on licensing examinations may be due to common domains (knowledge and cognition) and methods (multiple-choice questions). Assessments during medical
school that involved other domains (e.g., practical skills) and methods (e.g.,
observations, checklists) may reduce the validity of the MCAT as a
predictor of performance during medical school.

Albanese and colleagues (Albanese, Farrell & Dotti 2005) showed that 54% of
USMLE Step 1 failures, occurred in those with MCAT biological sciences subscale
score below 8, physical sciences subscale score below 7, verbal reasoning subscale
score below 7 and an overall MCAT score below 22. These thresholds are broadly
consistent with levels of MCAT performance previously found to be associated
with increased risk of academic difficulty (Huff & Fang 1999; Jones & Vanyur
1984; Julian 2005). However, more recently, Kozar and colleagues did not find
lower MCAT scores to be predictive of failure on the surgical specialty licensing
examination (Kozar et al. 2007). However, it is of note that the mean MCAT score
of both successes and failures on the exam in this study were well above the
threshold value of 22 described above. Indeed, despite numerous revisions and
extensive changes to the MCAT, the predictive ability of the current MCAT is
remarkable for its consistency with data reported for the original test in the 1930s
(McGaghie 2002).

In contrast to the abundance of data linking MCAT results to academic
performance, only a few authors have attempted to assess whether the MCAT
performance is related to non-cognitive characteristics. Haley and colleagues
examined the association between MCAT performance and personality
characteristics (Haley, Juan & Paiva 1971). No difference in benevolence,
leadership or support was observed. However, low MCAT scores were associated
with higher dogmatism and conformity, and lower independence. Haley concluded that personality variables, personal interests and motivations may affect MCAT performance.

More recently, a small, positive correlation between MCAT verbal and writing sample scores and clinical OSCE scores has been considered supportive of a direct relationship between communication skills and clinical performance during medical school (Stephens & Reamy 2009). Overall MCAT and performance on the verbal reasoning subscale have been shown to be useful predictors of performance on the communication skills and clinical reasoning (r=0.24) components of the Canadian licensing examination (Moruzi & Norman 2002; Violato & Donnon 2005). However, no correlation has been observed between scores in the biological sciences, physical sciences, and verbal reasoning sections of MCAT and empathy (Hojat et al. 2002).

2.3.3.2 The Graduate Australian Medical School Admissions Test (GAMSAT)

The GAMSAT is considered to be most validated and generalisable aspect of the selection process in Australia (Coates 2008). Nevertheless, observations that GAMSAT results vary significantly with candidate gender, age and highest degree level have led some authors to raise issues relating to equity when aiming to select students from diverse backgrounds (Aldous et al. 1997; Oates & Goulston 2012).

Correlation between GAMSAT score and performance during the early years of the medical course appears to be in the order of 0.3. Donnelly (Donnelly 2006) found unadjusted correlations of around 0.3 between GAMSAT scores and estimates of
performance in years 1 and 2. Coates also observed a correlation in the order of 0.3 between overall GAMSAT score and performance on year 1 medical school examinations. Groves and colleagues observed a weak correlation (r=0.24) between medical school candidates performances in the GAMSAT and their subsequent performance on preclinical barrier (year 2) examinations (Groves, Gordon & Ryan 2007). More recently, Wilkinson and colleagues reported similar results (r=0.20) for correlation between GAMSAT score and medical school performance (Wilkinson et al. 2008). Wilkinson and colleagues observed GAMSAT to be a better predictor of performance early in the medical course (r=0.25) and for performance on written examinations (r=0.28) (Wilkinson et al. 2008).

Despite data suggesting correlations of 0.25-0.3 with performance early in the medical course, the predictive validity of the GAMSAT in terms of performance as a doctor remains largely unexplored (Oates & Goulston 2012). GAMSAT scores appear to be weakly correlated with performance on clinical examination and communication skills assessments, but there does not appear to be a relationship between GAMSAT score and competency-based assessments (Donnelly 2006). Groves and colleagues compared GAMSAT performance with performance on two validated tests of clinical reasoning; Clinical Reasoning Problems (CRP) and the Diagnostic Thinking Inventory (DTI) (Groves, Gordon & Ryan 2007; Groves, Scott & Alexander 2002). There was no association between GAMSAT scores and CRP performance and a weakly negative correlation (r=-0.31 - 0.05) was observed for performance on the DTI.
2.3.3.3 The Undergraduate Medical and Health Sciences Admissions Test (UMAT)

Recent studies suggest that, whilst a relatively reliable selection tool, the predictive validity of the UMAT may limit its usefulness as a selection tool. Wilkinson and colleagues (2011) found only a weak correlation (0.15) between UMAT score and performance during only the first year the medical course. Although corrected for restriction of range, Griffin (2011) advised caution when interpreting these findings as a variety of other factors such as further range restriction, based upon high school performance and the failure to consider non-cognitive variables, small sample size and unreliability of the test and outcome measures. These are all valid concerns as the way these factors are addressed may significantly affect the correlation coefficients achieved. Nevertheless, several subsequent studies have also found UMAT to be a relatively weak predictor of performance during the medical course, accounting for less than 10% of variation in performance (Poole et al 2012).

It has been suggested that both gender and language background may to influence performance on UMAT to some degree. Indeed, Wilkinson and colleagues (2011) found that men performed better in sections 1 and 3 of the UMAT whilst women performed better on Section 2. These observations may be of limited significance however, as such differential performances appear to be in the direction predicted by the underlying constructs (Mercer & Chiavaroli 2007).
Despite these questions surrounding its usefulness as a selection tool, the UMAT is used by all but one undergraduate Australian medical course to select new medical students.

2.3.3.4 The UK Clinical Aptitude test (UKCAT) & the Biomedical Admissions Test (BMAT)

The BMAT appears to have predictive validity for performance on preclinical examinations. It has been suggested that, at Cambridge University, BMAT score is likely to be a better predictor than anything else available of performance at year one (Cassidy 2008). Both the aptitude and skills section and the scientific knowledge section of the test have been shown to correlate with year 1 and year 2 performance. Correlation coefficients for the aptitude and skills section have generally been between 0.1 and 0.2 for both first and second year performance. Those for the scientific knowledge section have been between 0.2 and 0.5, with values tending to be higher for the first year (Emery & Bell 2009). The longer term predictive validity of the study has not been assessed.

The predictive validity of the UKCAT is currently unclear and is likely to remain so for some time given the recent introduction of the test. Because of these uncertainties, several authors have questioned the usefulness of introducing the test into the selection process of UK medical schools (Cassidy 2008; McManus et al. 2005). Early data suggests that UKCAT total or subset scores do not correlate with medical school year 1 examination scores (Lynch et al. 2009).
Standardised entrance examinations vary somewhat in their design and aims. Data suggests that the MCAT and GAMSAT have some utility in predicting performance during the medical course. The MCAT appears to be the stronger predictor of performance during the medical course (r= 0.4) and is likely to account for approximately 20% of variability associated with medical school performance (Donnon 2007). The MCAT also appears to be a moderate to strong predictor of performance on licensing examinations (r=0.4-0.6; 20-45% of variability), however its utility in predicting real-life clinical performance is unclear.

2.3.4 Psychological tests

Specific psychological tests that provide information about personality or mental capacity have not been routinely employed in the selection of medical students. As a result, data relating specifically to psychological testing as part of medical student selection is limited. Early research provided mixed data regarding the usefulness of incorporating psychological testing in medical student selection. A battery of tests that assessed 35 variables related to IQ, personality, socio-political and cultural factors was found to be “uniformly disappointing” in predicting performance during the medical course, of low reliability and of limited use in identifying successful students (Brosin 1948). Mental ability as identified by IQ tests has been identified as a factor that predicts academic success. Candidates with an IQ below 130 were found to struggle to cope with the demands of the medical course and life as a doctor (Waggoner & Ziegler 1946). More recently, however, MacManus and colleagues found no significant association between performance on the AH5 intelligence test and career progression or drop-out of the UK medical registry (McManus et al. 2005).
Brosin considered projective methods such as the Rorschach test to hold the most potential for identifying personality traits that may be desirable or undesirable of doctors (Brosin 1948). During the middle of the 20th century, Rorschach testing was considered as a possible alternative to the interview for selecting candidates more likely to perform well as medical students (Mullen 1948; Shoemaker & Rohrer 1948). However, the test was unable to identify a consistent profile for medical students and was considered to have low predictive validity (Eron 1954). When interpreted by trained professionals, projective tests have high reliability when administered on both patient and non-patient populations. Nevertheless, the requirement for interpretation by trained professionals in conjunction with insufficient predictive validity data and poor face validity largely precludes their use in medical student selection.

More recently, personality testing has been reconsidered as a tool for medical school selection (Wilson et al. 2012). Identifying relationships between an applicant’s personality traits and their subsequent performance may be useful in ensuring selection of candidates that will meet the demands of the different careers within the profession (Wilson et al. 2012). It may also be useful in identifying individuals likely to struggle during the course or feel dis-satisfied with their chosen career following graduation.

Using the Hogan Developmental Survey (HDS), Knights and colleagues (Knights & Kennedy 2007) found that borderline/schizoid and narcissistic/antisocial characteristics were negatively correlated with academic success. Students who had a very low or very high level of anxiety performed worse academically than those
with a moderate level (Grover & Smith 1981). However, this has not been observed in individuals with anxious predispositions (Pamphlett & Farnill 1995; Stewart et al. 1999). Tartas and colleagues found that students with higher anxiety levels at admission and lower levels of anxiety during first year of the medical course were more likely to be satisfied with a career in medicine (Tartas et al. 2011). Higher levels of depression towards the end of the medical course were negatively correlated with satisfaction with the chosen career. Levels of anxiety and depression were also associated with likelihood of work stress and burnout.

A recent systematic review of the literature that compared performance in and beyond medical training with performance on personality and intelligence tests found a negative correlation ($r=-0.22$) between ‘dominance’ subscale score on the California Personality Inventory (CPI) and performance on MCQ examinations during the medical course, and a positive correlation ($r=0.22-0.32$) between ‘well-being and achievement via conformance’ subscale score on the CPI and oral examination performance (Ferguson, James & Madeley 2002). Ferguson and colleagues also found that students scoring highly on the conscientiousness subscale of the Big 5 personality test were more likely to do well in preclinical assessments ($r=0.51$) but performed less well during their clerkships ($r=-0.20$) (Ferguson, James & Madeley 2002). Using a measure of personality integration (the sense of coherence (SOC-29) tool, (Antonovsky 1993), Tartas and colleagues recently found the characteristics of comprehensibility and manageableness to be associated with success during the medical course and satisfaction with a career in medicine (Tartas et al. 2011).
Specific instruments, such as the Personal Qualities Assessment (PQA) have been developed to assess cognitive skills and particular personality/attitudinal traits relevant to health professional practice (Lumsden et al. 2005). Application of the tool to a Scottish cohort of students revealed no significant correlations between separate elements of the PQA assessment and student performance. However, students identified to have ‘non-extreme’ character types on the involved-detached and on the libertarian-communitarian moral orientation scales were found to rank significantly higher (p=0.049) in OSCEs (Dowell et al. 2011). Despite uncertainty relating to its predictive validity, the PQA will be used by one Australian medical school in 2012 to select applicants for entry into the medical course in 2013 (Wilson et al. 2012).

Psychological tests are commonly used for selection of employees by commercial, industrial and government sectors (Bore, Munro & Powis 2009). It has been asserted that psychological and behavioural factors are the principal determinants of success in modern humans (Falconer 1983). This implies that occupational success may be facilitated by the use of psychological tests to select for individuals with desirable qualities. Established psychological tests represent thoroughly tested, reliable tools that are validated for their intended purpose during development. However, psychological tests, including IQ tests, are constructed from items under the assumption they constitute a representative sample of environmental demands on the problem solving behaviour repertoires of individuals (Harrington 1997). This assumption is clearly false when such tests are employed for purposes of selection into specialised fields of training and raises concerns regarding the generalisability and context specific validity of such tests when used as selection tools. Although
the use of tests that incorporate items that are not representative of a particular job of interest may limit the validity of such tests in predicting job performance, validity findings for IQ and personality measures are broadly consistent across occupations (Schmidt & Hunter 1998; Schmidt, Hunter & Caplan 1981).

Performance on intelligence tests correlates moderately with occupational status and job performance, with validity correlation coefficients generally falling between 0.3 and 0.6 (Bore, Munro & Powis 2009). The degree of correlation between performance on intelligence tests and job performance appears to be dependent on complexity of job; performance in more complex occupations correlate more strongly with intelligence test performance than do less complex occupations (Hunter 1980). Most personality research has been organized around the Big Five model of personality (Falconer 1983). This model considers 5 personality domains; extraversion, neuroticism, agreeableness, openness and conscientiousness. Personality tests have generally been found to display low validity for predicting overall job performance. In a meta-analysis of personality test validity outcomes, Morgeson and colleagues observed uncorrected and corrected median validity correlation coefficients of 0.10 (interquartile range 0.5-0.17) and 0.18 (interquartile rage 0.9-0.26) respectively (Morgeson et al. 2007b). Although higher correlations have been calculated by other authors (eg Harrington 1997), it is likely that these estimates are inflated by overcorrecting for factors such as predictor unreliability (Morgeson et al. 2007b). Of the Big Five personality domains, conscientiousness appears to be the best predictor of job performance (Bore, Munro & Powis 2009; Harrington 1997; Schmit & Ryan 1993). Schmidt & Hunter also explored the effect of IQ on personality test performance and
concluded that, once IQ had been accounted for, conscientiousness was likely to be the only personality trait that contributed to career success (Bore, Munro & Powis 2009).

Morgeson and colleagues advocated for the judicious use of tests that measure psychological characteristics for selection purposes and advised that (Morgeson et al. 2007a, 2007b);

- Selection panels should be aware of the limitations and low predictive validity of such tools
- Tests should incorporate customized personality measures that are clearly job-related in face valid ways and should be used in conjunction with other tools
- Tests should not be used as stand-alone be used in conjunction with other selection tools or processes.

2.3.5 Problem-based learning tasks

Problem-based learning (PBL) tasks have been employed as a selection tool by institutions that utilize PBL techniques as part of their medical course. However, PBL tasks designed to assess communication skills and problem exploration ability did not predict overall licensing examination performance or performance on components of the examination designed to assess communication skills, problem exploration skills and data acquisition (Moruzi & Norman 2002). The reliability (intra-class correlation) coefficient for the test was 0.61 (problem exploration score) and 0.65 (communication skills score) respectively.
2.3.6 Selection interviews

Selection interviews for medical school have been notable for their diversity in terms of aims, format, structure, interviewer characteristics and contribution to the final ranking of candidates for admission. Huffcutt and colleagues proposed a means of classifying the structure of selection interviews: (a) degree of standardization of questions, (b) standardization of response evaluation and (c) standardization of method for combining ratings (see Table 2.3) (Huffcutt & Arthur 1994; Huffcutt & Woehr 1993). Each of these factors was found to have an independent effect on interview validity. Conway and colleagues reproduced these findings and also observed a profound effect of interviewer training on interview reliability and validity (Conway, Jako & Goodman 1995). In addition, basing questions on an analysis of requirements of the job on offer was found to improve reliability and validity through an indirect mechanism (Conway, Jako & Goodman 1995).

Table 2.3 Structure of selection interviews

<table>
<thead>
<tr>
<th>Variables Coded as Potential Moderators of Interview Reliability (Extracted from Variable and Codes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design:</td>
</tr>
<tr>
<td>0 = reliability based on separate interviews.</td>
</tr>
<tr>
<td>1 = reliability based on a panel interview.</td>
</tr>
<tr>
<td>Standardization of interview questions:</td>
</tr>
<tr>
<td>1 = no formal constraints.</td>
</tr>
<tr>
<td>2 = topical areas specified, and sample questions possibly provided.</td>
</tr>
<tr>
<td>3 = pool of primary questions provided, and interviewer choice allowed.</td>
</tr>
<tr>
<td>4 = primary questions specified, and follow-up probing allowed.</td>
</tr>
<tr>
<td>5 = primary questions specified, and no follow-up probing.</td>
</tr>
<tr>
<td>Standardization of response evaluation:</td>
</tr>
<tr>
<td>1 = formation of a single overall evaluation.</td>
</tr>
<tr>
<td>2 = formation of multiple evaluations along dimensions.</td>
</tr>
<tr>
<td>3 = evaluation of applicant's response to each individual question.</td>
</tr>
<tr>
<td>Standardization of method for combining ratings:</td>
</tr>
<tr>
<td>1 = ratings combined subjectively.</td>
</tr>
<tr>
<td>2 = ratings combined mechanically (e.g., summed or averaged).</td>
</tr>
<tr>
<td>Questions based on a job analysis:</td>
</tr>
<tr>
<td>0 = no.</td>
</tr>
<tr>
<td>1 = yes.</td>
</tr>
<tr>
<td>Interviewer training:</td>
</tr>
<tr>
<td>0 = no.</td>
</tr>
<tr>
<td>1 = yes.</td>
</tr>
</tbody>
</table>
Selection interviews for medical school may be conveniently classified within this framework. Until recently, selection interviews have generally been unstructured affairs conducted by a single interviewer or interview panel. Interviewers were commonly untrained and questions were, at best, only loosely related to the qualities required of medical students and doctors. Interview questions, response evaluations and methods for combining scores usually lacked a standardised approach. More recently, attempts have been made to increase the structure of selection interviews and most selection interviews attended by candidates for medical school are best described as semi-structured interviews (designed to assess pre-specified qualities but flexible in approach) or structured interviews (designed to assess pre-specified qualities using a consistent and pre-specified approach). These efforts have culminated in the development and application of the multiple mini-interview which is characterised by a degree of reliability that is comparable to that achieved by assessment practices commonly employed by medical schools.

2.3.6.1 Unstructured Interviews

Traditional, unstructured medical admissions interviews have been criticised heavily for being unreliable, lacking incremental validity, creating biases and wasting resources. (Gough 1979). Reported inter-rater reliability for unstructured selection interviews varies between 0.22 and 0.90 (Salvatori 2001). Most of the variance in interviewer ratings is likely to be due to interviewer variability, which in turn is related to interviewer experience (Harasym et al. 1996). Gough also observed interviewers to be heavily influenced by personal qualities that were not directly assessed by the interview and not themselves related to subsequent performance in medical training (Gough 1967).
Unstructured selection interviews in particular, appear to have particularly limited validity as a selection tool and are less useful than standardized test scores or premedical GPA in predicting future performance (Gough 1979). Correlations between performance at interview and subsequent performance are low due to factors such as poor test reliability, due to interviewer and interview variability, and failure to adequately identify and measure non-cognitive qualities that are required by good medical students and doctors, and are likely to add no more than 8% to prediction of subsequent performance (Goho & Blackman 2006; Gough 1979; Powis & Rolfe 1998).

Although some authors have reported a positive correlation between performance on unstructured selection interviews and academic and clinical performance during the medical course, a more common finding has been that the interview is not predictive of success (Salvatori 2001). A recent meta-analysis concluded that selection interviews for the health care professions (including medicine) have little value for predicting academic and clinical performance varies widely (r=0.17, range 0 to 0.65) (Goho & Blackman 2006). Kulatunga and colleagues found that McMaster panel interviews that assessed motivation for medicine, breadth of experience, and interpersonal skills had low predictive ability for the communication skills component of the Canadian licensing examination (r=0.24) but not for overall performance (r=0.01) (Kulatunga-Moruzi & Norman 2002). The inter-rater reliability for the overall interview score was 0.66. (Powis et al. 1988). DeVaul and colleagues (1987) found no meaningful differences in performance as an intern between medical students performing well or poorly on initial selection interviews (DeVaul et al. 1987).
2.3.6.2 Structured Interviews

Structured and semi-structured interviews are currently integrated into the selection processes of most medical schools. It is generally agreed that the use of a structured interview format and interviewer training improves reliability, but not necessarily validity (Salvatori 2001; Smith 1991). However, Axelson and colleagues have recently contended this view (Axelson et al. 2010). Improvements in test reliability are likely to raise the potential predictive validity of a test, however an increase in predictive validity also requires the identification and appropriate measurement of appropriate personal qualities. Indeed, it has become increasingly clear that training of assessors and explicit rating guidelines enhance test reliability and validity (Salvatori 2001). Consequently, reliabilities and validities are consistently higher for structured interviews than for unstructured interviews (Conway, Jako & Goodman 1995; Huffcutt & Arthur 1994). Indeed, correlation between interview performance and subsequent academic and on-the-job performance for structured interviews is commonly in the order of 0.5-0.6 and approaches that of standardised written tests (Huffcutt & Arthur 1994; Weisner & Cronshaw 1988). The correlations between unstructured interview ratings and subsequent performance are generally between 0.2 and 0.3 (Huffcutt & Arthur 1994; Weisner & Cronshaw 1988). Marchese and Muchinsky observed interview structuredness to correlate \((r=0.45)\) with interview validity (Marchese & Muchinsky 1993). However, it is likely that a ceiling effect exists, whereby increases in interview structure above a particular level have little effect on interview validity (Huffcutt & Arthur 1994).

Conway and colleagues conducted a meta-analysis of studies assessing interviews for selection into various positions. Weighted mean reliabilities (and maximum
achievable validities) were 0.67 for high structure, 0.56 for moderate structure, and 0.34 for low structure interviews (Conway, Jako & Goodman 1995). Given that these values were only slightly greater than validities reported by other recent meta-analyses (Huffcutt & Arthur 1994; Huffcutt & Woehr 1993; McDaniel et al. 1994), the authors concluded that the reason for low validities of unstructured interviews was not the criteria used but low reliability and that the best way to improve validity was most likely to be by increasing structure.

There is, however, little evidence to suggest that semi-structured interviews predict performance during the medical course. Performance on semi-structured panel interviews at a UK medical school designed to assess a range of similar non-cognitive domains (such as research into undergraduate course and career, non-academic achievements, teamwork and organisational abilities) did not correlate with first year exam performance (Lynch et al. 2009). Groves examined the relationship between candidate performance on a semi-structured panel interview designed to assess communication skills, decision-making ability, teamwork, motivation, and personal attributes such as empathy and self-awareness, and performance during the medical course at the University of Sydney and University of Queensland (Groves, Gordon & Ryan 2007). All interviewers had received formal training. Performance at interview did not correlate with performance on year 2 examinations at either medical school. However, Wilkinson and colleagues observed performance on the University of Queensland selection interviews to correlate with performance on clinical and ethics assessments late (but not early) in the medical course (Wilkinson et al. 2008). This suggests that the relevance of interview performance may increase over time.
2.3.6.3 The Multiple Mini-Interview (MMI)

The multiple mini-interview (MMI) is a multiple-station interview process that addresses some limitations of traditional selection interviews, such as limited content specificity, poor reliability and inadequate targeting of desirable personal qualities (Albanese et al. 2003). The MMI permits sampling across multiple stations designed to measure a range of desirable cognitive and non-cognitive skills and personal attributes resulting in improved objectivity and reliability (Eva, Reiter, et al. 2004b; Lemay et al. 2007). All selection interviews are resource intensive but the MMI has been shown to be cost-effective compared to traditional interviews (Eva, Rosenfeld, et al. 2004).

The MMI represents an approach to selection interviews that is similar in principle to the use of the objective structured clinical examination (OSCE) for assessing clinical performance (Eva, Rosenfeld, et al. 2004). Like, OSCE, MMI are able to attain a high degree of reliability. This is achieved because of the standardisation of questions, response evaluation, and method for combining ratings (Conway, Jako & Goodman 1995). Reliability coefficients between 0.65 and 0.8 have commonly been reported for MMI currently in use at various medical schools (Eva et al. 2009; Eva, Reiter, et al. 2004b; Eva, Rosenfeld, et al. 2004). However, it has been shown that the reliability of the MMI may be influenced by several factors, such as the number of interviewers and the number of stations employed. In general, increasing either of these factors will increase test reliability. However, optimization of these parameters so as to maximise reliability may result in unwieldy and impractical processes that outstrip available resources and produce participant fatigue. Because
of this, a balance must be struck between ideology and practicality so as to create interviews that are sufficiently reliable, but also practical.

As for OSCEs, the number of stations employed by the MMI is a key factor in determining MMI reliability. OSCE comprising as few as 4 and as many as 35 stations have been documented (Davis 2003; Hodges et al. 2002). Medical schools have generally adopted eight to ten station MMIs, based on evidence that this number of stations ensures acceptable (above 0.6) reliability (Axelson & Kreiter 2009; Eva, Reiter, et al. 2004b).

Most MMI employ stations of between five and ten minutes, however the effect of MMI station duration on MMI reliability has not been investigated. OSCE stations are typically 5 to 15 minutes in length, although stations ranging from 4 minutes to over an hour have been reported (Hodges et al. 2002). Station duration appears to have little effect on student performance at a variety of structured tasks (Schoonheim-Klein et al. 2007). Determining optimal interview duration is important to ensure meaningful, reliable results are achieved in an efficient and timely manner. Stations must be long enough to permit precise comparison for ranking purposes (Roberts et al. 2008). However, longer interviews may be associated with reduced decision quality and participant fatigue (Campion, Palmer & Campion 1997).

**Interviewers**

For panel interviews, reliability is improved by increasing the number of interviews to 2 or 3 and by ensuring independent scoring of interviewers (Mitchell, Mitchell &
Macgregor 1987; Shepard 1980). The impact of increasing number of interviewers per station on MMI reliability is less dramatic that the effect of increasing number of stations. Similar reliability (in the order of 0.75) can be achieved by using an MMI design that incorporates 9 stations, each manned by one interviewer, or by a design that incorporates 6 stations each manned by 3 interviewers (Axelson & Kreiter 2009; Eva, Reiter, et al. 2004b). In order to ensure efficient use of available resources, most medical schools employ a design that resembles the former. For the MMI, independent interviewer scoring is ensured and increasing the number of interviewers per mini-interview station has limited effect on MMI reliability.

Interviewers commonly engaged to interview potential medical student are derived primarily from three stakeholder groups: academics of the University conducting the interviews, relevant health care professionals, and members of the community. As potential peers, current medical students have also been included as interviewers at a number of institutions in Australia, Canada, the United States, the United Kingdom and Singapore. (Fruen 1983; Fulton 1979; Harris & Owen 2007; Parry et al. 2006; Tambyah 2005). Medical students have been seen to be valuable interviewers given their own recent experience in attending selection interviews and their familiarity with the expectations of the medical course (Roby 2008). In panel interviews, medical students were found to score interviews lower than other interviewer groups and were observed to be less decisive and less discriminating in their scoring than experienced interviewers (Koc, Katona & Rees 2008) However, the performance (reliability) of particular interviewer subgroups (academics, health care professionals, community members and medical students) in assessing MMI stations has not been assessed.
Scoring

Global rating scales or checklists have been used extensively to assess performance in OSCE, and may also be used to assess candidate performance at mini-interview stations. Checklists are dichotomously scored and criteria are either addressed or not whereas global scales provide measures of the general approach a candidate takes towards a particular problem or issue (Reznick et al. 1998). Scores awarded by checklists and global ratings tend to correlate reasonably well (Troncon 2006; Wilkinson & Fontaine 2002). However it has been proposed that checklists reward thoroughness rather than competence and efficiency and, as such, are more suited to assessing novices rather than experts (Hodges et al. 1999; Norman et al. 1985). Global rating scales have acceptable reliability and validity when used to assess patient-based encounters (Cohen et al. 1991; Reznick et al. 1998). For qualities commonly assessed during MMI such as communication skills and empathy, global rating scales have been found to have better construct validity and reliability than checklists (Hodges et al. 1999).

Early data suggests that performance on the MMI may predict performance during the medical course and on licensing examinations (Eva et al. 2009; Eva et al. 2004a; Reiter et al. 2007). However, these data are derived from a MMI employed by a single medical school and it is plausible that differences between MMIs employed at different medical schools may limit the generalisability of these results. Furthermore, the data available provide little information about the construct validity of the MMI as there have been no attempts to correlate performance at specific MMI stations with performance on subsequent assessable tasks that target the same qualities assessed during the MMI.
2.4 Selection methods: identifying and addressing limitations

Despite ongoing efforts to improve selection methods, the ability of university selection processes to provide useful and reliable information has often been questioned (Fulton 1979). Indeed, not all candidates admitted into medical school are “good medical students” and not all those that graduate would be considered by their patients to be ‘good doctors’. It would seem then that words voiced well over half a century ago, remain true today and that “no perfect subjective or objective method of estimating success has yet been inaugurated” (Douglass 1942).

There remains a key dilemma, selection involves a choice based on current preference or perceived value, whereas prediction involves a forecast, usually about future value (McGaghie 1990). It is not surprising that selection processes do not predict with 100% certainty future performance. No single test could be considered truly comprehensive in assessing all cognitive and non-cognitive qualities over all possible situations experienced in medical training or medical practice. Furthermore, selection tests cannot account for the effects of future training, events or experiences on decisions and behaviours.

Measuring the success of selection methods is also difficult. Performance during the medical course can be measured against the objectives of the course using standard measures such as examination performance or GPA. Assessor consistency can also be addressed and maximised. Evaluating performance beyond medical school is far more difficult and it is unlikely that any single model of a ‘good doctor’ could ever be agreed upon by all stakeholders. No assessment tool exists that permits a comprehensive and objective appraisal of physician performance.
Indeed, two reviews of measures used to assess physician performance have remarked on the marked heterogeneity of methods used and facets assessed (Hamdy et al. 2006; Wingard & Williamson 1973). Thus, it is not surprising that no single selection test has been shown to accurately predict future global performance as a medical student or as a doctor.

Douglass has suggested that the most valid method for finding out whether or not a candidate can achieve certain academic results or be successful in an occupation is for him or her to try it (Douglass 1942). Indeed many selection tools attempt to assess candidates in light of situations that are representative of those experienced by medical students or doctors. Past academic record provides a measure of academic ability and performance under exam conditions. Admission tests such as the MCAT, GAMSAT or UMAT assess problem solving ability and other analytical qualities commonly employed in medical school and beyond. Structured interviews enable trained interviewers to assess the responses of candidates to hypothetical scenarios that require forms of ethical and situational-based reasoning that are relevant to clinical decision making.

Gough advocated for selection processes that involved a variety of heterogeneous tests in order to select into medicine “a heterogeneous mixture of all the kinds of excellence medicine encompasses” (Gough 1979). Rather than hope for a single magic ‘bullet’ that accurately measures all qualities expected of a doctor, it would appear wiser to utilise a combination of selection methods that provide a multidimensional picture of each candidate.
2.5 Models for the Selection of Medical Students

A variety of tools have been used for the selection of applicants into medical training. Prior academic achievement and measures of aptitude appear to be relatively good predictors of early medical student examination performance, but have limited utility in predicting licensing examination and clinical performance. Conversely, non-cognitive variables appear to become more predictive as training progresses (Harding & Wilson 2008).

These observations must be considered within the contextual limitations that accompany the use of selection tools to assess cognitive and non-cognitive attributes considered desirable of doctors. These limitations include the following:

- the available selection tools assess proxy measures (desirable qualities and attributes) rather than performance in real life clinical settings
- medical courses all differ in course structure, content and assessment processes, these factors may limit the generalisability of results between medical schools
- the construct validity of the selection tool may be unclear (ie how well it measures what it intends to measure)
- both selection tools and subsequent assessments are not perfectly reliable which may attenuate correlations between the two. Although this may be corrected for, it does not address the unreliability of the test
- cohorts accepted into medicine generally represent high achievers and therefore data relating to performance within and beyond the medical course is only available for a restricted subset of the total relevant population. This
may further attenuate correlation between performance during and after selection

- correlation coefficients reported by different authors have been variably corrected for unreliability and/or restriction of range and therefore may not be readily comparable

Notwithstanding these limitations, incorporating multiple, complementary tools during the selection process improves the predictive validity of the process (Ferguson, James & Madeley 2002). Indeed, the combination of academic variables along with aptitude type factors noticeably improves predictions of performance (Basco et al. 2000; McGaghie 2002; Touron 1987). Roessler and colleagues observed the use of a combination of cognitive and non-cognitive predictors \( r=0.45-0.64 \) to be superior to cognitive predictors alone \( r=0.26-0.28 \) in predicting performance during medical school and on licensing examinations (Roessler et al. 1978). Non-cognitive variables were observed to account for at least as much variance in performance as cognitive predictors (Roessler et al. 1978). Use of the USMLE in addition to these two factors appears to add little to the predictive power of the combination (Julian 2005).

Several authors have proposed a model for selection based entirely on non-cognitive factors, provided an academic threshold has been achieved (Edwards, Elam & Wagoner 2001). Albanese and colleagues examined the effects of applying various academic thresholds for selection eligibility at the University of Wisconsin (Albanese, Farrell & Dotti 2005). Based on potential predictive ability, an MCAT
score of 24 and a science GPA of 3.00 were found to provide reasonable threshold levels for making admission decisions at that institution.

Kreiter considered the implications of utilising a model that involves using MCAT/GPA thresholds as eligibility criteria then selecting purely on non-cognitive criteria (Kreiter 2007). Unfortunately, the academic performance consequences of relying solely on non-cognitive factors for selecting applicants above the threshold have not been fully considered in the literature (Kreiter 2007). Kreiter concluded that there is no sound rationale for adopting a threshold approach that accepts lower marks.

Bore and colleagues (Bore, Munro & Powis 2009) recently proposed a comprehensive model for selection that included informed self-selection, academic achievement, general cognitive ability and aspects of personality and interpersonal skills. The model considers both cognitive and non-cognitive factors which is likely to maximize the probability of making accurate, fair and defensible selection decisions. It is unlikely that selection process can rely overly heavily on either cognitive or non-cognitive variables. High GPAs and aptitude test scores are not inextricably linked to exceptional academic performance during medical school nor do they guarantee the provision of particularly high quality health care as a practitioner. Similarly, non-cognitive measures are limited in their ability to predict performance during medical school, but their relevance increases as individuals move into the clinical realm. Indeed, non-cognitive qualities are better than measures of academic success and aptitude in providing information relating to
key facets of professionalism such as communication skills, compassion, integrity and commitment (Adam et al. 2012; Wagoner 2006).

It is reasonable to utilise a battery of selection tools that assess the cognitive and non-cognitive attributes. By combining selection methods, the limitations of each individual testing methods is addressed and the predictability of overall performance is increased. However, the benefits, limitations and costs of each test must be understood to maximise the fairness, efficiency, reliability and validity of the selection process. GPA and standardised aptitude tests are useful in ensuring that candidates chosen have the academic ability to cope with the demands of the medical course and to make clinical decisions that are informed and evidence-based. The usefulness of current methods that assess non-cognitive attributes is less certain. It is clear however, that professional and ethical clinical practice centres around a number of key non-cognitive characteristics that represent prime targets for assessment during the selection process. Current methods are increasingly rigorous but room for improvement clearly remains in relation to the identification and assessment of desirable qualities and the ability of selection tests to detect dysfunctional tendencies (Knights & Kennedy 2007). Methods for accurately assessing these characteristics are still being developed. The MMI is one such method that allows a targeted and objective assessment of desirable non-cognitive characteristics. However, further research is required to assess the influence of factors such as interview duration, interviewer type and interviewer training on the efficiency, reliability and validity of the MMI.
References


Douglass, H 1942, *The relation of vocational success and certain factors of academic success at the University of Minnesota*, University of Minnesota, Minneapolis.


Fulton, T 1979, 'The address to the students', paper presented to Opening of the Winter Session Royal Victoria Hospital, London, 4 October.


Harrington, G 1997, 'Psychological testing, IQ, and evolutionary fitness', *Genetica*, vol. 99, pp. 113-23.

Harris, S & Owen, C 2007, 'Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School', *Medical Education*, vol. 41, pp. 234-41.


Hodges, B, Regehr, G, McNaughton, N, Tiberius, R & Hanson, M 1999, 'OSCE checklists do not capture increasing levels of expertise', *Academic Medicine*, vol. 74, no. 10, pp. 1129-34.


Hughes, P 2002, 'Can we improve on how we select medical students?', *Journal of the Royal Society of Medicine*, vol. 95, pp. 18-22.

Jackson, D & Dawson-Saunders, B 1987, 'History of not completing courses as a predictor of academic difficulty among first year students', *Journal of Medical Education*, vol. 62, pp. 880-5.


Kreiter, C 2007, 'A commentary on the use of cut-scores to increase the emphasis of non-cognitive variables in medical school admissions', *Advances in Health Sciences Education*, vol. 12, pp. 315-9.

Kreiter, C & 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*, vol. 19, no. 2, pp. 95-100.


McGlinn, S & Jackson, E 1989, 'Predicting the medical school progress of minority students who participated in a preparatory program', *Academic Medicine*, vol. 64, no. 3, pp. 164-6.


Mercer, A, Chiavaroli, N 2007, ‘UMAT: a validity study. A review of the underlying constructs and an analysis of the content of the Undergraduate Medicine and Health Sciences Admission Test’, ACER & The University of Western Australia.


Morgeson, FP, Campion, MA, Diproye, RL, Hollenbeck, JR, Murphy, K & Schmitt, N 2007a, 'Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection', *Personnel Psychology*, vol. 60, pp. 1029-49.

Morgeson, FP, Campion, MA, Diproye, RL, Hollenbeck, JR, Murphy, K & Schmitt, N 2007b, 'Reconsidering the use of personality tests in personnel selection contexts', *Personnel Psychology*, vol. 60, pp. 685-729.


Mullen, F 1948, 'Selection of medical students', *Journal of the Association of American Medical Colleges*, vol. 23, p. 163.


Ogunyemi, D & Taylor-Harris, D 2005, 'Factors that correlate with the U.S. Medical Licensure Examination Step-2 scores in a diverse medical student population', *Journal of the National Medical Association*, vol. 97, no. 9, pp. 1258-62.


Reiter, HL, Eva, KW, Rosenfeld, J & Norman, GR 2007, 'Multiple mini-interviews predict clerkship and licensing examination performance', *Medical Education*, vol. 41, pp. 378-84.


Salvatori, P 2001, 'Reliability and validity of admissions tools used to select students for the health professions', *Advances in Health Sciences Education*, vol. 6, pp. 159-75.


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*, vol. 80, no. 10 Suppl pp s14-6


Yates, J & James, D 2006, 'Predicting the ‘strugglers’: A case- control study of students at Nottingham University Medical School', *British Medical Journal*, vol. 332, pp. 1009-13.
3. Attributes sought in candidates for medical school

The selection processes of medical schools attempt to select, from the available pool of applicants, those most likely to make good medical students and, more importantly, good doctors (Lowry 1992; Smyth 1946). The importance of selecting candidates with particular desirable intrinsic qualities has long been recognised. A natural disposition for medicine was required in ancient Greece and Rome (Calman 2007). Indeed, Galen considered that not all candidates were suitable for medical training and the training of one unsuited to the profession would not yield the desired results (Levey 1967, pp18-94). Those suited to medicine were considered by Al-Ruhawi “as a good wine that was only fit to be preserved in a vessel that would preserve the taste and purity of its colour, the goodness of its odour, and, in short, its other good qualities” (Al-Ruhawi 1967). Bauer recognized that there were core personal qualities that lent themselves to “goodness” in medical school and beyond. He felt that good students would often become good physicians regardless of the instruction they receive, and that those who struggled as students would make poor doctors (Bauer 1956).

However, despite the early recognition of the importance of personal attributes in determining suitability for a career in medicine, it was only 20 years ago that McGaghie declared:

_The purpose [of selection procedures] is clear: to decide who will and who will not be able to complete successfully a medical school program and become a good physician... It is the identification of the qualities that contribute to this success, that is missing_ (McGaghie 1990).
3.1 Desirable qualities and professional values

Students entering medical training have been selected according to characteristics or qualities likely to lead to successful completion of training and those expected of “good” practising doctors. However, the identification of core qualities that are required by successful medical students and good doctors is a challenging exercise. The *Huang di nei jing* suggested that medical teachers should look for candidates with personal characteristics akin to their own to ensure that the integrity of the profession would be maintained. In Ancient Greece, desirable characteristics of doctors included a healthy complexion, a well groomed appearance and a serious demeanor (Calman 2007). It was important that doctors were dedicated to their patients and put public good above their own wealth. Contracts drawn up between medical teachers and their apprentices, such as the Hippocratic Oath, provide further insights into those qualities considered desirable of medical students and practising physicians. The Hippocratic Oath espouses beneficence and non-maleficence and requires students of medicine to demonstrate excellence, collegiality, confidentiality and accountability. Students and practitioners are also expected to be respectful and to behave in a moral and reputable way (Staden 1996). Galen highlighted the requirement for a virtuous ethic, innate ability and for doctors to be motivated and dedicated to continuing education (Calman 2007). In Islamic cultures attempts were made to extend the phenotype of a good doctor so as to include physical characteristics (Ibn Hubal 1943-4). The *Letter to Arsenius*, composed during the latter part of the first millennium outlines the qualities expected of physicians during the Middle Ages (MacKinney 1952, p8):
First, [he] should test his personality to see that he is of a gracious and
innately good character, apt and inclined to learn, sober and modest,...
charming, conscientious, intelligent, vigilant and affable, in all details adept
and skilful... amiable, humble and benevolent ... not timid, turbulent or proud,
scornful or lascivious, or garrulous, a publican, or a woman-lover ... not
drunken or lewd, fraudulent, vulgar, criminal or disgraceful ... [He] should
not have faults, but instead discretion, taciturnity, patience, tranquility, and
refinement.

However, for almost the next one thousand years, candidates were admitted into
medical training largely on the basis of financial grounds and evidence of academic
achievement. Where consideration was given to the non-cognitive qualities
expected of doctors, it generally occurred upon entry into or during medical
practice, rather than entry into medical training. This is evident in the following
extract from an early manual for graduating medical students that stipulates that
graduates should be (Durling 1970, p23):

of good character and good memory, well formed, well behaved, daring in
diseases where nothing is to be feared, circumspect in dangerous cases, let
him flee severe diseases, be gracious to the sick, peaceable with his
colleagues, cautious in prognosis, chaste, sober, pious, compassionate, not
grasping or extortionate.

It was not until the first half of the twentieth century that attention was drawn back
to non-cognitive attributes and the development of tools designed to select for
candidates with desirable qualities and aptitude for a career in medicine. At the First World Conference on Medical Education, Johnson appealed to medical educators to consider not only intelligence, industry and aptitude for medicine, but also emotional stability, kindness and motivation (Johnson 1953). He did, however, acknowledge that the latter qualities were not readily measurable. The American Psychiatric Association highlighted the lack of data linking specific personal qualities with success as a doctor, stating “we do not know what special qualities make a good medical student or a proficient physician” (Bloomgarden 1957). Some 30 years later, Fruen reiterated these points, suggesting that the development of robust selection practises would require both the identification of the special qualities required of successful doctors qualities, and the development of tools that could adequately describe and measure them (Fruen 1983).

Since then, numerous attempts have been made to identify core qualities that contribute to the make-up of a “good doctor”. Academic aptitude as demonstrated by previous academic achievement has been the most consistently sought attribute and has, at times, been the only criterion used to select candidates into medical training. The need for medical students to possess related cognitive qualities such as decision-making ability, problem solving ability, and language proficiency has also become widely accepted. These qualities are now routinely assessed via standardised aptitude tests designed specifically to assess aptitude for the study of medicine. There has also often been an expectation that candidates have attained a requisite level of proficiency in particular disciplines integral to the study of medicine, such as biology, chemistry, physics and mathematics (Whitby 1956). Some authors have questioned these practises and suggest that “the nature of a
candidate’s premedical training is not important provided he has been truly educated and has the ability to think for himself” (sic) (Anonymous 1954). However, some medical schools still require candidates to have completed specific subjects to be eligible for selection, while others award bonuses to candidates who have completed desirable subjects.

Whitby has suggested that, in addition to cognitive ability, good doctors required a number of desirable non-cognitive qualities (Whitby 1956). Bauer highlighted the need to consider cognitive ability, personality and motivations and advocated the selection of candidates with “strong vigorous minds, stable personalities, and warm interest in the welfare of others” (Bauer 1956). Smyth drew attention to the diversity of qualities that may be needed in medicine (Smyth 1946). The list included; intellectual capacity, integrity, ability for hard work, conscientiousness, sympathy, tact, ability to deal with people, organizing ability, manual dexterity, cheerfulness, resourcefulness, decision-making ability, self-confidence, patience, enthusiasm, interest, endurance and common sense.

In 1964, Price and colleagues identified and measured 80 separate criteria of physician performance for general practitioners, specialists, and academics (Price et al. 1964). Despite the number of qualities proposed, the list was considered inadequate because it did not consider the dimension of patient care (Price et al. 1971)

Several years later, Price and colleagues extended the list to include 87 qualities considered desirable in medical practitioners (Price et al. 1971). The qualities were
ranked in order of importance by a diverse group of stakeholders that included medical practitioners, non-medical professionals, clergyman and recently discharged hospital patients. The full list of desirable qualities in order of importance were:

1. Has good clinical judgment (the ability to reach appropriate decisions regarding the care of patients)
2. Has thorough up-to-date knowledge of his own field of medicine
3. Has knowledge and ability to study patients thoroughly, and reach sound conclusions regarding diagnosis, treatment, and related problems
4. Readily refers patients when it is to their advantage to do so
5. Habitually makes as thorough an examination of each patient as may be required for accurate diagnosis and proper treatment
6. Is wise, thoughtful; is able to get at the heart of a problem; is able to separate important points from details
7. Is strict about honoring confidences; avoids and discourages gossip
8. Is adaptable; is able to adjust to new knowledge and changing conditions
9. Provides treatment appropriate to the condition of each of his patients, with (in general) satisfactory immediate and long-range results
10. Is able to convert acquired information into working knowledge
11. Inspires confidence in his patients
12. Has intellectual honesty (incompatible with bluffing, cheating, assuming poses for ulterior purposes, trickery, claiming undue credit, assuming knowledge not really possessed, transferring blame unfairly, etc.) and forthrightness
13. Keeps completely honest records
14. Is alert, observant
15. Is able to be his own teacher; to learn from books and journals, from meetings and informal discussions, from experience and his own mistakes, etc., thus adding continually to his own education
16. Keeps full and accurate clinical records
17. Is emotionally stable
18. Has sustained genuine concern for patients during their illness and convalescence
19. Has awareness of emotional and psychosomatic factors in dealing with patients and their diseases
20. Is decisive; is able without undue delay to reach conclusions and act upon them
21. Is a stable, calming influence in critical or stormy situations
22. Is conscientious; strives for perfection in his work
23. Is equipped with an orderly mind; mentally efficient; logical
24. Is willing to take needed time to listen to patients' problems sympathetically and helpfully
25. Establishes good doctor-patient relationships
26. Is able to communicate well in everyday work and relationships (with patients, relatives, assistants, students, colleagues, the public, etc.)
27. Is studious; attends appropriate medical meetings or refresher courses; keeps abreast of progress in medical knowledge and practice, especially in his own field
28. Is capable of independent thinking; able to reach his own conclusions
29. Has foresight, the ability to anticipate problems
30. Works effectively with patients in a doctor-patient team approach to combat illness and to promote better health
31. Is able and willing to learn from others (colleagues, nurses, students, etc.)
32. Demonstrates irreplaceable behavior in regard to his handling of patients
33. Is considerate of others; is alert to patients' convenience and comfort; courteous, tactful
34. Is available when needed, even at the cost of personal convenience
35. Is an understanding sort of person
36. Is skilled in handling disconsolate, antagonistic, or emotionally upset people
37. Is of unquestionable integrity, high-principled (so that low, mean, dishonest, immoral, uncharitable, selfish courses of action seem foreign to his nature)
38. Patients and their relatives are generally satisfied by the care provided by the physician
39. Is unusually intelligent; mentally quick; bright, keen
40. Has a well-equipped office, with needed diagnostic facilities present or available
41. Works effectively with patient's family in giving complete medical care
42. Sees his staff (nurses, technicians, and clerical people) as teammates in giving the best health services to the public
43. Is adaptable to all sorts of people; acceptable to them and at ease with them, irrespective of their economic status, social standing, race, amount of education, degree of culture, etc.
44. Is frank and open; takes patients into his confidence
45. Strives to educate and inform patients about treatment and good health practices; follows through by checking with patients later about the effectiveness of the information he offered
46. Strives to educate and inform patients about treatment and good health practices; follows through by checking with patients later about the effectiveness of the information he offered
47. In charging specific fees, he is sensitive to possible economic difficulties of his patients
The list is extensive, including a variety of cognitive and non-cognitive qualities. However, some authors have suggested that candidates do not require a full set of ideal personality traits, provided critical qualities are present. A sense of vocation and motivations based on professional achievement rather than financial gain or social security were identified early as key qualities for consideration (Anonymous 1954). Weingartner suggested that in addition to the essential cognitive traits of
intelligence and imagination, it was important that those entering medical training had good judgment and good character (Weingartner 1980). In 1990, McGaghie nominated 10 key characteristics sought in candidates for medical school entry; character and integrity, knowledge, leadership, geographic preferences, gender, race and religious beliefs, work habits and motivation for study, personality and attitude, orientation towards service, altruism and personal effectiveness (McGaghie 1990). The usefulness of lists such as this is questionable however as many of the characteristics put forward are themselves vague and subjective terms that are not easily agreed upon or measured.

Little consensus has been reached regarding the characteristics medical schools should seek in candidates and the literature provides little guidance regarding the best ways to measure these qualities (Parry et al. 2006). Disagreement about which personal qualities should be evaluated, therefore, remains a major issue (McGaghie 2002). A host of non-cognitive qualities considered desirable for those embarking on a career in medicine have been nominated. Similarly, there is no shortage of terms associated with “good medical practice” and “professionalism”, but these vary in their underlying conceptual and philosophical bases and much disagreement has been reported regarding the qualities intrinsic to professionalism.

The diversity of qualities considered desirable in doctors and medical students is unsurprising given the nature of medical practice and the varying expectations of patients, colleagues and other key stakeholders. Furthermore, the definition of medical professionalism tends to change with time and the attributes of a doctor may differ between countries with distinct medical histories (Kang et al. 2004). In
recent years, medical professional organizations such as the Accreditation Council for Graduate Medical Education, the American Board of Medical Specialties, and the Royal College of Physicians and Surgeons of Canada have emphasized the multiplicity of physician roles such as medical expert, collaborator, manager, health advocate, scholar, professional, and communicator (Donnon 2007). Lowry has questioned whether an essential common core of knowledge, attitudes and skills can be defined and has suggested that “a profession like medicine may contain niches for all regardless of their particular interests, skills, and weaknesses” (Lowry 1992). Smyth suggested that the qualities considered desirable may further depend upon the branch or specialty area of medicine pursued (Smyth 1946).

Differences in the importance that different stakeholder groups place upon different qualities further complicate the issue. Wagner and colleagues identified differences in the way that different stakeholder groups perceived the relative importance of professional qualities with patients and students placing more emphasis upon the establishment of relationships in contrast to faculty and residents who placed more emphasis upon skills and knowledge (Wagner et al. 2007). Green and her associates further explored the relative importance of different personal qualities to different stakeholder groups through focus groups and surveys (Green, Zick & Makoul 2009). Focus groups consisting of patients, nurses and physicians were used to identify 68 key items (qualities or behaviours), which then formed the basis of surveys sent out nationally to members of each stakeholder group, asking them to rate the importance of each item. The authors noted considerable consistency in terms of how different stakeholder groups viewed the importance of the items and 53 of the 68 items were deemed very important signs of professionalism by at least
75% of respondents to the patient, nurse, and/or physician surveys. Of note was that ethical practice and honesty were rated most highly by all groups. Some differences in the way health professionals and patients viewed a number of items were apparent however. A higher proportion of patients than doctors and/or nurses considered certain patient-centred items as very important, a higher proportion of nurses rated highly those items relating to patient advocacy and respect, while a higher proportion of physicians considered items relating to accountability and commitment to life-long learning to be very important.

In a survey of consumers in Dunedin, Hutchinson and Reid investigated the relative importance that patients and consumers place upon different personal qualities and professional values (Hutchinson & Reid 2011). The authors noted significant variability in the level of importance placed by participants on each of 18 nominated professional qualities. Qualities concerning social justice and appearance were rated as significantly less important than patient autonomy and patient welfare. The professional qualities that were most commonly ranked in the top five by respondents related to honesty, respect, competence and the ability to listen.

3.2 Classifying desirable qualities and professional values

More recently, authors have tried to identify key professional domains or themes that incorporate qualities considered desirable for those embarking on a career in medicine. Wagner and colleagues developed a model of professional values using focus groups that comprised of academic faculty, residents, medical students and patients (Wagner et al. 2007). The model included three primary and 3 secondary themes based on common values identified by different focus groups. Primary
themes were described as knowledge/technical skills, patient relationship and character virtues, whilst secondary themes included medicine as a unique profession, congruency between personal characteristics and outward appearance and behaviour and the importance of peer relationships.

Hur and Kim used a Delphi survey to identify core elements of medical professionalism that would develop into professional competence that should be sought in medical school applicants (Hur & Kim 2009). One hundred-six responses from medical school professors and 230 completed questionnaires from medical students produced 1,580 elements that were then reclassified into 3 domains, ‘professional knowledge’, ‘professional skills’, and ‘professional attitude’ containing eight subordinate categories and a total of 27 core elements (Table 3.1).

Table 3.1 Core elements of medical professionalism (extracted from Hur & Kim 2009)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Category</th>
<th>Core elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional knowledge</td>
<td>Understanding basic science</td>
<td>Knowledge of basic science</td>
</tr>
<tr>
<td></td>
<td>Understanding human-socijal science</td>
<td>Basic knowledge of human-social science</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Understanding the characteristics of Korean society</td>
</tr>
<tr>
<td>Professional skills</td>
<td>Self management skills</td>
<td>Self-restraint &amp; risk management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Planning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Physical &amp; mental health</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Life-long learning skills</td>
</tr>
<tr>
<td></td>
<td>Human relationships</td>
<td>Teamwork</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leadership</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communication skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foreign language skills</td>
</tr>
<tr>
<td></td>
<td>Multiple thinking skills</td>
<td>Logical &amp; critical thinking skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem solving skills</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decision making skills</td>
</tr>
<tr>
<td>Professional attitudes</td>
<td>Service attitude</td>
<td>Service oriented</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Respect for others</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Humanity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Etiquette</td>
</tr>
<tr>
<td></td>
<td>Monitoring attitude</td>
<td>Ethical thinking &amp; behaviour</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self-confidence, trust, autonomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Integrity, diligent, honesty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sense of duty</td>
</tr>
<tr>
<td></td>
<td>Progressive attitude</td>
<td>Self-examination</td>
</tr>
</tbody>
</table>


Using semi-structured interviews with faculty preceptors, residents, interns, nurses and patients associated with different hospital departments, Leung, Hsu and Hui recently identified 30 desirable qualities that were then classified into three major themes; expectations of a professional doctor, work values and patient care (Table 3.2) (Leung, Hsu & Hui 2012).

Table 3.2 Desirable qualities of doctors (extracted from Leung, Hsu & Hui 2012)

<table>
<thead>
<tr>
<th>Theme</th>
<th>Qualities</th>
</tr>
</thead>
</table>
| Expectations of a Professional Doctor | Accountability  
Excellence  
Medical knowledge  
Good conduct  
Personal appearance  
Being respected |
| Work Values                   | Altruism  
Acting for patients’ best interests  
Doing no harm to patients  
Integrity  
Being a responsible person  
Being ethical in research  
Treating patients without discrimination  
Being objective  
Self-confidence  
Emotion management and empathy  
Being enthusiastic at work  
Good communication with colleagues  
Knowing self limitations  
Perseverance  
Respecting the profession  
Having a strong sense of mission  
Team spirit  
Willingness of accepting others’ opinions  
Working seriously |
| Patient Care                  | Communication with patients and patient family members  
Taking care of patients’ psycho-social well-being  
Respecting patients  
Having patience  
Having a loving heart |
Wilkinson and colleagues performed a systematic review of the literature and identified the following 5 key themes that encompassed all qualities and values reported in the literature between 1996 and 2007 (Wilkinson, Wade & Knock 2009):

1. Adherence to ethical practice principles; including qualities such as honesty, integrity, confidentiality, moral reasoning and respect of privileges and codes of conduct

2. Effective interactions with patients and with people who are important to those patients; including respect for diversity, politeness and courtesy, patience, empathy, manners, inclusion of patients in decision-making and the maintenance of professional boundaries

3. Effective interactions with other people working within the health system; including teamwork, respect for diversity, politeness and courtesy, manners, the maintenance of professional boundaries and balancing availability for others with care of oneself

4. Reliability; including accountability, punctuality, taking responsibility and being organised

5. Commitment to autonomous maintenance and continuous improvement of competence; including reflectiveness, lifelong learning, leadership and advocacy

A literature search by the author, utilising the terms medical AND (professional OR professionalism OR non-cognitive) AND (qualities OR attributes) in PubMed identified 648 published articles of which approximately half provided information on qualities considered desirable in medical students and/or doctors. An assessment of these papers, their references and published codes of conduct developed by
professional medical organisations revealed over 400 different words or phrases to describe qualities considered to be desirable in those embarking on a medical career.

Despite the number and diversity of qualities reported desirable in doctors and doctors-to-be, common themes could be identifiable and the qualities identified appeared to be most conveniently classified into 4 major domains: values and attitudes (describing intrinsic virtues), behaviours and decorum (describing the way individuals interact with, and are viewed by, others), ethics and decision making (describing qualities that underpin decisions to do what is considered ‘right’ or ‘just’) and responsibilities (including obligations perceived to be associated with medical practice). These domains and examples of commonly reported qualities that fit within each domain are shown in Table 3.3.

Table 3.3 Domains of professionalism and associated qualities

<table>
<thead>
<tr>
<th>Values &amp; Attitudes</th>
<th>Behaviour &amp; Decorum</th>
<th>Ethics &amp; Decision making</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altruism</td>
<td>Trustworthy/maintains</td>
<td>Justice/resource use/mediation</td>
<td>Medical expert</td>
</tr>
<tr>
<td>Honourable</td>
<td>Confidentiality</td>
<td></td>
<td>Clinical/ &amp; Scientific</td>
</tr>
<tr>
<td>Honesty</td>
<td>Respectful/polite</td>
<td>Beneficence &amp; Non-maleficence</td>
<td>Competence/excellence</td>
</tr>
<tr>
<td>Caring/compassion</td>
<td>Manner/dress</td>
<td>Regulation (self/peer)</td>
<td>Interests of others</td>
</tr>
<tr>
<td>Empathy</td>
<td>Awareness, Reflective &amp; Mindfulness</td>
<td>Patient autonomy</td>
<td>Rule abiding (law/ethical codes)</td>
</tr>
<tr>
<td>Integrity</td>
<td>Collaborative skills &amp; Teamwork</td>
<td>Indiscriminate care</td>
<td>Advocacy role &amp;</td>
</tr>
<tr>
<td>Responsibility</td>
<td></td>
<td>Appropriate relationships</td>
<td>Improving access &amp; quality of care</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humility</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Therefore, a pool of qualities commonly considered to be integral to good medical practice is readily available. Approximately 30 qualities associated with these terms appear commonly in the literature and are likely to be generally accepted as being desirable in candidates for medical school entry. These qualities have been ranked highly by a variety of key stakeholders involved in medical student selection and are likely to be broadly accepted (Green, Zick & Makoul 2009; Price et al. 1971; Wagner et al. 2007). Furthermore, despite suggestions that cultural or historical factors may affect the attributes of a doctor considered desirable (Kang et al. 2004), these qualities have been consistently nominated by authors that have gathered their information from different cultural contexts.

It is not practical or possible to assess applicants for medical school for all possible qualities that could bear positively on a career in medicine. Even reducing the pool of items from several hundred candidate characteristics to a set of 30 or so broadly accepted qualities does not leave medical schools with manageable task if they were to attempt to assess all of them. To assess applicants for so many qualities would be resource intensive and exhausting for applicants and assessors alike. Furthermore, the measurement of desirable non-cognitive qualities also remains a major challenge. Many of these qualities are difficult to define or context specific and not readily assessed in applicants for medical school.
Nevertheless, a pool of key qualities considered desirable of medical students and doctors have been consistently identified by a number of different authors using a variety of techniques such as focus groups, surveys and systematic reviews of the literature. These qualities may be conveniently classified into a small number of key domains. Although differences exist in the way authors have grouped certain qualities, the classification system used may be less important than ensuring that all (or most) domains in a given model are covered. Given the limitations described above, a reasonable approach may be for medical schools to select qualities from these domains that are measurable, and particularly pertinent to the underlying philosophy and objectives of the school and to then select medical students who are likely to achieve those outcomes. This would ensure that selection processes consider essential elements of the professionalism ‘sphere’ while also serving to increase the diversity of strengths possessed by candidates.

Some medical schools have recently stated their objectives and outcomes more broadly, taking into account desirable non-cognitive qualities. In many cases, medical schools have taken an outcome-based approach to the selection of medical students - assessing applicants against core values expected of junior doctors. This is the case at Deakin University, where all candidates are assessed against 10 key outcomes of the medical course (Deakin University 2009). These outcomes include; communication skills, evidence-use, health promotion, teamwork, motivation for a career in medicine, self-directed learning, social justice, professionalism, resource use and an awareness of rural issues. These outcomes provide a multi-faceted view of applicants and can be classified according to any of the previously discussed models for classifying personal qualities. The classification of Deakin outcomes...
according to the domains nominated by 4 different authors is shown in Table 3.4. The Deakin outcomes provide information relating to all domains proposed by Leung, Hsu and Hui (2012) and my own proposed domains, and all but one (Reliability) of those proposed by Wilkinson and colleagues (2009).

Table 3.4 Deakin outcomes classified according to various proposed models of professionalism

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication skills</td>
<td>Effective interactions</td>
<td>Patient Care</td>
<td>Behaviour &amp; Decorum</td>
</tr>
<tr>
<td>Evidence use</td>
<td>Commitment to</td>
<td>Work values</td>
<td>Ethics &amp; Decision-making/</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Responsibilities</td>
</tr>
<tr>
<td>Health promotion</td>
<td>Effective interactions</td>
<td>Patient Care</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Responsibilities</td>
</tr>
<tr>
<td>Teamwork</td>
<td>Effective interactions</td>
<td>Work Values</td>
<td>Behaviour &amp; Decorum</td>
</tr>
<tr>
<td>Career motivation</td>
<td>Commitment to</td>
<td>Work Values</td>
<td>Values &amp; Attitudes</td>
</tr>
<tr>
<td></td>
<td>autonomous maintenance &amp; continuous</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>improvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-directed learning</td>
<td>Commitment to</td>
<td>Work Values</td>
<td>Responsibilities</td>
</tr>
<tr>
<td></td>
<td>autonomous maintenance &amp; continuous</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>improvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social justice</td>
<td>Adherence to ethical</td>
<td>Expectations of a</td>
<td>Ethics &amp; Decision-making</td>
</tr>
<tr>
<td></td>
<td>principles</td>
<td>professional doctor</td>
<td></td>
</tr>
<tr>
<td>Professionalism</td>
<td>Adherence to ethical</td>
<td>Expectations of a</td>
<td>Ethics &amp; Decision-making</td>
</tr>
<tr>
<td></td>
<td>principles</td>
<td>professional doctor</td>
<td></td>
</tr>
<tr>
<td>Resource use</td>
<td>Adherence to ethical</td>
<td>Expectations of a</td>
<td>Ethics &amp; Decision-making</td>
</tr>
<tr>
<td></td>
<td>principles</td>
<td>professional doctor</td>
<td></td>
</tr>
<tr>
<td>Rural awareness</td>
<td>Commitment</td>
<td>Work Values</td>
<td>Responsibilities</td>
</tr>
</tbody>
</table>
References


Hur, Y & Kim, S 2009, 'What qualities do medical school applicants need to have?', *Yonsei Medical Journal*, vol. 50, no. 3, pp. 427-36.


Johnson, V 1953, 'Requirements for entry into medical schools: the selection of students', *British Medical Journal*, vol. 2 no. 4834 p 467.

Leung, D, Hsu, E & Hui, E 2012, 'Perceptions of professional attributes in medicine: a qualitative study in Hong Kong', *Hong Kong Medical Journal*, vol. 18, no. 4, pp. 318-24.


4. Introduction to Experimental work

The previous sections provide a consolidated historical account of the evolution of processes employed to select individuals into medical training and critically appraise currently available tools used to control entry into medical school. Selection of applicants for admission into medical schools is commonly based on measures of aptitude and academic performance in conjunction with a structured interview. Previous academic achievement as a marker of academic ability has been consistently and justifiably used to provide information about the cognitive ability of candidates for medical school. A number of tools such as the GAMSAT have been specifically developed to assess a candidate’s aptitude for medicine. Desirable non-cognitive qualities have proven more difficult to identify, qualify and quantify. Interviews represent the most commonly employed selection tool for assessing the non-cognitive attributes of medical school candidates, however the ability of traditional unstructured panel interviews to provide useful information about such qualities is limited by a number of factors such as poor reliability, inconsistency and interviewer subjectivity. The MMI was developed to address these limitations and represents a promising, but insufficiently researched, selection tool. All selection interviews are resource intensive but the MMI has been shown to be cost-effective compared to traditional interviews (Eva, Rosenfeld, et al. 2004).

The MMI represents an unique form of selection interview where candidates are assessed at a series of short interview stations, each manned by an independent interviewer. Each interview station is designed to assess a specific quality or outcome considered important in the make-up of a good doctor.
By way of its inherent structure, the MMI addresses limitations of traditional selection interviews, such as limited content specificity, poor reliability and inadequate targeting of desirable personal qualities (Albanese et al. 2004). The MMI permits sampling across multiple stations designed to measure a range of desirable cognitive and non-cognitive skills and personal attributes resulting in improved objectivity and reliability (Eva, Reiter, et al. 2004; Lemay et al. 2007).

The MMI is widely employed by medical schools, including Deakin University, however various factors, such as the length of interviews, the qualities assessed and the number of stations employed, vary between medical schools. Medical schools have adopted eight to ten station MMIs, based on evidence that this number of stations ensures acceptable reliability (Eva, Reiter, et al. 2004). However, the significance of variables such as station duration, interviewer scoring characteristics and qualities assessed have not been investigated.

Currently, most MMI employ stations of between five and ten minutes. The qualities assessed at each station vary between schools but generally include a variety of cognitive and non-cognitive attributes considered important for doctors and assessment is usually undertaken by trained interviewers that belong to key stakeholder groups. How these factors are incorporated into selection MMI is likely to influence the reliability and validity of the process.
The Deakin Medical School (DMS) selection process includes an MMI comprising ten stations of eight minutes duration that address core DMS outcomes that align with moral issues such as ethical values, resource use and professional behaviours. More specifically, each station assesses one of 10 core Deakin outcomes: communication skills, professionalism, social justice, evidence-use, self-directed learning, teamwork, effective use of resources, career motivation, health promotion and rural awareness.

The content for the stations, and their format, are largely based upon MMI under license from McMaster University. Potential scenarios are considered with respect to their relevance to the 10 core Deakin outcomes and then reshaped so as to focus upon the particular quality that is to be assessed. When relevant, scenarios are also adapted to the Australian context. In cases where no suitable McMaster questions are available, scenarios are created and formatted appropriately to ensure a consistent semi-structured approach across all interview stations.

Interviewers for the MMI are derived from three key stakeholder groups: health professionals, academics and community members. All interviewers are trained and provided with the opportunity to score and then discuss pre-recorded interviews prior to taking part in the MMI. On the day of the MMI, all interviewers attends a briefing session at which each interviewer is allocated one MMI station that is directed towards a particular quality. The interviewer receives an interviewers pack which contains a copy of the scenario to be discussed at that station, as well as background information, suggested prompting questions, a scoring guide, and a list
of the applicants that will pass through the station during a given interview round. The interviewer then assesses each applicant in turn as they cycle through his/her station.

Performance at each station is scored using a six-point interval scale where 1 = unsatisfactory, 2 = borderline, 3 = satisfactory, 4 = good and 5 = excellent. A score of zero is given to applicants whose performance raised questions about their suitability for a career in medicine. Although scoring sheets all follow the same global scale, the scoring guide for each station is tailored to the specific quality being assessed at that station. Descriptor terms for each numeric value provide interviewers with a graded way of assessing the quality of interest within the context of the particular scenario. Scores obtained at each station are summed together to produce a total interview score for each applicant that is then combined with GPA and GAMSAT score.

Subsequent sections of this thesis include original experimental work that aims to explore the reliability and validity of the Deakin Medical School interview process with a particular focus on the effects of interview duration and interviewer characteristics.

**Hypotheses**

Eight central hypotheses underlie the experimental work conducted. The first 4 hypotheses relate to reliability characteristics of the MMI, the last 4 relate to MMI validity. In selection MMI, that incorporate multiple separate interviews, test
reliability reflects both consistency in applicant performance across interviews (internal consistency) and the level of agreement between interviewers (inter-rater reliability) (Conway, Jako & Goodman 1995).

Chapter 5 considers the internal consistency of the Deakin MMI and addresses the first 2 hypotheses. The first hypothesis draws upon previous research that at least eight eight-minute MMI stations, each manned by a single interviewer, provides a selection tool of acceptable reliability and stipulates that;

1. The incorporation of 10 professionalism-related outcomes into a MMI will produce a selection tool that is of acceptable reliability

The second hypothesis relates to the duration of MMI stations. Given the need to interview substantial number of medical school applicants, the duration of MMI station bears heavily upon the resources required to run selection interviews. In addition, stations must be long enough to permit precise comparisons for ranking purposes but not so long as to cause reduced decision quality and participant fatigue (Campion, Palmer & Campion 1997; Roberts et al. 2008). Medical school MMIs commonly utilise stations of between 6 and 8 minutes duration. Consistent with this, the duration of each Deakin MMI interview station is 8 minutes. It has been suggested, however, that individuals are likely to make early judgements in social encounters that incorporate many of the non-cognitive qualities assessed in the Deakin MMI (Ambady & Rosenthal 1992; Haidt 2001). Thus shortening the duration of MMI stations is unlikely to impact on interview reliability. This gives rise to Hypothesis 2:
2. Shortening interview duration from 8 minutes to 5 minutes will have minimal impact on interview reliability and overall candidate rankings.

Chapter 6 examines the inter-rater reliability of interviewers recruited to assess candidates during selection interviews. Interviewers utilised during the Deakin MMI fall into 3 main groups; community member, health professionals and academics. Current medical students represent an additional resource and key stakeholder group that could potentially contribute to the interview process. The inter-rater reliability of specific subgroups of interviewers (community members, current medical students, clinicians and academics) is explored across stations and the effect of previous experience in assessing or being assessed at stations is considered. No data describing the performance of these interviewer subgroups is currently available.

The Deakin MMI assesses candidates’ performance at 10 stations, each of which assess a distinct quality or outcome. It is common practice to “match” interviewers to certain stations so as to maximise interviewer comfort and enhance face validity. For example, clinicians are commonly placed on stations assessing outcomes most commonly experienced in clinical practice, such as health promotion and clinical ethics. Similarly, academics are placed on stations that assess qualities such as teamwork and self-directed learning, and community members on stations that assess outcomes such as communication skills and motivation for a career in medicine. However, many of the outcomes measured represent positive qualities that fulfil broad socio-cultural norms and are therefore likely to be viewed (and scored) similarly by all key stakeholder groups. Consistency in scoring would be
further promoted by interviewer training and the provision of interview guides for
interviewers. This premise forms the basis of Chapter 6 which investigates the
following two related hypotheses, that:

3. **The IRR of a representative Deakin MMI would be relatively high**

4. **IRR will be similar for all standard interviewer subgroups (community
members, clinicians and academics), and for a novel group of interviewers
composed of current medical students.**

The final four hypotheses relate to the validity of the MMI and are investigated in
Chapter 7. Hypotheses 5, 6 and 7 stem from the notion that the MMI provides an
opportunity to assess cognitive and non-cognitive qualities commonly considered to
lie within the broad definition of professionalism, as well as attitudes towards social
responsibility, cultural awareness and communication skills. The predictive validity
of the MMI is examined in relation to its ability to provide useful information about
subsequent performance in tasks encountered throughout the medical course that
draw upon the cognitive and non-qualities targeted in the MMI. Therefore it is
hypothesized that:

5. **Performance at specific interview stations will correlate with performance at
medical school assessment tasks designed to measure the same quality (eg
communication skills)**
The curriculum throughout the Deakin Medical Course is organised into four themes: *Doctor and Patient (DP)*; *Knowledge of Health and Illness (KHI)*; *Doctors, Cultures, Peoples and Institutions (DPCI)*; and, *Ethics, Law and Professional Development (ELPD)*. In broad terms, scientific knowledge is assessed within the KHI theme, public health, social responsibility and cultural awareness, epidemiology and biostatistics within the DPCI theme, communication skills, clinical reasoning and procedural skills within the DP theme, whilst professional values were largely assessed within the ELPD theme. When these factors are considered, it follows that:

6. **Performance at the MMI will correlate best with overall performance in the DPCI, DP and ELPD themes of the Deakin medical course and will correlate better with performance in later years of the course, where there is more emphasis upon clinical and professional interactions**

It was also hypothesised that:

7. **Performance on MMI taken in combination with performance on GPA and GAMSAT would better predict performance on the course overall than performance on the MMI alone,**

as these three selection tools are likely to provide useful but complementary information about ability to cope with the medical course.
The final hypothesis considers the effect of shortening interview duration on the predictive validity of the MMI. Given that the MMI assesses enduring traits that are likely to be consistently present throughout each 8 minute MMI stations, as such;

8. Shortening the duration of MMI stations will not affect the predictive validity of the MMI

The experimental data was collected in three phases:

1. Firstly, data relating to the reliability (internal consistency) of the Deakin MMI (Chapter 4 - Hypotheses 1 & 2) was collected during routine selection interviews during 2008.

2. Secondly, data relating to inter-rater reliability of the Deakin MMI (Chapter 5 - Hypotheses 3 & 4) was obtained from scripted interviews that were recorded and viewed by all participants.

3. Thirdly, assessment of MMI validity (Chapter 6 - Hypotheses 5-8), compared scores obtained by applicants during the 2008 MMI to scores obtained by the same students on relevant assessment tasks in subsequent years of the Deakin medical course.

In all cases, ethics approval and informed consent was obtained from participants.
References


5. Reliability of the Deakin MMI

5.1 Introduction

Determining optimal interview duration is important to ensure meaningful, reliable results are achieved in an efficient and timely manner. Stations must be long enough to permit precise comparison for ranking purposes (Roberts et al. 2008). However, longer interviews may be associated with reduced decision quality and participant fatigue (Campion, Palmer & Campion 1997).

Given previous observations that eight or more eight-minute MMI stations, each manned by a single interviewer, provides a selection tool of acceptable reliability, it was hypothesised that the incorporation of 10 professionalism-related outcomes into a MMI will produce a selection tool that is of acceptable reliability.

It has been suggested that individuals are likely to make early judgements in social encounters that involve observation of expressive behaviours (Ambady & Rosenthal 1992), and those that target emotive and moral qualities (Haidt 2001). It was therefore hypothesized that MMI interviewers would reach a decision regarding applicant performance relatively early in each interview and that shortening interview duration from 8 minutes to 5 minutes will have minimal impact on interview reliability and overall candidate rankings.
5.2 Methods

5.2.1 Deakin MMIs

One hundred and twenty places in the Deakin medical course were available in 2008. Applicants were selected for interview based on a composite score that incorporated undergraduate academic grades and performance in the Graduate Australian Medical Schools Admission Test (GAMSAT) (Aldous et al. 1997). The study was approved by the Deakin University Human Research Ethics Committee.

One hundred and seventy five applicants were interviewed at 10 stations in 18 cohorts over one week resulting in a total of 1750 mini-interviews. A total of 81 interviewers (faculty academics, clinicians and community members) participated. As standard procedure for the Deakin MMI, interviewers observed, scored and discussed several pre-recorded mock-interviews at a training and information session one week before interviews. Immediately prior to participating in interviews, interviewers and applicants attended independent briefing sessions. Interviewers were provided with an interviewer pack that contained score sheets, a scoring guide and suggested prompting questions for their station. Applicants provided written informed consent but remained blinded to the status (control or experimental) of each station. Applicants were informed that only scores at eight minutes would be used for selection ranking.

The study occurred within the context of the standard Deakin MMI of ten eight minute stations with two minutes preparation time between stations. At each station, a scenario was provided for discussion that was designed to address one of the ten core DMS outcomes: communication skills, professionalism, social justice,
evidence-use, self-directed learning, teamwork, effective use of resources, career motivation, health promotion and rural awareness. The outcome assessed at each of the ten stations remained constant for all interview cohorts.

The five experimental stations for each interview cohort were either stations 1, 3, 5, 7, and 9 (odd) or stations 2, 4, 6, 8 and 10 (even). This sequence alternated between interview cohorts. The scenario and interviewer at each station remained constant for two successive cohorts. Thus, each interviewer assessed 20 applicants on the same scenario, 10 under experimental conditions and 10 under control conditions.

A bell was rung to signal the beginning and end of each eight-minute mini-interview. At experimental stations, interviewers were alerted at the five-minute point by a sign shown through an open doorway. The arrangement within the room ensured that the signal was visible only to the interviewers and allowed the interview to proceed without the possibility of applicants being distracted by the signal.

At the conclusion of each eight minute mini-interview, all applicants were scored using the Deakin six-point scale (outlined in Chapter 4). At experimental stations, interviewers were asked to provide an additional score at the five-minute mark of the interview. After scoring two interview cohorts, interviewers attended a debriefing session where they provided verbal feedback regarding interview duration.
5.2.2 Statistical methods

Mean scores at eight minutes (experimental and control stations) and at five minutes (experimental stations) were calculated for each station and for all stations combined (pooled data). Mean cumulative scores across five control and five experimental stations were then calculated. Unpaired t-tests were used to compare mean scores at experimental and control stations and paired t-tests were used to compare five and eight minute scores at experimental stations. Correlations between five-minute and eight-minute scores at experimental stations were assessed using Pearson Correlation Coefficients (Pearson & Filon 1898). Rankings of applicants based on cumulative five-minute and eight-minute scores at experimental stations were compared using Spearman Rank Order Coefficient (Spearman 1904).

Generalisability theory (Cronbach et al. 1972) was used to estimate variance components via a random-effects, nested two-facet model (Applicants*Interviewers:Stations). A generalisability coefficient was calculated by dividing the estimated applicant variance component by estimated observed score variance for eight and five-minute scores. Confidence intervals and significance data were calculated to facilitate comparison of coefficients (Fan & Thompson 2001). Statistical calculations were performed using SPSS 15.0 software (SPSS Inc.). Minimum Norm Quadratic Estimation (MINQUE) was used for estimating variance components.
5.3 Results

5.3.1 MMI scores

Mean eight-minute scores at control stations and mean five and eight minute scores at experimental stations for each DMS outcome are shown in Table 5.1. For each experimental station the mean score at five minutes was lower than the mean score at eight minutes (p <0.05). For all stations except those assessing communication skills (S1), career motivations (S5) and social justice (S6), the difference between five and eight minute score was significant at the 0.01 level.

Table 5.1 Summary of mean scores achieved by applicants at 5 minutes (5E) and 8 minutes (8E) at experimental stations and at eight minutes (8C) at control stations, and Pearson correlation coefficients (Cp 8E/5E)) for 5 and 8 minute scores at experimental stations. Brackets denote 95% confidence intervals.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>8E (95% CI)</th>
<th>5E (95% CI)</th>
<th>8C (95% CI)</th>
<th>Cp 8E/5E (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 Communication Skills</td>
<td>3.45 (3.24-3.67)</td>
<td>3.32* (3.14-3.51)</td>
<td>3.50 (3.28-3.72)</td>
<td>0.85* (0.77-0.90)</td>
</tr>
<tr>
<td>S2 Evidence Use</td>
<td>3.76 (3.56-3.95)</td>
<td>3.36** (3.16-3.55)</td>
<td>3.61 (3.40-3.82)</td>
<td>0.85* (0.76-0.90)</td>
</tr>
<tr>
<td>S3 Health Promotion</td>
<td>3.55 (3.35-3.74)</td>
<td>3.27** (3.08-3.46)</td>
<td>3.79 (3.61-3.97)</td>
<td>0.82* (0.73-0.88)</td>
</tr>
<tr>
<td>S4 Teamwork</td>
<td>3.86 (3.67-4.04)</td>
<td>3.69** (3.52-3.87)</td>
<td>3.86 (3.65-4.06)</td>
<td>0.87* (0.81-0.91)</td>
</tr>
<tr>
<td>S5 Career</td>
<td>3.92 (3.73-4.11)</td>
<td>3.82* (3.62-4.01)</td>
<td>3.94 (3.76-4.11)</td>
<td>0.90* (0.85-0.93)</td>
</tr>
<tr>
<td>S6 Social Justice</td>
<td>3.95 (3.79-4.11)</td>
<td>3.86* (3.69-4.03)</td>
<td>3.78 (3.54-4.01)</td>
<td>0.91* (0.87-0.94)</td>
</tr>
<tr>
<td>S7 Self-directed learning</td>
<td>3.60 (3.41-3.79)</td>
<td>3.36** (3.18-3.55)</td>
<td>3.68 (3.50-3.87)</td>
<td>0.83* (0.75-0.89)</td>
</tr>
<tr>
<td>S8 Professionalism</td>
<td>3.66 (3.45-3.87)</td>
<td>3.49** (3.28-3.70)</td>
<td>3.56 (3.32-3.80)</td>
<td>0.88* (0.83-0.92)</td>
</tr>
<tr>
<td>S9 Resource Use</td>
<td>3.60 (3.41-3.79)</td>
<td>3.36** (3.16-3.57)</td>
<td>4.05** (3.89-4.21)</td>
<td>0.83* (0.75-0.89)</td>
</tr>
<tr>
<td>S10 Rural Awareness</td>
<td>3.57 (3.38-3.76)</td>
<td>3.39** (3.20-3.57)</td>
<td>3.40 (3.18-3.62)</td>
<td>0.86* (0.80-0.90)</td>
</tr>
</tbody>
</table>
The mean five and eight-minute scores across all experimental stations (pooled data) were 3.50 and 3.70 (p<0.01) respectively (Table 5.1). There was no difference between the scores at five minutes and eight minutes at 634 (72.5%) of the 875 experimental stations. Scores at eight minutes were one mark higher at 206 (23.5%) stations and one mark lower at 34 (4%) stations. Applicants who received a score between one and four after five minutes were more likely to receive a higher score after eight minutes (Figure 5.1).
Figure 5.1. Effect of reduced MMI station duration on station score.

Difference (mean +/- sem) between 5 and 8-minute scores at experimental stations in relation to five-minute score.

Mean cumulative scores based on five-minute and eight-minute scores at five experimental stations were 17.50 and 18.50 respectively (p < 0.01). For 45 applicants (26%) cumulative five minute and eight minute scores were identical, for 116 (66%) the cumulative eight-minute score was higher; by one mark for 61 (35%), by two marks for 42 (24%), by three marks and for 9 (5%) and by four marks for 4 (2%). Fourteen (8%) applicants had a cumulative eight-minute score that was one mark lower than the cumulative five-minute score. There was no relationship between the cumulative five-minute score and the difference between the cumulative scores at five minutes and eight minutes (Pearson correlation
coefficient \(-0.053; p = 0.48\). For all but the highest scoring applicants, improvement in cumulative score between five and eight minutes was similar regardless of performance at five minutes (Figure 5.2).

Figure 5.2. Effect of reduced MMI station duration on cumulative score.
Difference (mean +/- sem) between cumulative 5 and 8-minute scores at experimental stations in relation to cumulative 5-minute score.

There were strong and highly significant correlations between five minute and eight minute scores at all experimental stations (0.82 to 0.91, \(p<0.01\)) for all stations, and between the cumulative scores at five minutes and eight minutes (0.92, \(p<0.01\)) (Table 5.1).
There was no difference between the mean eight-minute scores at each station under experimental and control conditions, except at Station 9, which assessed attitudes to effective use of resources in medical practice. There was no significant difference between mean eight-minute scores at control and experimental stations (3.73 v 3.70; p = 0.56), or between cumulative eight-minute scores under experimental and control conditions (18.63 v 18.50; p = 0.55) (Table 5.1).

5.3.2 Applicant Ranking

A comparison of applicant rankings based on cumulative 5 minutes and 8 minutes scores for the experimental stations showed very little difference (Figure 5.3). The Spearman rank-order coefficient for rankings based on cumulative five-minute and eight-minute scores was 0.92 (95%CI 0.89-0.94). For one third of applicants, ranking did not change. The rankings of the remaining applicants changed by one to three positions. Changes were most pronounced for applicants with the highest and lowest rankings (Figure 5.4).
Figure 5.3. Effect of reduced MMI station duration on applicant ranking.

Changes in ranking observed when applicants were ranked using cumulative 5 and 8-minute scores

Figure 5.4 Effect of reduced MMI station duration on applicant ranking.

Ranking of applicants based on cumulative 5 and 8-minute scores
5.3.3 Reliability

Table 5.2 shows estimated variance components and generalisability coefficients for eight and five-minute scores. Three major variance components, together accounting for approximately 90% of total variance, were identified (Applicants, Applicant*Stations, and Applicant*Interviewer:Station) for both eight and five-minute scores. Generalisability coefficients were 0.78 and 0.75 for scores obtained at eight and five minutes respectively.

Table 5.2. MMI Reliability: Variance components and Generalisability coefficients (95% CI) for scores awarded at 5 and 8-minutes

<table>
<thead>
<tr>
<th></th>
<th>8 minute scores</th>
<th>5 minute scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 ) Applicant</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>( \sigma^2 ) Station</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>( \sigma^2 ) Interviewer:Station</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>( \sigma^2 ) Applicant*Station</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>( \sigma^2 ) Applicant*Interviewer:Station</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>Generalisability coefficient</td>
<td>0.78 (0.73-0.82)</td>
<td>0.75 (0.70-0.80)</td>
</tr>
</tbody>
</table>

5.3.4 Feedback

Interviewers who had manned one of the less complex stations such as communication skills and career motivation, agreed that five minutes was ample time to provide an accurate assessment of applicant performance. Several interviewers at these stations reported that most applicants had fully addressed the scenario before the eight-minute bell rang to end the station. Interviewers at more
complex stations felt that mini-interview length could be reduced to five minutes if more direction was provided to applicants who did not immediately address the station outcome. Most felt that appropriate prompting would have enabled applicants to score higher at the five-minute mark at these stations.

5.4 Discussion

This study has investigated the effects of scoring MMI stations at five and eight minutes for 175 graduates applying for selection into the Deakin Medical School. Mean scores at individual stations and mean cumulative scores across multiple stations were slightly higher when applicants were assessed after eight minutes. The study had sufficient power (0.996 for pooled data and 0.923 for cumulative data) to demonstrate mean scores that were on average 0.2 marks higher per station. Strong, highly statistically significant correlations were found between five and eight-minute scores at single stations and between cumulative five and eight-minute scores. Applicant rankings based on scores awarded after five and eight minutes were almost identical.

For the majority of applicants scores awarded at five and eight minutes were identical. However, the final three minutes were beneficial to a minority, most of whom did not perform well in the first five minutes (Figure 5.2). These applicants may have pursued a more indirect path in addressing key criteria of the station or were able to react to prompts from interviewers during the final three minutes. A small minority, mostly those who had high scores at five minutes, lost marks in the final three minutes. There was no relationship between cumulative five-minute score and cumulative improvement in score during the final three minutes of
interviews, indicating that longer interviews did not provide extra benefit for lower performing applicants (Figure 5.2). The overall effect of an additional three minutes for each station was to increase the cumulative scores of all but the highest scoring applicants with minimal effect on ranking of applicants. When changes in ranking did occur, they were most pronounced for the highest and lowest ranking applicants. This has important implications, as these groups are least likely to be affected by subtle changes in rankings. In our study, where 175 applicants competed for 120 places, only one applicant ranking within the top 120 positions based on eight-minute scores dropped out of the top 120 when applicants were ranked according to five-minute scores (Figure 5.4). The applicant returned to the top 120 when scores from control stations were combined with five-minute scores prior to ranking applicants.

The reliability for the 10 station, 8-minute MMI was 0.78. This figure represents a degree of reliability that is acceptable and compares favourably with reliability coefficients for MMI reported elsewhere. Reliability coefficients and variance components were by and large unaffected by duration of stations. In both the five and eight-minute model, most variance was due to Applicant, Applicant*Station and Applicant*Interviewer:Station interactions, reflecting differences in the ability of applicants, the context specificity of different DMS outcomes and interviewer variation. Applicant*Interviewer:Station interactions accounted for less variance than Applicant*Station interactions. This is likely to reflect a consistent approach to scoring by different interviewers that was facilitated by interviewer training sessions and scoring guides.
The nature of the scenario chosen at an MMI station may influence the optimal duration of that station. Depending on the course outcome addressed, stations varied in their complexity and in their ability to elicit moral issues and expressive behaviours. The power of the study was insufficient to discriminate a mean difference of less than 0.4 between five and eight-minute scores at stations addressing different outcomes. However, differences this small are unlikely to be practically significant. In general the differences between five and eight minute scores were less and the correlations between the scores were stronger for stations based on less complex scenarios although this was not statistically significant. No differences were observed between stations more or less likely to be affected by moral intuition or expressive behaviours. This may have been due to our attempts to enhance scenario structure by providing interviewers with training sessions and detailed scoring guides as both of these have been shown to minimize interviewer reactions and improve the psychometric properties of selection interviews (Campion, Palmer & Campion 1997).

In line with previous studies (Eva, Rosenfeld, et al. 2004), qualitative feedback received from interviewers supports the notion that shortening the duration of MMI stations from eight to five minutes does not affect the outcome of the MMI. No feedback was collected from applicants, and such feedback may not have been useful as they were not aware which stations were conducted under experimental or control conditions. However, changing the duration of MMI stations may impact upon the acceptability of the test to applicants and represents an area for further study. Eva and colleagues (Eva, Reiter, et al. 2004) reported that applicants found that eight minutes was an appropriate station duration. Stations perceived as too
short by applicants may adversely affect performance by limiting the time for interviewer and applicant to develop rapport and by pressuring applicants to formulate hurried responses. Adequate preparation time between stations may help to reduce time pressures and facilitate effective brief interviews.

A limitation of this study was our inability to blind interviewers. It is possible that a score given at five minutes could influence the score given by the same interviewer at eight minutes. This was addressed by the design of the study, where each station was manned by the same interviewer for two successive rounds of interviews and the same scenarios were used for both rounds. In each round ten applicants were interviewed under experimental conditions (scoring at five and eight minutes) and ten under control conditions (scoring at eight minutes). The absence of a significant difference between eight-minute scores under experimental and control conditions at nine of the ten stations suggests that scores at eight minutes were not influenced by scores at five minutes. Furthermore there was no difference between cumulative eight-minute scores from control and experimental stations.

The observations reported here for performance on the MMI correspond with those observed for performance on other short structured tasks such as the Objective Structured Clinical Examination (OSCE). The MMI represents an approach to selection interviews that is similar in principle to the use of the OSCE for assessing clinical performance (Eva, Rosenfeld, et al. 2004). For OSCEs, the number and duration of stations often represents a compromise between feasibility (available resources and participant fatigue) and reliability. OSCE comprising as few as 4 and as many as 35 stations have been documented (Davis 2003; Hodges et al. 2002).
OSCE stations are typically 5 to 15 minutes in length, although stations ranging from 4 minutes to over an hour have been reported (Hodges et al. 2002). Station duration appears to have little effect on student performance at a variety of OSCE-based structured tasks (Schoonheim-Klein et al. 2007).

Determining optimal interview duration is important to ensure meaningful, reliable results are achieved in an efficient and timely manner. This study demonstrates that reducing the duration of MMI stations from eight to five minutes provides a means of conserving scarce resources with minimal effect on applicant ranking and without compromising reliability. Further studies are required to establish the acceptability of five-minute stations to applicants and the ability of five-minute stations to predict future performance.

Further studies are also required to assess the optimal duration of MMI stations and to assess how interview questions and scoring guides influence applicant performance and test reliability.
References


6. Scoring characteristics and inter-rater reliability of interviewers for the MMI

6.1 Introduction

In interviews that incorporate multiple separate interviews, test reliability reflects both the level of agreement between interviewers (inter-rater reliability), and consistency in applicant performance across interviews (internal consistency) (Conway, Jako & Goodman 1995). Applicant performance has been addressed in a previous section that showed the internal consistency of the Deakin MMI to be acceptable and comparable to reliability coefficients reported elsewhere (Eva et al. 2004; Dodson et al. 2009). This chapter examines the inter-rater reliability of interviewers recruited to assess candidates during selection interviews.

Inter-rater reliabilities (IRR) reported for panel interview and separate interviews are generally in the order of 0.7-0.9 and 0.4-0.6 respectively, with more structured interviews yielding higher reliability coefficients (Conway, Jako & Goodman 1995). However, Axelson and colleagues have recently observed higher IRR for unstructured medical student selection interviews and observed that IRR was highest for interviews that incorporated both structured and unstructured elements in approximately equal proportions (Axelson et al. 2010). This observation led the authors to contend that unstructured interviews were potentially useful components of medical student selection processes. The authors recognised however that the IRR of such interviews were likely to be context dependent and influenced by interview questions and context.
These findings have important implications for MMI used as part of the selection process for applicants to medical school where they are routinely employed to assess various cognitive and non-cognitive qualities. However, the qualities assessed, interview questions, interview format, number of stations, interview duration, and the scoring processes vary widely. A number of studies have considered IRR of MMIs used for selection into specialist training schemes and report values between 0.36 and 0.7 (Bandiera & Regehr 2004; Blouin 2010; Finlayson & Townson 2011; Gilbart, Cusimano & Regehr 2001). However, the IRR of selection MMIs for entry into medical school remains largely explored. In particular, little is known regarding the IRR of interviewers sourced from different stakeholder groups, the effects of experience and the effect of factors such as outcomes assessed.

The Deakin MMIs are best categorised as semi-structured interviews in that applicants are largely assessed on an equivalent set of 10 standardised questions through free-flowing dialogue with the interviewer. No two interviews at a particular station are the same, however interviewers are trained and provided with guidance material so that the same principles are covered for all applicants at a given station. A scoring guide is provided to aid scoring against a global scale, but interviewers are not required to complete a detailed assessment rubric.

Three key stakeholder groups are employed routinely in Deakin Medical School selection interviews; members of the academic staff of Deakin University, healthcare professionals, and community members. Community members are allocated stations that assess communication skills, social justice and resource use
(stations 1, 6 and 9), clinicians are allocated stations that assess health promotion, career/motivation, professionalism and rural awareness stations (Stations 3, 5, 8 and 10) and academics are allocated stations that assess evidence use, teamwork and self directed learning (Stations 2, 4 and 7). The matching of interviewers with certain stations is done to maximise face validity and ensure interviewers are comfortable with the content of the station. However, the value of this process in terms of IRR is unknown.

In Chapter 2, a paucity of data relating to the relative performance of different interviewer subgroups in medical student selection interviews was identified. The tendency for some medical schools to engage medical students as an additional stakeholder group was also highlighted (Fruen 1983; Fulton 1979; Harris & Owen 2007; Parry et al. 2006; Tambyah 2005). Although some evidence suggests that medical students may be less decisive and less discriminating in their scoring than more experienced interviewers (Koc, Katona & Rees 2008). There is no data currently available that examines how medical student scoring for the MMI compares with that of other key stakeholder groups.

The study was designed to assess the IRR of interviewers for the Deakin MMI and to compare the scoring characteristics and IRR of standard interviewers (health professionals, academics and community members), and junior medical students on a common series of scripted MMIs. Given this combination of structural and non-structural elements, it was hypothesised that the IRR of a representative Deakin MMI would be satisfactory.
It was also hypothesized that the IRR of interviewer subgroups (including medical students) would not be greatly affected by station theme or previous exposure to a given station. This was expected as the qualities assessed at many stations were underpinned by broad socio-cultural norms and the appropriateness of candidate responses would therefore be viewed similarly by most interviewers irrespective of occupation, and because all interviewers would have received standardised training. These factors considered, IRR will be similar for all standard interviewer subgroups (community members, clinicians and academics), and for a novel group of interviewers composed of current medical students.

Understanding the IRR of MMI interviewers, and the effects of factors such as previous exposure to a particular scenario and the nature of the outcomes assessed, is important in order to inform the judicious use of limited resources (interviewers), and to ensure selection interviews are valid and reliable.

6.2 Methods

Ten five-minute MMI were constructed; one for each of the 10 key outcomes of the Deakin Medical Course. For each station a script was constructed for interviewer and applicant such that the applicant would achieve a predetermined level of performance between 1 and 5. Predetermined performance levels reflected performance during the routine Deakin selection MMI. As such, most stations were scripted to a score between 3 and 5. The total scripted score (summed across all 10 stations) was 33 (4,4,3,4,5,2,4,3,3,1).
All ten scripted stations were video recorded. Existing first and second year students played the roles of candidates and interviewer. The interviewer was held constant for all interviews to minimise any “interviewer variation” that may have influenced scoring. In order to replicate routine selection interviews, each interview was terminated after 5 minutes regardless of whether the applicant had finished. A bell was rung to signal the conclusion of the interview.

Junior (first and second year) medical students at Deakin University and standard interviewers for the 2009 Deakin University selection MMI were invited to participate. Standard interviewers consisted of three main groups of stakeholders: community members, clinicians and academics.

All participants were trained in identical fashion. This involved a presentation and observation/scoring of two mock interviews. The scenario of the first mock interview was the same as that used at Station 9 of the experiment. For interviewers, the experiment was conducted following routine training for selection interviews.

For each station, participants were provided with the following routine documentation available to interviewers during selection MMI:

- A description of the scenario
- Relevant background information
- A scoring guide
- A score sheet
Although interviewers are matched with appropriate stations during routine selection MMI to maximise face validity and prevent undue interviewer stress, for the purposes of the experiment, interviewers were asked to score all 10 interviews. Participants were asked to watch all 10 interviews in succession. Participants were provided with 2 minutes between interview for scoring and preparation. Participants were instructed not to discuss scores until all interviews had been scored and score sheets had been collected. Participants’ scores were excluded from subsequent analysis if they had scored fewer than 5 interviews.

As in routine DMS selection MMI, scores were awarded on a global 6 point scale where 1 = unsatisfactory, 2 = borderline, 3 = satisfactory, 4 = good and 5 = excellent, and a score of zero was given to applicants whose performance raised questions about their suitability for a career in medicine.

Three of the stations included in the study had been used previously during routine selection interviews. This was reflective of routine practice where approximately two-thirds of the stations used during the Deakin selection MMI are new stations and one third of stations have been used previously. As a result, some interviewers had gained experience scoring certain scenarios during routine selection interviews and some students had been interviewed at certain scenarios during their own selection interviews. As this reflects the situation in routine Deakin interviews, both “experienced” and “naive” interviewers were included in the principle analysis and the effects of “experience” were considered as a sub-analysis. In order to enable such an analysis, participants were asked to indicate whether they had experienced
the scenario (either as an interviewer in the case of standard interviewers, or as an applicant in the case of student interviewers) during routine selection interviews.

The study was powered to discriminate differences in score of approximately 0.5 points or differences in overall IRR of greater than 0.05. Data from student scorers was compared to pooled standard interviewer data and data from each subgroup of standard interviewers.

The distribution of total scores for the 10 station MMI for each interviewer subgroup were recorded and compared. Mean MMI scores awarded by interviewer subgroups were also calculated and compared using Spearman’s correlation coefficient. Spearman’s coefficient was chosen over Pearson’s coefficient because the data were not normally distributed. Individual station scores awarded by each participant group were compared for all stations using a number of measures appropriate for interval data. Frequency histograms were computed to enable a comparison of the distribution of scores awarded.

IRR was calculated using Krippendorff alpha coefficient. Krippendorff’s alpha coefficient is a statistical measure of the extent of agreement among coders that has been used increasingly in the literature because it provides sufficient flexibility to account for any number of coders, incomplete (missing) data, unequal sample sizes and to any number of values available for coding a variable (Hayes & Krippendorff 2007). The computation of Krippendorff’s alpha is complex and has been fully outlined by Krippendorff (Krippendorff 2011). The measure ($\alpha$) is based on the underlying principle that:
\[ \alpha = 1 - \frac{D_o}{D_e} \]

where \( D_o \) is the observed disagreement amongst values assigned to units of analysis and \( D_e \) is the disagreement expected if the coding of units is attributable to chance rather than the properties of the units. \( D_o \) and \( D_e \) are calculated using the frequencies of values in coincidence matrices.

SPSS 1.6.0 was used to calculate Krippendorff’s alpha for interval data using a macro described by Andrew Hayes, Ohio State University (Hayes 2011). The IRR of each interviewer subgroup was compared for the MMI as a whole, for each subset of interviewer-matched stations, and for experienced and naïve interviewers.

6.3 Results

Sixty four standard interviewers and 72 student interviewers participated in the study. Two standard interviewers scored fewer than 5 stations and were excluded from the analysis, leaving a total of 62 standard interviewers (21 community members, 23 clinicians and 18 academics) and 72 student interviewers that were included in the analysis. Sixteen of the standard interviewers and 58 of student interviewers had been previously exposed (interviewed at or been interviewed at) to at least one of the 10 station scenarios included in the study.

6.3.1 Scoring characteristics of interviewers

Figure 6.1 shows mean total score (summed across all 10 stations) for students and standard interviewers (as a whole and for each subgroup). For all participants, total scores awarded were within 8 points (25-39) of the scripted score and mean score
for each participant group was within 3 points of the scripted score. Table 6.I shows further details regarding the scoring characteristics of each participant group for the MMI as a whole and for each of the ten stations.

Figure 6.1 Mean total scores for interviewers of the Deakin MMI

![Figure 6.1 Mean total scores for interviewers of the Deakin MMI](image)

Table 1. Mean scores (total and individual stations) for interviewers of the Deakin MMI

<table>
<thead>
<tr>
<th></th>
<th>Interviewers</th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>31.37 (30.48-32.27)</td>
<td>30.62 (29.06-32.18)</td>
<td>32.48 (30.97-33.99)</td>
<td>30.83 (29.98-32.58)</td>
<td>33.76 (33.03-34.49)</td>
</tr>
<tr>
<td>Station 1 (4)*</td>
<td>3.82 (3.64-4.01)</td>
<td>3.71 (3.33-4.10)</td>
<td>3.83 (3.54-4.11)</td>
<td>3.94 (3.58-4.31)</td>
<td>4.33 (4.19-4.48)</td>
</tr>
<tr>
<td>Station 2 (4)</td>
<td>3.85 (3.70-4.01)</td>
<td>3.81 (3.5-4.12)</td>
<td>3.91 (3.69-4.16)</td>
<td>3.83 (3.53-4.14)</td>
<td>4.08 (3.04-4.22)</td>
</tr>
<tr>
<td>Station 3 (4)</td>
<td>3.73 (3.53-3.92)</td>
<td>3.76 (3.41-4.11)</td>
<td>3.73 (3.37-4.11)</td>
<td>3.67 (3.33-4.01)</td>
<td>4.44 (4.26-4.62)</td>
</tr>
<tr>
<td>Station 4 (4)</td>
<td>4.21 (4.02-4.40)</td>
<td>4.24 (3.89-4.59)</td>
<td>4.26 (3.96-4.56)</td>
<td>4.11 (3.70-4.52)</td>
<td>4.31 (4.13-4.48)</td>
</tr>
<tr>
<td>Station 5 (5)</td>
<td>4.48 (4.33-4.64)</td>
<td>4.33 (4.03-4.63)</td>
<td>4.65 (4.40-4.90)</td>
<td>4.44 (4.14-4.75)</td>
<td>4.71 (4.59-4.83)</td>
</tr>
<tr>
<td>Station 6 (2)</td>
<td>1.98 (1.77-2.20)</td>
<td>1.95 (1.59-2.31)</td>
<td>1.96 (1.58-2.34)</td>
<td>2.06 (1.62-2.49)</td>
<td>1.75 (1.56-1.94)</td>
</tr>
<tr>
<td>Station 7 (4)</td>
<td>3.79 (3.58-4.00)</td>
<td>3.76 (3.33-4.19)</td>
<td>4.13 (3.83-4.43)</td>
<td>3.39 (3.04-3.74)</td>
<td>4.04 (3.84-4.24)</td>
</tr>
<tr>
<td>Station 8 (2)</td>
<td>2.32 (2.10-2.54)</td>
<td>2.00 (1.68-2.32)</td>
<td>2.52 (2.09-2.95)</td>
<td>2.44 (2.05-2.83)</td>
<td>2.54 (2.31-2.77)</td>
</tr>
<tr>
<td>Station 9 (3)</td>
<td>2.56 (2.36-2.77)</td>
<td>2.57 (2.20-2.94)</td>
<td>2.70 (2.34-3.05)</td>
<td>2.39 (2.00-2.78)</td>
<td>3.07 (2.86-3.28)</td>
</tr>
<tr>
<td>Station 10 (1)</td>
<td>0.61 (0.42-0.80)</td>
<td>0.48 (0.20-0.75)</td>
<td>0.78 (0.37-1.19)</td>
<td>0.55 (0.25-0.86)</td>
<td>0.49 (0.37-0.60)</td>
</tr>
</tbody>
</table>

*Brackets denote scripted score
Mean total MMI scores for each subgroup were within 3 points of the total scripted score. Total student scores were significantly higher (two-tailed unpaired t test) than community members ($p=0.0007$), academics ($p=0.035$) and standard interviewers ($p=0.0001$) as a whole but not clinicians ($p=0.130$). Students scored higher than interviewers at all stations except 6 and 10 which were the lowest scored stations by all subgroups. These differences were significant for Station 1 (communication skills), Station 2 (evidence use), Station 3 (health promotion), Station 5 (career/motivation) and Station 9 (resource use). For Station 3 mean student score was also significantly higher than each subgroup of standard interviewers. Total scores awarded by subgroups of standard interviewers did not differ significantly ($p=0.08 - 0.85$).

A high degree of correlation was observed between mean total scores awarded by students and standard interviewers (Spearman coefficient 0.84). Correlations (Spearman coefficients) between all subgroups of interviewers, including students, ranged between 0.79 and 0.96, and were highly significant ($P<0.005$) in all cases (see Table 6.2). The relationship between mean scores awarded by different interviewer subgroups is shown in Figure 6.2. The figure shows the consistency of scoring for all interviewer subgroups across all stations regardless of level of applicant performance.
Table 6.2 Spearman coefficients comparing means scores at all stations for all interviewer subgroups. (p< 0.005 in all cases)

<table>
<thead>
<tr>
<th></th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinicians</td>
<td>0.96</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academics</td>
<td>0.91</td>
<td>0.90</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>0.85</td>
<td>0.79</td>
<td>0.90</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 6.2 Scatterplot showing means responses for all stations and all interviewer subgroups. Each panel contains 10 data points, one for each of the 10 MMI stations, and compares means scores awarded by two interviewer subgroups at each MMI station. For example, the top row of panels compares scores awarded by community members, clinicians and academics (X axis) with those awarded by students (Y axis) for all 10 stations.
6.3.2 Inter-rater reliability of the Deakin MMI

Percentage agreement with scripted station score ranged between 26% and 74% and is shown for all interviewer subgroups in Table 6.3. Figure 6.3 shows the distribution of scores awarded by members of each participant group at a typical station (Station 4).

Table 6.3 Interviewer percentage agreement with scripted station score

<table>
<thead>
<tr>
<th>Station</th>
<th>Scripted score</th>
<th>All Standard interviewers</th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>53.23</td>
<td>38.10</td>
<td>69.57</td>
<td>60.00</td>
<td>50.00</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>67.74</td>
<td>66.67</td>
<td>69.57</td>
<td>50.00</td>
<td>68.06</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>48.39</td>
<td>52.38</td>
<td>47.83</td>
<td>44.44</td>
<td>26.39</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>45.16</td>
<td>38.10</td>
<td>47.83</td>
<td>50.00</td>
<td>36.11</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>54.84</td>
<td>42.86</td>
<td>69.57</td>
<td>50.00</td>
<td>73.61</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>35.48</td>
<td>38.10</td>
<td>26.09</td>
<td>44.44</td>
<td>34.72</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>41.94</td>
<td>38.10</td>
<td>52.17</td>
<td>33.33</td>
<td>38.89</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>43.55</td>
<td>52.38</td>
<td>39.13</td>
<td>38.89</td>
<td>40.28</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>41.94</td>
<td>47.62</td>
<td>39.13</td>
<td>38.89</td>
<td>50.00</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>41.94</td>
<td>38.10</td>
<td>43.48</td>
<td>44.44</td>
<td>48.61</td>
</tr>
</tbody>
</table>
Figure 6.3 Distribution of scores awarded at Station 4 by all interviewer subgroups

For all stations, and for all groups and subgroups, over 85% of scores fell within +/-1 of the scripted score. The distribution of total scores (summed across all ten stations) for each interviewer subgroup is shown in Figure 6.4. Within each interviewer subgroup, the spread of scores awarded was similar, extending over 12 points for students, over 13 points for community members and over 11 points for clinicians and academics.
Figure 6.4 Total MMI scores (summed over all ten stations) awarded by interviewers

The IRR for standard interviewers of the Deakin MMI was 0.70. The IRR standard interviewer subgroups (community members, clinicians, and academics) and student interviewers are shown in Table 6.4. The IRR of student interviewers was significantly higher than that of standard interviewers as a whole, and higher than any standard interviewer subgroup.

Table 6.4 Krippendorff’s alpha for interviewers of the Deakin MMI

<table>
<thead>
<tr>
<th>Group</th>
<th>Students</th>
<th>All standard interviewers</th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
</tr>
</thead>
<tbody>
<tr>
<td>K alpha</td>
<td>0.76 (0.75-0.76)</td>
<td>0.70 (0.69-0.70)</td>
<td>0.71 (0.69-0.72)</td>
<td>0.69 (0.67-0.71)</td>
<td>0.71 (0.69-0.73)</td>
</tr>
</tbody>
</table>
6.3.3 Inter-rater reliability for interviewer subgroups and matched-stations

Table 6.5 shows the IRR for each interviewer subgroup for each cluster of interviewer-matched stations. IRR was highest for all interviewer subgroups for the clinician-matched cluster of stations. For this cluster (Stations 3: Health promotion, 5: Career/motivation, 8: Professionalism and 10: Rural awareness), the IRR of clinician interviews was significantly lower than for other interviewer subgroups, while the IRR of student interviews was significantly higher than all other interviewer subgroups except community members.

<table>
<thead>
<tr>
<th>Group</th>
<th>Interviewers</th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Members</td>
<td>0.4817 (0.4643-0.4981)</td>
<td>0.4356 (0.3764-0.4882)</td>
<td>0.4827 (0.4345-0.5275)</td>
<td>0.5138 (0.4576-0.5684)</td>
<td>0.6454 (0.6340-0.6561)</td>
</tr>
<tr>
<td>Clinicians</td>
<td>0.7914 (0.7838-0.7990)</td>
<td>0.8308 (0.8150-0.8453)</td>
<td>0.7387 (0.7114-0.7648)</td>
<td>0.8245 (0.8068-0.8416)</td>
<td>0.8466 (0.8419-0.8512)</td>
</tr>
<tr>
<td>Academics</td>
<td>0.0502 (0.0142-0.0845)</td>
<td>0.0382 (-0.0667-0.1348)</td>
<td>0.0214 (-0.0736-0.1117)</td>
<td>0.1192 (-0.0024-0.2390)</td>
<td>0.0149 (-0.0159-0.0457)</td>
</tr>
</tbody>
</table>

IRR was modest for all interviewer subgroups at the cluster of stations matched to community members (Stations 1: Communication skills, 6: Social justice and 9: Resource use). IRR was significantly higher for student interviewers than for all other interviewer subgroups. As a subgroup, community members displayed the lowest IRR, although this was significant only compared to student interviewers.
Inter-rater reliability was poor for all interviewer subgroups for the academics-matched cluster of stations (Stations 2: Evidence use, 4: Teamwork and 7: Self directed learning). The IRR for academics was higher than other interviewer subgroups, however this difference was not statistically significant.

6.3.4 Inter-rater reliability for experienced and naïve interviewers

The IRR of participants with experience interviewing (standard interviewers) or being interviewed at specific stations included in the MMI were compared using Krippendorff’s alpha. For both standard and student interviewers previous exposure to a station improved IRR (Table 6.6).

Table 6.6 Krippendorff’s alpha for experience as interviewer or interviewee

<table>
<thead>
<tr>
<th>Group</th>
<th>All</th>
<th>Experienced*</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer</td>
<td>0.6973</td>
<td>0.8491</td>
<td>0.6985</td>
</tr>
<tr>
<td></td>
<td>(0.6910-0.7031)</td>
<td>(0.7434-0.9358)</td>
<td>(0.6918-0.7055)</td>
</tr>
<tr>
<td>Students</td>
<td>0.7569</td>
<td>0.8268</td>
<td>0.7277</td>
</tr>
<tr>
<td></td>
<td>(0.7527-0.7612)</td>
<td>(0.8141-0.8395)</td>
<td>(0.7210-0.7344)</td>
</tr>
</tbody>
</table>

*Experience for interviewers= previously interviewed at same station during selection MMI
*Experience for students= previously had been interviewed at that station during selection MMI

6.4 Discussion

In keeping with the hypotheses, this chapter has shown the IRR of standard interviewers for the Deakin MMI is satisfactory (Krippendorff’s alpha = 0.70) with no significant variation in IRR between standard interviewer subgroups. This section has also shown medical students to be a valid and reliable group of interviewers, with an IRR that was significantly higher than achieved by each
standard interviewer subgroup and standard interviewers as a whole. Prior exposure to a particular MMI station improves IRR but matching stations to interviewers does not.

Scores awarded for the ten MMI stations were highly consistent between interviewer subgroups and with scripted scores. Students tended to score higher on stations scripted to a higher level of performance and lower on those scripted to lower performance levels. This may reflect lesser awareness of the potential range of interviewee performances at MMI stations secondary to a relative lack of experience as interviewers. Outside of standard interviewer training, this was the first opportunity for students to score performance at MMI stations. In contrast, most of the participating standard interviewers had scored multiple applicants during routine Deakin selection MMIs. This experience would have provided interviewers with exposure to a range of high and low scoring interview performances, providing benchmarks that resulted in a moderation of scoring during the study.

Percentage agreement with scripted scores and spread of scores for each subgroup of interviewers was broadly similar. Percentage agreement with scripted score varied between 26% and 74% and was generally in the order of 40-50%. The usefulness of this as a measure of inter-rater agreement is limited however, as the scripted score was not always the most commonly chosen score by members of an interviewer subgroup.
The use of Krippendorff’s alpha to assess IRR helped overcome these limitations and permitted the calculation of a reliability coefficient that could cope with the complexities of the collected data. The overall IRR of standard interviewers for the Deakin MMI was 0.70 which compares favourably with IRR reported for residency selection interviews in orthopaedics and emergency medicine (Bandiera & Regehr 2004; Blouin 2010; Finlayson & Townson 2011; Gilbart, Cusimano & Regehr 2001). The value of 0.70 was reflective of the inter-rater reliabilities of the three standard interviewer subgroups (0.71 for community members and academics, and 0.69 for clinicians).

The IRR of student interviewers was significantly higher than each subgroup of standard interviewers and all standard interviewers combined. This was an unexpected finding, particularly as many standard interviewers had participated in routine Deakin MMIs on multiple occasions and had received training on each occasion. The most likely explanation for this is related to another finding of the study; that previous exposure to a station improves IRR. Indeed, 170 of 720 data points (interviewers*station) for students involved “experienced” interviewers compared to only 19 of 620 data points for standard interviewers. It is likely that the increased number of ‘experienced’ student interviewers contributed to the improved IRR in that group compared to standard interviewers.

For both standard interviewers and student interviewers, prior exposure to a station was found to improve IRR. It is possible that the process of working through a scenario on previous occasions helps interviewers develop a better reference against which to score subsequent interviews. If this were the case, one would
expect the improvement in IRR to increase for interviewers with multiple previous exposures. The observation that standard interviewers appeared to gain more benefit from previous exposure to a station than did student interviewers supports this notion. Standard interviewers are asked to assess a given station 20-30 times during routine interviews, whereas experienced student interviewers would only have encountered a station once (as an interviewee).

Matching interviewers to specific MMI stations was not found to improve IRR. A number of factors may have contributed to this observation; the provision of background material and scoring guides provide interviewers with sufficient depth of knowledge to assess at each station and consistent guidelines for scoring, all interviewers are also exposed to standardised interviewer training and peer review. Furthermore, assessments of good and bad performance at MMI stations are also likely to be consistent across interviewer subgroups as acceptable approaches towards the value-loaded scenarios assessed during the MMI are broadly accepted within contemporary society irrespective of occupation.

IRR was highest for the ‘Clinicians’ cluster of stations. It is possible that the qualities measured at those stations, or more effective scoring guides thereof, may have promoted a more consistent approach to scoring. However, a more likely explanation is in the nature of the experimental method and that the clinicians cluster consisted 4 stations (rather than three) that were characterised by a generous spread of scores. Both these factors are likely to have increased IRR. Conversely, IRR for academic stations was poor. The most likely explanation being that the scripted score for all three academic stations was four. This means that slight
differences in scoring between interviewers across these stations would have resulted in a profound reduction in IRR. It is also possible that scoring guides and background material provided at these stations was insufficient.

The results of this study indicate that interviewer matching does not improve IRR and this is also evident when attempts are made to re-allocate stations amongst the four interviewer groups (community members, clinicians, academics and students) (see Table 6.7). However, factors other than IRR must also be considered and failure to match interviewers to stations may result in a loss of face validity and patronage as interviewer volunteers may not be so forthcoming if they feel out of depth on the stations they are asked to man. Nevertheless, where resources are limited, interviewers from any subgroup who have been trained and feel comfortable with station content represent a pool of reliable interviewers that may be called upon to man other stations.

Table 6.7 K-alpha for proposed stations allocated to interviewer subgroups including students

<table>
<thead>
<tr>
<th>Group</th>
<th>Students</th>
<th>Interviewers</th>
<th>Community</th>
<th>Clinicians</th>
<th>Academics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community members</td>
<td>0.7999 (0.7936-0.8067)</td>
<td>0.6729 (0.6617-0.6849)</td>
<td>0.6290 (0.5936-0.6654)</td>
<td>0.7154 (0.6874-0.7435)</td>
<td>0.6555 (0.6143-0.6974)</td>
</tr>
<tr>
<td>Clinicians</td>
<td>0.8114 (0.8047-0.8179)</td>
<td>0.7180 (0.7050-0.7301)</td>
<td>0.7895 (0.7681-0.8113)</td>
<td>0.6251 (0.5800-0.6622)</td>
<td>0.7718 (0.7427-0.7978)</td>
</tr>
<tr>
<td>Academics</td>
<td>0.3057 (0.2750-0.3365)</td>
<td>0.4541 (0.4283-0.4786)</td>
<td>0.4038 (0.3150-0.4774)</td>
<td>0.4391 (0.3637-0.5098)</td>
<td>0.5142 (0.4354-0.5866)</td>
</tr>
<tr>
<td>Students</td>
<td>0.0196 (-0.0148-0.0544)</td>
<td>0.0588 (0.0166-0.1003)</td>
<td>0.0513 (-0.0525-0.1565)</td>
<td>-0.0133 (-0.1264-0.0977)</td>
<td>0.1658 (0.0318-0.2976)</td>
</tr>
</tbody>
</table>
This study has identified students as an additional reliable subgroup of potential assessors at MMI stations. The best way of utilising these extra resources requires consideration. Allocating students to stations that assess qualities such as teamwork or self-directed learning would appear to be associated with high face validity. However, students were identified as the most reliable interviewers across the ten MMI stations and are likely to be the most readily available subgroup of interviewers. They are also the subgroup most likely to be faced with potential conflicts of interest in the event of encountering friends or past classmates during routine selection interviews. These factors suggest that students may be most valuable as a pool of interviewers that could be called upon to man any MMI stations at times of need, such as when insufficient numbers of standard interviewers (community member, clinician and academic interviewers) have volunteered to participate in selection interviews or when an interviewer belonging to one of these subgroups is unable to attend a scheduled round of interviews at late notice.

As prior exposure to an MMI scenario was found to improve IRR, consideration must also be given to the possibility of training interviewers specifically on the stations they will man during selection MMIs. This is likely to be impractical, however, due to the resources required to ensure every interviewer is trained at his/her station. Another approach may involve making efforts to match interviewers with stations manned in previous years. This would serve to ensure interviewers have been previously exposed to the stations they are asked to man. However this is unlikely to be possible in all cases as there is a degree of interviewer turnover for the MMI, with some interviewers failing to return for interviews in subsequent
years, and others volunteering for the first time. Repeated interviewing at the same station by ‘experienced’ interviewers is also likely to result in interviewer boredom and reduced motivation during interviews, thereby directly contributing to increased interviewer turnover.

This chapter has demonstrated that the IRR for interviewers for the Deakin MMI is satisfactory and that students are a valid and reliable group of interviewers. The chapter also revealed that prior exposure to a particular MMI station improves IRR but matching stations to interviewers does not. Incorporating the findings of this study into a process of quality improvement for the MMI is complex and requires consideration of a variety of additional factors, however these findings help inform the judicious use of limited resources (interviewers) whilst ensuring that selection interviews remain valid and reliable.
References


Blouin, D 2010, 'Reliability of a structured interview for admission to an emergency medicine residency program', *Teaching and Learning in Medicine*, vol. 20, no. 4, pp. 246-50.


Fulton, T 1979, 'The address to the students', paper presented to Opening of the Winter Session Royal Victoria Hospital, London, 4 October.


Harris, S & Owen, C 2007, 'Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School', *Medical Education*, vol. 41, pp. 234-41.


Hayes, A & Krippendorff, K 2007, 'Answering the Call for a Standard Reliability Measure for Coding Data', *Communication Methods and Measures*, vol. 1, no. 1, pp. 77-89.


Tambyah, P 2005, 'Selection of medical students in Singapore: a historical perspective', *Annals of the Academy of Medicine Singapore*, vol. 34, pp. 147C-51C.
7. Validity of the Deakin MMI

7.1 Introduction

This chapter aims to determine whether the Deakin Medical School (DMS) multiple-mini interview (MMI) represents a valid selection tool for the selection of medical students, and whether the validity of the tool is influenced by interview length.

In order for a test to be useful for selection it must be reliable and valid so as to provide an accurate measure of a pertinent trait or quality (Fruen 1983). Reliability provides an assessment of how well an instrument measures a particular quality and two important forms of reliability, internal consistency and inter-rater reliability, have been assessed for the Deakin MMI in other sections of this thesis. Validity indicates how well test performance correlates with the quality it is believed to measure (O'Brien et al. 2011). This implies that performance on tests that are valid will predict performance in real-life situations that incorporate the characteristics or qualities measure by the test. As outlined in Chapter 2, reliability and validity are inextricably linked. Validity is affected by test reliability but is also affected by variations in relationships between predictor and criterion constructs (Conway, Jako & Goodman 1995).

Selection tests for medical school are valid if they assess cognitive and non-cognitive attributes that are truly related to performance as a medical student and doctor. As outlined in Chapter 2, the ability of tests that assess cognitive attributes to predict medical school examination success has been studied extensively.
However, less is known about their ability to predict performance beyond medical school. Still less is known about the ability of tests for non-cognitive traits, to predict performance during and after the medical course.

The Deakin Medical School (DMS) selection process considers applicants’ GPA, GAMSAT score and performance on a selection MMI. GPA and GAMSAT provide information relating to the academic ability of applicants to succeed in the medical course. The Deakin MMI, comprising ten stations of eight minutes duration\(^1\) that address core DMS outcomes, attempts to identify qualities commonly considered to be desirable in doctors. These include qualities that are likely to underpin effective and professional encounters with patients and colleagues (communication skills, professionalism, health promotion, teamwork and evidence use), maintenance of professional standards (career motivation and self-directed learning), social responsibility (social justice and effective use of resources), and the desire to service areas of need in the Western Victorian region (rural awareness). Interviewers for the MMI are derived from three key stakeholder groups: health professionals, academics and community members. This approach creates an interview process with high acceptability and face validity (Grey \textit{et al.} 2001; O’Brien \textit{et al.} 2011). However, no quantitative assessment of the criterion validity of the Deakin MMI has been undertaken. This chapter aimed to assess the criterion validity of the Deakin MMI by comparing performance on the MMI with performance during the Deakin medical course.

\(^{1}\) Reduced to 5 minutes on the basis of results of Chapter 5 of this thesis.
The curriculum throughout the Deakin Medical Course is organised into four themes: *Doctor and Patient (DP)*; *Knowledge of Health and Illness (KHI)*; *Doctors, Cultures, Peoples and Institutions (DPCI)*; and, *Ethics, Law and Professional Development (ELPD)*. In broad terms, scientific knowledge is assessed within the KHI theme, public health, epidemiology and biostatistics within the DPCI theme, communication skills, clinical reasoning and procedural skills within the DP theme, whilst professional values, including legal and ethical issues, are largely assessed within the ELPD theme. It is therefore possible to map the qualities assessed at Deakin MMI stations with themes that access these qualities during the Deakin medical course (Figure 7.1).

**Figure 7.1 Qualities assessed during the Deakin MMI and their place within the Deakin Medical Course curriculum**

<table>
<thead>
<tr>
<th>Knowledge of Health &amp; Illness</th>
<th>Doctor &amp; Patient</th>
<th>Doctors, Cultures, Peoples &amp; Institutions</th>
<th>Ethics, Law &amp; Professional Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Skills</td>
<td>Health Promotion Evidence Use Effective use of resources</td>
<td>Professionalism Social justice Teamwork (Interprofessional learning)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Career motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-directed learning</td>
</tr>
<tr>
<td>Teamwork</td>
</tr>
<tr>
<td>Rural awareness</td>
</tr>
</tbody>
</table>
It was hypothesized that MMI score would predict performance in tasks encountered during the medical course that assessed the same qualities. Since the outcomes assessed at some MMI stations were to be directly assessed during the Deakin medical course (such as communication skills and teamwork), it was hypothesized that performance at specific interview stations would correlate with performance in these medical school assessment tasks. In contrast, it was expected that overall performance during the course would correlate best with career motivation, teamwork and self-directed learning as these were essential drivers for ongoing achievement throughout the course.

As most qualities assessed during the MMI fit broadly within the sphere of ‘professionalism’, ‘communication skills’ and ‘social responsibility’, it was hypothesized that performance at the MMI would correlate best with overall performance in the themes of the Deakin medical course that taught and assessed these areas and that MMI performance would correlate better in later years of the course, where there is more emphasis upon clinical and professional interactions. As the MMI assesses enduring traits that are likely to be consistently present throughout each of the MMI stations, it was also hypothesized that shortening the duration of MMI stations would not affect the predictive validity of the MMI. A final hypothesis was that performance on MMI taken in combination with performance on GPA and GAMSAT would better predict performance on the course overall than performance on the MMI alone as these three selection tools are likely to provide complementary information about the ability of students to cope with the medical course.
7.2 Methods

The performance of successful applicants on the Deakin MMI was compared to performance attained on relevant assessment tasks completed during the Deakin Medical Course.

7.2.1 Data collected

The collection of data regarding performance during the Deakin MMI occurred within the context of routine selection interviews in 2008 (outlined in Chapter 5). However, additional selection data (GPA and GAMSAT performance) were also included in this analysis. Data relating to participants’ performance on assessments completed during the Deakin medical course was extracted from the Deakin University, School of Medicine Assessment Database. Performance on assessment tasks completed during the Deakin medical course was reviewed and categorised, according to the type of information available into: (i) outcome-specific performance (for example performance on tasks specifically assessing an outcome also assessed at an MMI station such as communication skills); (ii) theme-specific performance (that included combined performance over all assessments completed within one of the four medical course themes); (iii) year-specific performance (averaged over all themes for each year of the course); and (iv) overall performance (final combined score for all themes and all year levels averaged over the course). The available intra-course results are outlined in Table 7.2.
Table 7.2 Deakin medical course results available for analysis

<table>
<thead>
<tr>
<th>Outcome specific</th>
<th>Theme specific</th>
<th>Year specific</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication skills</td>
<td>KHI</td>
<td>Year 1</td>
<td>Final</td>
</tr>
<tr>
<td>Teamwork</td>
<td>DP</td>
<td>Year 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DPCI</td>
<td>Year 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ELPD</td>
<td>Year 4</td>
<td></td>
</tr>
</tbody>
</table>

Performance during the Deakin medical course is assessed by a variety of means, including written short answer and multiple choice question examinations, written assignments, rater-based assessments of performance in OSCE and other patient encounters, and presentations. All assessment, are rigorously evaluated in terms of their reliability and efforts are made to maximise validity by crosslinking assessment content with course learning objectives. Much of the learning during the course is self-directed and requires students to regularly work together in groups. Assessments specifically assessing communication skills consisted of global score from OSCE and communication skills modules. Teamwork was specifically assessed within the Inter-professional Learning component of the ELPD theme.

7.2.2 Participants

Final year students of the Deakin medical course in 2011 were invited to participate in the project, since they represented the same cohort of students that had provided data for the study described in Chapter 5. All 105 eligible students (ie those that had participated in the 2008 MMI, and had progressed through the course) were invited to participate in the study. Full participation would have provided sufficient power to discriminate correlations between MMI and course performance that were
in the order of 0.2. However, the 33 (29%) that consented to this analysis instead gave power to discriminate moderate to strong correlations (above approximately 0.4) between MMI and course performance.

7.2.3 Statistical Analyses

Pearson correlation coefficients were used to compare performance on individual MMI stations, the MMI as a whole, and the MMI in combination with GPA and/or GAMSAT score with performance during the Deakin medical course. p values and 95% confidence intervals were calculated using standard formulae.

As the students admitted into the Deakin medical course were explicitly selected on the basis of GPA, GAMSAT and MMI performance, the students participating in the study are likely to be at the top end of the potential range of scores for previous academic performance and are also likely to do well in their medical school training. In order to adjust for attenuation associated with restriction of range, correlation coefficients were corrected for restriction of range. As participants’ performance on selection predictor variables (GPA, GAMSAT score and MMI score), the variance of each predictor variable in the total population of applicants, the variances of the predictor and course success variables, and the unadjusted correlation between the two variables in the restricted range population were known, the standard formula for explicit selection, first derived by Pearson (Pearson 1903) was used to correct for restriction of range (Held & Foley 1994):

\[
R_{xyr} = \frac{r_{xy} \left(S_y/s_x \right)}{\left[1 - r_{xy}^2 + r_{xy}^2 \left(S_y/s_x \right)^2 \right]^{1/2}}
\]
where $R_{xy}$ is the corrected correlation coefficient, $S_x$ is the standard deviation of the predictor variable in the unrestricted pool of applicants, $s_x$ is the standard deviation of the predictor in the restricted sample, and $r_{xy}$ is the observed correlation between the two variables within the restricted population. The strength of the relationships between predictor and outcome variables following appropriate correction were categorised as weak ($\leq 0.35$), moderate ($0.36 - 0.67$), or strong ($\geq 0.68$) (Taylor 1990).

Unreliability associated with selection tools and medical school assessments may further limit the size of the correlations between predictors and outcomes (Ferguson, James & Madeley 2002). Spearman provided the following equation to correct for unreliability in calculation of the correlation coefficient (Spearman 1904):

$$\rho_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$$

Where $\rho_{xy}$ is the corrected validity coefficient, $r_{xy}$ is the obtained validity coefficient, $r_{xx}$ is the reliability of the predictor, and $r_{yy}$ is the reliability of the outcome measure.

Correcting in this way removes the potential diluting effects of measurement error upon the correlation coefficient and provides a more accurate picture of the "true" relationship between the predictor variable and outcome in the population (Nunnally 1967).
However, selection tests and medical course assessments are not 100% reliable. Correcting correlation coefficients for test unreliability may also be seen to artificially inflate the ‘real-world’ correlation coefficient between a predictor variable (as measured by an imperfect tool) and an outcome because it ignores a key limitation (unreliability) of the test itself. This may in turn have the effect of falsely increasing confidence in the ability of the tool to predict the outcome (Murchinsky 1996). For this reason, coefficients corrected for only restriction of range (and not unreliability) were included in the final analysis and used assess the validity of the MMI within the context of the Deakin medical course.

As \( r \) distributions are non-normal, confidence intervals were calculated in a three-step process whereby correlation coefficients were first converted to \( Z \) values by using Fisher’s \( r \) to \( Z \) transformation.

\[
z = \frac{1}{2} \log_e \left( \frac{1+r}{1-r} \right)
\]

This allowed the calculation of upper and lower confidence interval limits for each \( z \) value using standard formulae. The confidence interval limits were then converted back to \( r \) values by solving Fisher’s \( r \) to \( Z \) transformation for \( r \) to give the following formula:

\[
r = \frac{(e^{2z}-1)}{(1+e^{2z})}
\]

For moderate and strong correlations (\( r > 0.35 \)), the degree of variability explained by the predictor variable (sample coefficient of determination) was then calculated by squaring the adjusted correlation coefficient.
7.3 Results

7.3.1 Participant characteristics

Thirty-three of the eligible participants returned consent forms (response rate 29%) and were included in the analyses. The mean characteristics of this subset compared to those of the total eligible population are shown in Table 2. There were no significant differences ($p \geq 0.05$ in all cases; two-tailed unpaired t test) between the GPA, GAMSAT score and MMI scores for the participant group compared to those of the total eligible population, suggesting that the characteristics of participant group were representative of the total eligible population

| Table 7.3 Predictor characteristics for participant group and total eligible population |
|---------------------------------|---------------------------------|
|                                | Participants mean value (standard deviation) | Total eligible population mean values (standard deviation) |
| MMI Overall Score              | 78.85 (6.54)                      | 78.59 (7.09)                      |
| MMI Station 1                  | 4.00 (1.03)                       | 3.75 (0.96)                       |
| MMI Station 2                  | 3.70 (0.77)                       | 3.90 (0.87)                       |
| MMI Station 3                  | 3.85 (0.80)                       | 3.93 (0.80)                       |
| MMI Station 4                  | 4.18 (0.64)                       | 4.08 (0.83)                       |
| MMI Station 5                  | 4.24 (0.83)                       | 4.09 (0.79)                       |
| MMI Station 6                  | 4.00 (0.75)                       | 4.06 (0.82)                       |
| MMI Station 7                  | 3.91 (0.91)                       | 3.84 (0.80)                       |
| MMI Station 8                  | 3.76 (0.76)                       | 3.89 (0.85)                       |
| MMI Station 9                  | 3.97 (0.73)                       | 4.02 (0.79)                       |
| MMI Station 10                 | 3.82 (0.73)                       | 3.77 (0.81)                       |
| GAMSAT Overall Score           | 60.30 (5.10)                      | 60.10 (4.81)                      |
| GAMSAT Section 1               | 60.45 (5.33)                      | 58.56 (5.05)                      |
| GAMSAT Section 2               | 62.03 (7.14)                      | 62.48 (6.66)                      |
| GAMSAT Section 3               | 59.36 (6.48)                      | 59.70 (7.28)                      |
| GPA                            | 87.63 (5.68)                      | 88.11 (6.07)                      |

* $p < 0.05$
Table 7.4 shows the mean and standard deviations achieved by the participant group for each theme, year of the course and the course as a whole.

**Table 7.4 Mean and standard deviation for course themes, years and overall course mark**

<table>
<thead>
<tr>
<th></th>
<th>KHI</th>
<th>DP</th>
<th>DPCI</th>
<th>ELDP</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>67.06</td>
<td>67.22</td>
<td>72.47</td>
<td>68.77</td>
<td>70.83</td>
<td>74.31</td>
<td>76.12</td>
<td>72.43</td>
<td>73.65</td>
</tr>
<tr>
<td>SD</td>
<td>4.20</td>
<td>3.03</td>
<td>5.57</td>
<td>3.33</td>
<td>4.08</td>
<td>3.70</td>
<td>3.25</td>
<td>4.53</td>
<td>5.34</td>
</tr>
</tbody>
</table>

7.3.2 Analysis of Outcome-Specific Performance

Correlations were investigated between two specific outcomes (communication skills and teamwork) that were assessed both in the MMI and subsequently in the Deakin medical course (Table 7.5). No significant associations ($p \geq 0.05$) were observed between performance at the communication skills and teamwork MMI stations and performance on tasks designed to assess these qualities during the medical course.

**Table 7.5 Correlation coefficients (adjusted) between performance at the communication skills and teamwork MMI stations and performance on tasks designed to assess these qualities during the medical course**

<table>
<thead>
<tr>
<th>MMI Station</th>
<th>Unadjusted correlation coefficient</th>
<th>Adjusted correlation coefficient</th>
<th>Variability in course assessment explained by performance on MMI station</th>
<th>PPV/NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Skills</td>
<td>0.05</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork</td>
<td>0.18</td>
<td>0.20</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

7.3.3 Ability of MMI station performance to predict course performance
Correlations were also investigated between performance at specific MMI stations and performance in each year of the medical course and the course as a whole (Table 7.6). Although no statistically significant correlations were observed, performance at the teamwork (r = 0.18, 95% confidence interval -0.18 - 0.49) and career motivation stations (r = 0.24, 95% confidence interval -0.11 - 0.61) appeared to correlate best with course performance.

Table 7.6 Correlation coefficients (adjusted) between performance at MMI stations and performance in the Deakin medical course.

<table>
<thead>
<tr>
<th>MMI Station</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Skills</td>
<td>-0.31</td>
<td>0.29</td>
<td>-0.25</td>
<td>0.12</td>
<td>-0.25</td>
</tr>
<tr>
<td>Evidence Use</td>
<td>-0.16</td>
<td>0.05</td>
<td>-0.24</td>
<td>-0.22</td>
<td>0.03</td>
</tr>
<tr>
<td>Health Promotion</td>
<td>-0.21</td>
<td>-0.05</td>
<td>-0.32</td>
<td>-0.10</td>
<td>-0.22</td>
</tr>
<tr>
<td>Teamwork</td>
<td>-0.03</td>
<td>0.36</td>
<td>0.01</td>
<td>0.24</td>
<td>0.18</td>
</tr>
<tr>
<td>Career motivation</td>
<td>-0.12</td>
<td>0.21</td>
<td>0.18</td>
<td>0.31</td>
<td>0.24</td>
</tr>
<tr>
<td>Social Justice</td>
<td>-0.02</td>
<td>0.28</td>
<td>0.03</td>
<td>-0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Self-directed Learning</td>
<td>0.02</td>
<td>0.16</td>
<td>0.03</td>
<td>0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>Professionalism</td>
<td>-0.06</td>
<td>0.14</td>
<td>-0.15</td>
<td>0.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Resource Use</td>
<td>0.28</td>
<td>0.09</td>
<td>0.29</td>
<td>0.29</td>
<td>0.07</td>
</tr>
<tr>
<td>Rural awareness</td>
<td>-0.07</td>
<td>0.26</td>
<td>0.03</td>
<td>0.23</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

Comparison between performance on specific MMI stations and theme performance revealed a statistically significant negative association between performance at the health promotion MMI station and performance in the KHI theme (r = -0.36, p = 0.047). Other correlations that approached statistical significance were between rural awareness and DPCI performance (r = 0.31, p = 0.10), resource use and DP performance (r = 0.31, p = 0.22), and communication skills and KHI performance (r = -0.30, p = 0.07) (Table 7.7).

Table 7.7 Correlation coefficients (adjusted) between performance at MMI stations and theme performance.
stations and performance in the four themes of the Deakin medical course.

<table>
<thead>
<tr>
<th>MMI Station</th>
<th>KHI</th>
<th>DP</th>
<th>DPCI</th>
<th>ELPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication Skills</td>
<td>-0.30</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Evidence Use</td>
<td>-0.04</td>
<td>-0.25</td>
<td>-0.06</td>
<td>0.11</td>
</tr>
<tr>
<td>Health Promotion</td>
<td>-0.36*</td>
<td>-0.02</td>
<td>-0.22</td>
<td>0.12</td>
</tr>
<tr>
<td>Teamwork</td>
<td>0.24</td>
<td>0.17</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>Career motivation</td>
<td>-0.11</td>
<td>0.27</td>
<td>-0.01</td>
<td>0.25</td>
</tr>
<tr>
<td>Social Justice</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Self-directed Learning</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
<td>-0.06</td>
</tr>
<tr>
<td>Professionalism</td>
<td>-0.18</td>
<td>-0.02</td>
<td>0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>Resource Use</td>
<td>0.19</td>
<td>0.31</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Rural awareness</td>
<td>-0.15</td>
<td>0.16</td>
<td>0.33</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

* p < 0.05

7.3.4 Ability of total MMI performance to predict course performance

Performance at the MMI provided little information about overall performance during the medical course. However, a significant correlation was observed between MMI performance and performance during the second year of the course (r = 0.45, p = 0.01) (Table 7.8a), indicating that MMI performance may account for approximately 20% of variation in second year results.

No significant correlations were observed between total MMI score and performance in the four themes of the Deakin medical course. Performance on the MMI appeared to best predict performance in the DP theme (r = 0.21, p = 0.27), and negatively predict performance in the KHI theme (r = -0.31, p = 0.10) (Table 7.8b).
Table 7.8 Correlation coefficients (adjusted) between performance on the MMI and performance during the Deakin medical course

a) yearly and overall performance

<table>
<thead>
<tr>
<th></th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI (TOTAL SCORE)</td>
<td>-0.21</td>
<td>0.45*</td>
<td>-0.13</td>
<td>0.24</td>
<td>0.10</td>
</tr>
</tbody>
</table>

b) performance in the four themes of the Deakin medical course

<table>
<thead>
<tr>
<th></th>
<th>KHI</th>
<th>DP</th>
<th>CLINICAL</th>
<th>DPCI</th>
<th>ELPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI (TOTAL SCORE)</td>
<td>-0.31</td>
<td>0.21</td>
<td>-0.04</td>
<td>0.05</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

* p < 0.05

7.3.5 Ability of composite measures (MMI in combination with GAMSAT and/or GPA) to predict course performance

Correlations were investigated between scoring on the three selection tools utilised by Deakin Medical School and performance during the medical course (Tables 7.9 & 7.10). No significant correlations were observed for any individual selection tool. Of the three selection tools, GPA appeared to best predict overall performance during the medical course (r = 0.29, p = 0.12) and was the most consistent predictor of performance across years and themes of the medical course (Tables 7.9 & 7.10). MMI score predicted performance in the second year of the course (r = 0.45, p = 0.01), while GAMSAT score did not appear to provide useful information about how well applicants would perform during the medical course (Table 7.9). Combining GPA with MMI score produced improved predictive validity for overall performance, yielding a correlation coefficient of 0.49 that bordered on significance (p = 0.07) and suggesting that 25% of variation in performance during the medical course may be accounted for by the combination of GPA and MMI score. GPA in combination with MMI score was also found to be particularly useful in predicting
performance in the second and final years of the medical course (Table 7.9),
accounting for 40% and 36% of variation in performance respectively, and in
predicting performance in the DP theme (Table 7.10), accounting for 25% of
variation in performance. The addition of GAMSAT performance to GPA and/or
MMI did not appear to improve the predictive validity of the selection process.

Table 7.9 Correlation coefficients (adjusted) between GPA, MMI and
GAMSAT performance, and performance (yearly and overall) during the
Deakin medical course

<table>
<thead>
<tr>
<th>Selection Tool</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI TOTAL</td>
<td>-0.21</td>
<td>0.45*</td>
<td>-0.13</td>
<td>0.24</td>
<td>0.10</td>
</tr>
<tr>
<td>GAMSAT TOTAL</td>
<td>0.10</td>
<td>-0.21</td>
<td>-0.08</td>
<td>-0.27</td>
<td>-0.13</td>
</tr>
<tr>
<td>GPA</td>
<td>0.25</td>
<td>0.05</td>
<td>0.28</td>
<td>0.25</td>
<td>0.29</td>
</tr>
<tr>
<td>MMI + GPA</td>
<td>0.01</td>
<td>0.64*</td>
<td>0.18</td>
<td>0.60*</td>
<td>0.49</td>
</tr>
<tr>
<td>MMI + GAMSAT</td>
<td>-0.13</td>
<td>0.29</td>
<td>-0.22</td>
<td>0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td>GPA + GAMSAT</td>
<td>0.30</td>
<td>-0.14</td>
<td>0.17</td>
<td>-0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>MMI + GPA + GAMSAT</td>
<td>-0.13</td>
<td>0.46</td>
<td>0.08</td>
<td>0.34</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table 7.10 Correlation coefficients (adjusted) between GPA, MMI and
GAMSAT performance, and performance in the four themes of the Deakin
medical course

<table>
<thead>
<tr>
<th>Selection Tool</th>
<th>KHI</th>
<th>DP</th>
<th>DPCI</th>
<th>ELPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI TOTAL</td>
<td>-0.31</td>
<td>0.21</td>
<td>0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td>GAMSAT TOTAL</td>
<td>0.23</td>
<td>-0.24</td>
<td>0.00</td>
<td>-0.24</td>
</tr>
<tr>
<td>GPA</td>
<td>0.27</td>
<td>0.16</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>MMI + GAMSAT</td>
<td>-0.12</td>
<td>0.00</td>
<td>0.09</td>
<td>0.31</td>
</tr>
<tr>
<td>MMI + GPA</td>
<td>-0.12</td>
<td>0.49*</td>
<td>0.26</td>
<td>0.29</td>
</tr>
<tr>
<td>GPA + GAMSAT</td>
<td>0.42</td>
<td>-0.07</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td>MMI + GPA + GAMSAT</td>
<td>0.20</td>
<td>0.23</td>
<td>0.39</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

* p < 0.05,
7.3.6 Effect of shortening station duration on validity

To examine the effects of shortening the MMI stations, five-minute data was available from odd-numbered stations for 13 participants, and from even-numbered stations from 20 participants. Substituting total 5-minute score for 8-minute score produced correlation coefficients between MMI alone and in combination with other selection tools, that were broadly similar to those obtained using 8-minute scores (Tables 7.11 and 7.12). However, the meaningfulness of these observations were limited by the low number of data points which resulted in wide confidence intervals.

Table 7.11 Comparison of correlation coefficients (adjusted) for MMI using 8-minute and 5-minute scores alone and in combination with GPA and GAMSAT, and performance in each year of the Deakin course and the course overall

<table>
<thead>
<tr>
<th>Selection Tool</th>
<th>MMI station duration</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI TOTAL</td>
<td>8 min</td>
<td>-0.21</td>
<td>0.45*</td>
<td>-0.13</td>
<td>0.24</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.14</td>
<td>0.86*</td>
<td>0.28</td>
<td>0.51</td>
<td>0.72</td>
</tr>
<tr>
<td>MMI + GAMSAT</td>
<td>8 min</td>
<td>-0.13</td>
<td>0.29</td>
<td>-0.22</td>
<td>0.01</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.30</td>
<td>-0.19</td>
<td>0.17</td>
<td>-0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>MMI + GPA</td>
<td>8 min</td>
<td>0.01</td>
<td>0.64*</td>
<td>0.18</td>
<td>0.60*</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.52</td>
<td>0.50</td>
<td>0.53</td>
<td>0.53</td>
<td>0.65*</td>
</tr>
<tr>
<td>MMI + GPA + GAMSAT</td>
<td>8 min</td>
<td>-0.13</td>
<td>0.46</td>
<td>0.08</td>
<td>0.34</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.55</td>
<td>0.19</td>
<td>0.35</td>
<td>0.07</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Table 7.12 Comparison of correlation coefficients (adjusted) for MMI using 8-minute and 5-minute scores alone and in combination with GPA and GAMSAT, and performance in the four themes of the Deakin medical course

<table>
<thead>
<tr>
<th>Selection Tool</th>
<th>MMI station duration</th>
<th>KHI</th>
<th>DP</th>
<th>DPCI</th>
<th>ELPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMI TOTAL</td>
<td>8 min</td>
<td>-0.31</td>
<td>0.21</td>
<td>0.05</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>-0.08</td>
<td>0.20</td>
<td>0.25</td>
<td>0.03</td>
</tr>
<tr>
<td>MMI + GAMSAT</td>
<td>8 min</td>
<td>-0.12</td>
<td>0.00</td>
<td>0.09</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.04</td>
<td>0.06</td>
<td>0.22</td>
<td>-0.09</td>
</tr>
<tr>
<td>MMI + GPA</td>
<td>8 min</td>
<td>-0.12</td>
<td>0.49*</td>
<td>0.26</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.10</td>
<td>0.28</td>
<td>0.36</td>
<td>0.34</td>
</tr>
<tr>
<td>MMI + GPA + GAMSAT</td>
<td>8 min</td>
<td>0.20</td>
<td>0.23</td>
<td>0.39</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>5 min</td>
<td>0.15</td>
<td>0.12</td>
<td>0.29</td>
<td>0.02</td>
</tr>
</tbody>
</table>

7.4 Discussion

There were several key findings to be drawn from this chapter. Firstly, the relationships between individual MMI stations and performance during the Deakin medical course were generally absent or weak. However, a moderate negative correlation was observed between performance at the Health Promotion MMI station and in the KHI theme of the medical course. Performance at the Teamwork and Career Motivation MMI stations appeared to best predict overall performance during the Deakin medical course, although in both cases the association was weak and not statistically significant. Secondly, of the three key selection tools used during the selection process for entry into the Deakin medical course, GPA was the best predictor overall. The MMI provided limited information about course performance, however, a moderate and statistically significant correlation between MMI performance and performance during the second year of the course was observed. Of note was that performance on the GAMSAT did not appear to provide useful information regarding performance during the Deakin medical course.
Thirdly, combining the results of multiple selection tools was found to be beneficial in predicting performance during the Deakin medical course. The combination of GPA and MMI score was the best predictor of course performance. Finally, the predictive validity of the MMI does not appear to be decreased by reducing MMI stations from 8 to 5 minutes.

The hypothesis that that MMI score would predict performance in tasks encountered during the medical course that assessed the same qualities was not supported. Although overall performance during the course did appear to correlate best with performance at the Career Motivation and Teamwork stations, these correlations were weak and not statistically significant. It may be that these qualities are essential drivers for ongoing achievement throughout the course.

The hypothesis that performance in the MMI would correlate better with performance in the ELPD, DCI and DP themes than the KHI theme was partially supported. However, relationships between MMI performance and all themes were weak or absent and not statistically significant. This could be because many of the qualities assessed in the MMI were required (although not always formally assessed) in all themes of the Deakin medical course. For example, all themes required significant amounts of group work, in the form of group projects and presentations and problem-based learning (PBL) tutorials. The hypothesis that the relationship between MMI performance and course performance would improve in later years of the course was not supported. Rather, MMI performance best predicted performance during the second year of the course, accounting for 20% of variation in second year mark. There are several potential reasons for this. For
example, second year represents the year of the medical course when there is maximal convergence of clinical and basic sciences and may therefore be the part of the course most closely related to MMI performance. Second year is also the point at which students engage in their first clinical and professional encounters. Therefore it is possible that second year represents a window where clinical performance during the course reflects natural abilities, prior to equalization of skills through ongoing exposure and teaching in the medical course.

It was also hypothesized that MMI performance taken in combination with GPA and GAMSAT performance would be a better predict performance on the course overall than performance on the MMI alone. This was the case for GPA and MMI, suggesting that these two tools may provide useful and complementary information about how applicants are likely to cope with the medical course. In contrast, GAMSAT did not appear to provide useful information either alone or in combination with other selection tools. This finding was unexpected as authors have consistently reported positive correlation coefficients in the order of 0.2-0.3 (eg (Donnelly 2006; Groves, Gordon & Ryan 2007; Wilkinson et al. 2008). The reason for this discrepancy is unclear, although it is possible that this may be at least partially explained by characteristics of the participant group. Indeed, previous observations have found that GAMSAT results may vary significantly with candidate gender, age and highest degree level (Aldous et al. 1997; Oates & Goulston 2012). Aldous and colleagues found that male applicants scored more highly on GAMSAT in reasoning in the humanities and reasoning in the sciences than female medical students, older students perform less well in reasoning in the sciences but better in written communication occurred with age, and honours
students recorded better overall GAMSAT scores than those holding only Bachelors degrees. The type of first degree completed was also found to have a strong influence on GAMSAT performance. Those with a background in arts and social science performed well with reasoning in the GAMSAT humanities test and written communication, whilst those with science degrees performed better in reasoning in the GAMSAT physical sciences test. The participant group included a low proportion of males (approximately one third of participants) and included individuals with a diversity of backgrounds and ages ranging between 20 and 42. These factors may have served to dilute any relationship between GAMSAT score and course performance. It is also possible that factors intrinsic to the Deakin course and its assessment methods, such as the early introduction of clinical encounters and the emphasis on professionalism and clinical skills training from early in the course. These components become even more important in the final years of the course and may explain the observation that, in this thesis, as in previous studies, GAMSAT appeared to correlate better with performance early in the course than with performance in later years of the course.

The hypothesis that shortening the duration of MMI stations from 8 to 5 minutes would not affect the predictive validity of the MMI, draws some support from this study. Due to the small number of data points available or 5-minute stations (20 data points for odd-numbered stations and 13 for even-numbered stations) no meaningful data was gathered in relation to the ability of individual stations to predict performance during the medical course. However, total 5-minute score alone and in combination with other selection tools produced correlation coefficients broadly consistent with those observed for 8-minute scores. In both
cases, MMI in combination with GPA score was the best predictor of course performance. This was an expected finding due to the likelihood that characteristics demonstrated during the interview were considered to reflect inherent personal qualities likely to have been apparent throughout the interview, an observation supported by the findings of Chapter 5 in which interview scores were shown to vary little in the last 3 minutes of an 8 minute interview.

The research outlined in this chapter is characterised by a number of limitations. Firstly, the study findings are based upon a small sample size that resulted in a situation where sufficient power was present to detect only moderate to strong correlations. Although unintended, this was not unexpected as the participant response rate from the eligible population of 29% was consistent with other studies relying upon a pool of graduating medical practitioners. Invitations and consents for participation in the study were sent out to eligible individuals in the final year of the medical course, a time at which many students were on clinical rotations, electives and/or focusing on final exams. Only, 22 consent forms were returned following the original mail out. Further efforts to recruit more participants over a period of 6 months using email, social media and reminders at clinical schools increased this to 33. Although the poor response rate reduced the power of the study, sufficient power remained to detect the more meaningful correlations (r >0.35). Even if there had been sufficient power to detect smaller correlations, the usefulness of these findings would have been questionable as such correlations would have accounted for 10% or less of variation in the predictor variable.
Secondly, there is a lack of consensus regarding the best way to correct for correlation coefficients. Therefore, the coefficients reported here should be considered within context. In the literature, some coefficients are reported uncorrected, others corrected for restriction of range only, and others corrected for both restriction of range and unreliability. The decision was made in this study to correct correlation coefficients for restriction of range but not for unreliability. It was felt that this would provide the most meaningful correlation coefficient that was representative of a cohort applying for admission into the medical course, but also took into account ‘real-world’ limitations associated with admission and course tests such as imperfect reliability.

Other limitations of the study relate to its external validity or generalisability. The data were obtained from a single cohort, incorporating only three sets of potential MMI station scenarios. The MMI stations utilised in the study were considered representative of the Deakin type of scenarios, format and scoring process for the assessment of each outcome or quality, and the assessments encountered during the course were standard for the Deakin medical course. Similarly, the outcome variables assessed in this study (such as year and theme scores) are heavily embedded within the Deakin medical course philosophy, structure, and assessment processes. Therefore it is also important to consider the potential impact of the course and its assessments on the observed validity data. This is particularly important as the validity of many of the assessments used to assess non-cognitive qualities during the course have themselves not been rigorously evaluated in terms of their own construct and criterion validity.
With these limitations in mind, it is possible that the relationships observed here may not be readily extrapolated to other medical schools with medical selection interview processes and/or medical courses that are structured differently, assessed differently or geared towards different goals. Further studies that incorporate larger sample sizes, different scenarios, and different yearly cohorts will help overcome these limitations.

Notwithstanding the above limitations, the findings of this chapter provide valuable information regarding the usefulness of the MMI. Individual MMI stations appear limited in their ability to predict performance in specific elements of the medical course. Therefore, caution should be taken when attempting to draw meaningful information from performance at individual stations. In contrast, performance across all stations of the MMI does provide useful information about performance during the medical course but is most valuable in predicting performance when used together with GPA. Shortening the duration of MMI stations does not appear to reduce the predictive validity of the MMI. Finally, GAMSAT may not be a useful adjunct in providing information about student performance in the medical course and may be better used as a threshold measure rather than a component for determining ranking.
References


8. Discussion & directions for further research

This thesis has provided an historical account of the approaches used for the selection of medical students throughout the ages as well as a detailed overview of the selection tools commonly employed by medical schools around the world at the beginning of the 21st century with a focus on Australian medical schools. The value of these tools in providing useful and reliable information about candidates’ ability to cope with the medical course, and perform well as medical students and doctors was then assessed in the context of the Deakin Medical School.

Throughout history, a variety of methods have been used to assess the suitability of candidates for a career in medicine. Selection processes have commonly considered cognitive ability and non-cognitive qualities. For much of history, the assessment of these factors was the responsibility of a potential mentor, prior to the entry into a master-apprentice type relationship. However, as medical training became integrated into universities, the responsibility shifted to the institution and it became necessary to use selection tools that were more objective and defensible. For the assessment of cognitive ability, this was readily achieved through the use of records of past academic performance and dedicated entrance exams. Measures of past academic performance, such as GPA, and specific entrance examinations, such as the GAMSAT, are now widely employed to assess candidates’ cognitive ability and broadly appear to be reliable and to provide useful information about performance, particularly early, in the medical course. The usefulness of these tools in predicting performance later in the course, or during the early years as a doctor, is less clear.
An objective assessment of non-cognitive qualities has proved more difficult. A variety of methods, including aptitude tests, personality tests, personal biographies and interviews have been used to assess the non-cognitive qualities of candidates. However, the objective evaluation of non-cognitive traits remains a problematic area due to limited research, poor reliability and questionable validity surrounding tools such as interviews that have been used to assess these qualities. The MMI has been identified as a promising, but under-researched tool that may provide a means of assessing non-cognitive qualities in a reliable and valid way.

Deakin Medical School utilises a 10 station MMI, specifically designed to assess the 10 key outcomes of the Deakin medical course, with each station lasting 8 minutes with 2-minutes preparation time between stations. Each interview station is manned by a trained interviewer derived from one of three subgroups; clinicians, academics and community members. The work described in this thesis specifically addressed 8 hypotheses that assessed the reliability and validity of the Deakin MMI.

This work has yielded the following key findings:

- The Deakin MMI is a reliable tool for the selection of medical students.
- The inter-rater reliability (IRR) of the Deakin MMI, when manned by a composite of clinicians, academics and community members is acceptable and not influenced by interviewer type.
- Current students represent an additional subgroup of interviewers, characterised by inter-rater reliability that is at least comparable that of standard interviewers.
• The Deakin MMI has some predictive validity as a selection tool and is particularly useful in predicting performance during the second year of the medical course.

• Of the selection tools employed by Deakin medical school, MMI used in combination with GPA is the best predictor of success during the medical course.

• Shortening the duration of MMI stations from 8 to 5-minutes did not adversely affect the reliability or validity of the tool.

The use of MMI in student selection is a resource intensive process. This thesis has identified four practical methods for improving the efficiency and sustainability of the Deakin MMI process. Firstly, the current practices of using only interviewers from three subgroups (clinicians, academics and community members) is not necessary. Secondly, current medical students represent an additional pool of reliable interviewers who can be called upon as interviewers. Thirdly, the practice of ‘matching’ interviewers to stations is not strictly necessary to ensure reliable assessment of applicants, although it may be beneficial in terms of maximising face validity and interviewer comfort. Finally, MMI stations may be shortened from 8 to 5-minutes without adversely affecting the reliability of validity of the process. For the Deakin MMI, which characteristically involve approximately 200 applicants (20 MMI, representing 20 rounds of 10 applicants), this equates to a saving of 10 hours (20*30 minutes), reducing the process from nearly 5 days to just over 3 days.

Although this thesis provides valuable information about the reliability and validity of the MMI, further research is needed. The research presented here is characterised
by several limitations. These have been discussed in detail in the relevant sections of this thesis but include three points of particular significance. Firstly, the experiments relating to the internal consistency and validity of the MMI (Chapter 5 and 7) relied upon data extracted from only one cohort of students and one MMI. This may limit the generalisability of the study to other MMIs and Deakin student cohorts. Secondly, the qualities assessed by the Deakin MMI, the way the MMI is structured, and the way interviewers are trained may all differ from those assessed elsewhere and therefore the findings of this thesis may not be readily generalizable to MMI employed by different medical schools. Thirdly, the experiment assessing the validity of the Deakin MMI (Chapter 7) was characterised by low participant numbers that limited the ability to identify weak correlations between MMI performance and performance during the medical course.

Further studies that include data from multiple student cohorts will help overcome these limitations by providing information across different years’ MMIs and different groups of applicants. This will also increase the eligible population of participants and therefore potentially increase the number of participants in the study. As studies conducted by other medical schools provide information about the reliability and validity of their MMIs, the generalizability of the findings of this thesis will become clearer.

This thesis has also provided several avenues for further study. This thesis indicates that MMI stations of 5-minutes produce an MMI that is just reliable as an MMI composed of 8-minute stations. Given the resource intensive nature of the MMI, it may be useful to determine the shortest MMI stations possible to obtain reliable
results. Moreover, since other studies have shown that 8 stations are required to produce an MMI of acceptable reliability (Axelson & Kreiter 2009; Eva, Reiter, et al. 2004) it could also be worthwhile to assess whether reliability is affected by station type. These approaches may enable additional increases in the efficiency of the MMI process, enabling further conservation of resources. Any such research would, however, need to consider the potential sequelae of such changes. For example, interviews that are perceived as ‘too short’ may decrease face validity and the exclusion of stations that assess qualities considered important to a medical school may decrease the meaningfulness of the process.

Interviewer training is a standard component of preparation for the MMI. During interviewer training, large groups of interviewers are asked to view two recorded interviews conducted at typical MMI stations. After the first interview has been viewed and scored, interviewers are provided the opportunity to discuss their scores before the second interview is viewed. The spread of interviewer scores following viewing of the second interview is almost always less than that of the first (data not shown). Although considered necessary to ensure consistency across the pool of interviews, the true value of interviewer training has not been assessed in this thesis or elsewhere. One finding of this thesis was that experience (either as interviewer or interviewee) at a particular MMI station improves IRR at that station. This implies that specifically training each interviewer at their allocated station may improve the reliability of the MMI. However, providing personalised training to interviewers would be a resource intensive process and further investigation is required to clarify the value of such specific training compared to the generic training currently delivered to all interviewers.
MMI performance at individual stations also requires further investigation. This thesis failed to find support for the use of individual station scores in predicting performance in tasks assessing similar qualities that were encountered during the medical course. Further study is required to assess the content validity of the MMI stations. This should ideally involve a comparison of the performance of applicants at MMI stations with performance on validated measures of the qualities assessed. This may be difficult, however, as validated tools are not readily available for the qualities measured in the MMI.

Data relating to the usefulness of the MMI in predicting performance beyond the medical course remains limited and provides a fertile area for further research. Although not possible within the timeframe of this thesis, the data presented here provides a useful base upon which to add information relating to performance as young doctors. Such information may include, for example, performance ratings by supervisors, patients and/or peers, as well as future success on specialty training examinations. It is also possible that performance on the MMI or certain stations on the MMI may also predict vocational choices within medicine. For example, high performance on the Rural Awareness station may be associated with a career in rural medical practice, high performance on the Health Promotion station may be associated with a career in public health and a high score on the Career Motivation station may help identify those individuals committed to a long career in medicine.

Finally, this thesis has compared the usefulness of the MMI, GAMSAT and GPA in predicting course performance. The combination of MMI score and GPA provided particularly useful information about performance during the medical course,
suggesting that the two selection tools provide complementary information about applicants and supporting a multifaceted approach to selection that considers both cognitive and non-cognitive qualities. Other measures such as portfolios, personal statements, referee reports and psychological tests may also contribute to the selection process. These tools require further investigation to clarify their usefulness alone, and in combination with other selection tools, for predicting success during the medical course. This is necessary in order to determine the combination of selection tools that provides the information most useful to medical schools when selecting new medical students, and the combination of selection tools that select for the best medical students, and even more importantly, the competent, caring and professional doctors.
References

水漲船高

‘When the tide rises, boats float higher’

Chinese proverb