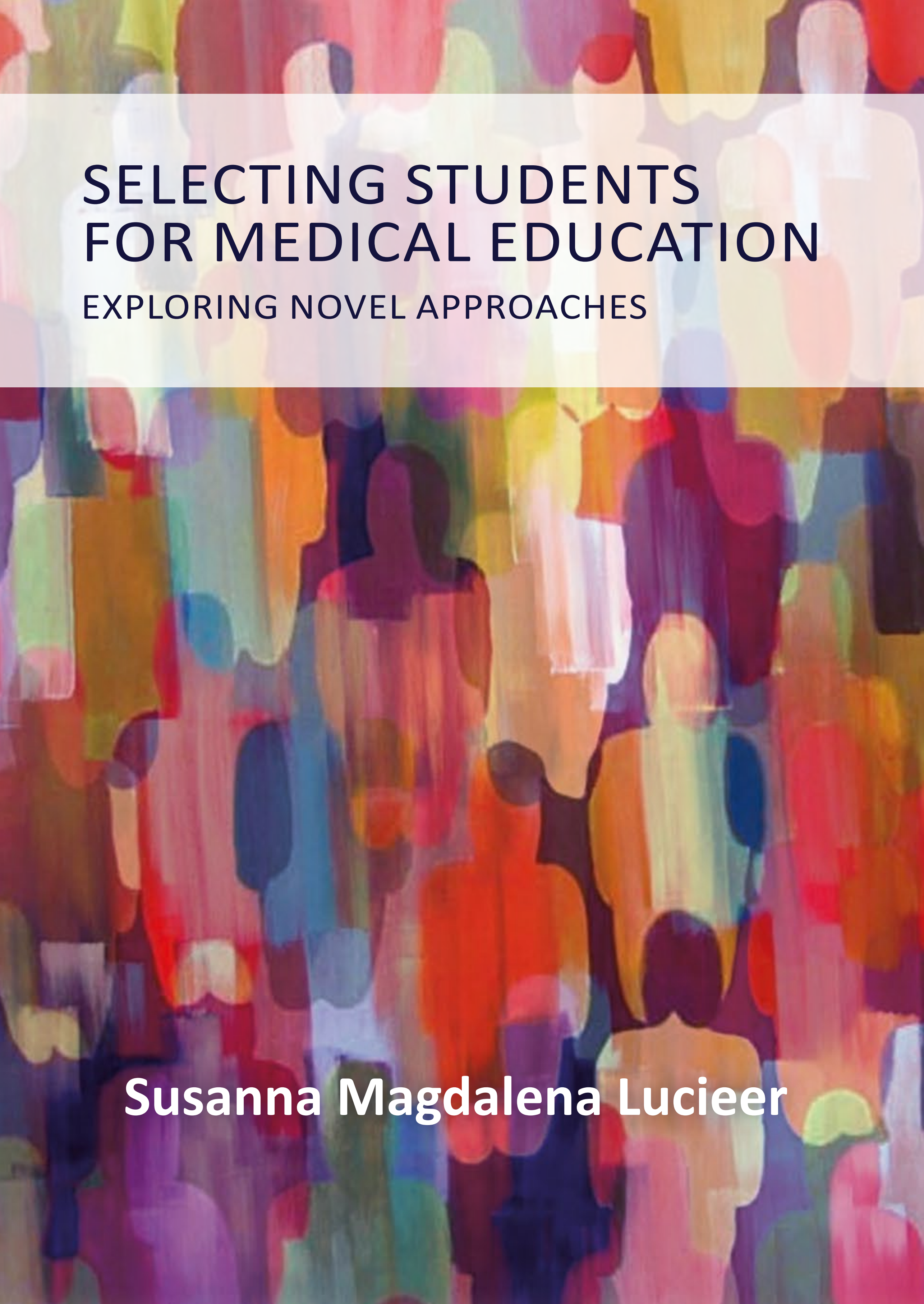
SELECTING STUDENTS FOR MEDICAL EDUCATION

EXPLORING NOVEL APPROACHES



Susanna Magdalena Lucieer

Selecting Students for Medical Education

Exploring novel approaches

Susanna Magdalena Lucieer

Susanna Magdalena Lucieer
Selecting Students for Medical Education: Exploring novel approaches
Thesis, Erasmus University Rotterdam, the Netherlands

ISBN: 978-94-6299-165-1

Cover design: Maya Krijgsman, Eigenwijs Ontwerp
Lay-out: Maya Krijgsman, Eigenwijs Ontwerp
Printing: Ridderprint BV, Ridderkerk

Selecting Students for Medical Education

Exploring novel approaches

Het selecteren van studenten voor de geneeskundeopleiding

Verkenning van nieuwe benaderingen

Proefschrift
ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College van Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 6 januari 2016 om 15.30 uur

door

Susanna Magdalena Lucieer

geboren te Rotterdam

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren	Prof.dr.ir. A.P.N. Themmen Prof.dr. R.M.J.P. Rikers
Overige leden	Prof.dr. M. de Hoog Prof.dr. G.W.C. Paas Prof.dr. G.R. Norman

Contents

GENERAL INTRODUCTION

Chapter 1

Chapter 1 General introduction

Medical schools have the responsibility of educating future doctors who are able to provide optimal healthcare (Bullimore, 1992; Hughes, 2002). However, for various reasons only a limited number of places is available to educate future doctors: 1) a high quality education needs to be ensured (Burch, 2009; Lievens, Ones, & Dilchert, 2009; Mitchell, 1990), 2) the education is expensive, and 3) the number of places to educate future doctors needs to be in tune with the eventual needs for society (Whitehouse, 1997). Since the number of applicants highly exceeds the number of places available, medical schools all over the world have to select their students. In Canada, the United States, the United Kingdom, and the Netherlands, only 1:3 – 1:17 applicants are accepted for medical school (DUO, 2015; Razack, Maguire, Hodges, & Steinert, 2012). It is important for medical schools that the applicants they do admit, will be able to successfully complete the medical programme and deliver optimal healthcare (Albanese, Snow, Skochelak, Huggett, & Farrell, 2003; Basco, Lancaster, Gilbert, Carey, & Blue, 2008; Bullimore, 1992; Burch, 2009; Kulatunga-Moruzi & Norman, 2002; Mitchell, 1990; Prideaux, et al., 2011).

The stakes of medical student selection are thus high, but the selection is a challenging and complex process. The majority of the applicants are still adolescents while medical schools aim to select applicants who are able to provide optimal healthcare at least six or more years later. Furthermore, all applicants share many characteristics, which makes it difficult to distinguish relevant characteristics in the applicants. Without a suitable preparatory training or high grades, applicants are not allowed to participate in the selection (McManus, Smithers, Partridge, Keeling, & Fleming, 2003; ten Cate, 2007). The homogeneous character of the applicant pool increases the difficulty of deciding who should be accepted, and who should be rejected for medical school. In order to demonstrate the complexity of medical schools' selection processes, this general introduction will provide an overview of the methods used in medical student selection and will discuss the methods' strengths and shortcomings. Subsequently, since the current selection methods are not optimal, concepts from outside the medical field will be discussed.

MEDICAL SCHOOL SELECTION: AN OVERVIEW

Medical student selection has a long standing history. Medical schoolsstarted to select applicants in the early twentieth century, after Abraham Flexner (Flexner, 1909) recommended to increase the quality of medical education by insisting that applicants should have a strong, scientific basis (Anderson, 2011; Irby, Cooke, & O'Brien, 2010; Krupa, 2010). These days, medical schools have created a selection policy based on desired characteristics of medical school entrants in terms of academic standards and personal qualities (Powis, 1994). These selection policies entail a wide range of selection methods that aim to measure cognitive and non-cognitive attributes, in order to identify candidates who possess the required characteristics, and to reject those who do not (Powis, 1994).

Methods that are frequently used to examine the cognitive capabilities of medical school applicants include the pre-university (secondary school) and undergraduate grade point average (GPA; Cohen-Schotanus, et al., 2006; Ferguson, James, & Madeley, 2002) and aptitude tests, such as the Medical College Admission Test (MCAT; Eskander, Shandling, & Hanson, 2013). Studies have indicated that both the pre-university GPA (pu-GPA; Baars, 2009; Cohen-Schotanus, et al., 2006; Lambe & Bristow, 2011; McManus, et al., 2005) and the undergraduate GPA (Ferguson, et al., 2002; Harnett & Willingham, 1980) provide significant predictive value for medical school performance, especially during earlier years. However, a decreasing predictive effectiveness in later years of education make the pu-GPA and the undergraduate GPA a less valuable measure for future career performance (McManus, et al., 2003; Mitchell, 1990; Salvatori, 2001). As to the MCAT, this test aims to measure the applicants' preparedness for medical school (Donnon, Paolucci, & Violato, 2007; Eskander, et al., 2013). The MCAT shows a reasonable good predictive value for academic performance in medical school (Mitchell, 1990). Combined with the undergraduate GPA, it accounts for up to 23% of the variance in scores obtained at medical school examinations (Ferguson, et al., 2002; Violato & Donnon, 2005), still leaving a large proportion of the variance unexplained.

From the 1970s onward, methods aiming to measure non-cognitive capabilities were introduced in the selection processes of medical schools since non-cognitive skills have become increasingly important for medical doctors (Benbassat & Baumal, 2007; Bullimore, 1992; Dore, et al., 2006; Eva, Rosenfeld, Reiter, & Norman, 2004; Hughes, 2002; Lievens, et al., 2009; Mahon, Henderson, & Kirch, 2013; Nieuwhof, 2004). Patients want to be involved in the decision-making process, ask questions and discuss treatments with their physicians. This requires a medical doctor to possess strong communication skills, to be empathic and have professional integrity (Haidet, et al., 2002; Pulvirenti, McMillan, & Lawn, 2014; Stewart, 2001). In addition, team-based working has become the standard for medical doctors, also increasing the importance of good communication skills (Mahon, et al., 2013). However, in comparison to selection tests that measure cognitive attributes, more controversy exists concerning the reliability and validity of selection methods that aim to measure non-cognitive attributes (Hughes, 2002). Selection methods aiming to measure the non-cognitive capabilities of applicants are, among others, interviews (Eva, Reiter, Rosenfeld, & Norman, 2004a; Harasym, Woloschuk, Mandin, & Brundin-Mather, 1996; Kreiter, Yin, Solow, & Brennan,2004; Lambe & Bristow, 2011; Salvatori, 2001), letters of motivation and recommendation (Ferguson, James, O'Hehir, Sanders, & McManus, 2003; Leichner, Eusebio-Torres, & Harper, 1981), tests that aim to measure integrity and empathy (Dore, et al., 2009; Patterson, Ashworth, et al., 2012), and questionnaires that measure an applicants' personality and motivation (Balogun, Karacoloff, & Farina, 1986; Bore, Munro, & Powis, 2009; Cremonini, 2010; Morgeson, et al., 2007; Powis, 2015).

The uncertainty concerning the reliability and validity of these non-cognitive measures arises from several sources. Although some methods look promising, most methods have many shortcomings. Concerning interviews (Kreiter, et al., 2004; Roberts, Zoanetti, & Rothnie, 2009), a large range in interviewer variability exists and personal factors such as the applicants' gender and background as well as the interviewer expectations are potential sources of bias (Eva, Reiter, et al., 2004a; Salvatori, 2001). However, there is evidence that the use of more structured interviews increases the validity (Courneya, et al., 2005). In addition, the Multiple Mini Interview (MMI) technique (Eva, Reiter, et al., 2004a; Eva, Rosenfeld, et al., 2004; Salvatori, 2001), with many short interviews, multiple examiners per applicant and training of the examiners (Bullimore, 1992; Harasym, et al., 1996) will increase the reliability (Burch, 2009; Eva, et al., 2009; Hofmeister, Lockyer, & Crutcher, 2009). Nevertheless, the ability of the MMI to predict clinical performance is unclear and further research is required (Burch, 2009; Lambe & Bristow, 2011). In addition, a limitation of interviews is that it is not practical to invite the large number of applicants who apply. To limit the number of applicants that needs to be interviewed, an extra pre-interview non-cognitive screening test could be used, such as the Computer-based Multiple Sample Evaluation of Noncognitive Skills (CMSENS). The CMSENS is also known as CASPer and shows satisfactory correlations with the MMI (Dore, et al., 2009). Motivation letters and letters of reference (Ferguson, et al., 2003; Leichner, et al., 1981) have a very low predictive validity (Ferguson, et al., 2003; Leichner, et al., 1981; Salvatori, 2001), as it turns out to be difficult to establish a decent scoring system to compare the letters (Ferguson, et al., 2002) and they rarely distinguish one applicant from another (Spina, Smith, Marciani, & Marshall, 2000). Nevertheless, the Situational Judgment Test (SJT), which aims to measure professional attributes such as professional integrity (Becker, 2005; Whetzel & McDaniel, 2009) and empathy (Lievens & Patterson, 2011; Patterson, Ashworth, et al., 2012), is promising. In an SJT, applicants are presented with descriptions of situations they might encounter during medical school or on the job. The applicants are asked to judge the appropriateness of given responses (Oostrom, Born, Serlie, & van der Molen, 2010; St-Sauveur, Girouard, & Goyette, 2014). The SJT shows good levels of reliability, predictive validity and incremental validity for integrity and empathy (Patterson, Ashworth, et al., 2012). In contract to the SJT, personality and motivation questionnaires are criticized because of their possible lack of reliability and the opportunity for applicants to respond in a socially desirable manner (Arthur, 2001). Nonetheless, some studies have shown a positive correlation between personality and academic performance (Ackerman, Chamorro-Premuzic, & Furnham, 2011; Chamorro-Premuzic & Furnham, 2003; Furnham, Chamorro-Premuzic, & McDougall, 2002; Lievens, Coetsier, de Fruyt, & de Maeseneer, 2002; Lievens, et al., 2009), and others have stated that questionnaires that determine the student's motivation could be of interest as well (Ferguson, et al., 2002; Gagne & St Pere, 2002).

However, an important critical note needs to be made about the evaluation of the predictive value of selection methods. The predictive validity of selection methods is mostly examined by investigating the relation between selection measures and grades obtained during medical school, i.e. on examination of basic science and clinical knowledge. While the relation between cognitive measures and academic performance makes sense, the relation between non-cognitive measures and medical school performance is less appropriate. A student does not need to possess non-cognitive qualities to pass the majority of the examinations, and a high predictive validity between non-cognitive qualities and academic performance should thus not be expected (Powis, 2009). Only, a better criterion to judge the predictive validity of non-cognitive measures is not easily available.

Concluding, selecting applicants with the desired characteristics for medical school appears to be challenging. Cognitive measures are able to predict up to 23% of the variance in academic performance (Ferguson, et al., 2002; Violato & Donnon, 2005), still leaving a large proportion unexplained. Additionally, the majority of the non-cognitive measures appears to lack the desired reliability and validity (Balogun, et al., 1986; Morgeson, et al., 2007; Salvatori, 2001; Siu & Reiter, 2009) and methods to measure the predictive validity should be reconsidered (Powis, 2009). Therefore, further research on selection methods that are able to select those applicants who do possess the required academic standards and personal qualities is essential.

THINKING OUTSIDE THE BOX

Obviously, the challenge of selecting the right person is not limited to the context of medical school admission. Extensive research on selection has been performed in the area of personnel assessment and organizational psychology. Here, multiple factors have been identified that relate to outstanding performance, and that can be used to select the most promising applicant for a specific job. Personality, in particular the Big Five personality trait conscientiousness, and general cognitive ability show a consistent positive relation with job performance (Schmidt, 2002; Schmitt, 2012; Tracey, Sturman, & Tews, 2007). The importance of cognitive ability even increases with job complexity, providing the highest validity for highly complex professions, including attorneys and pilots (Hunter & Hunter, 1984; Schmitt, 2012). In addition, other factors that are found to relate to general job performance are self-efficacy and self-monitoring (Bhatti, Battour, & Ismail, 2013; Holton, Bates, & Ruona, 2000; Stajkovic & Luthans, 1998). While the importance of cognitive ability is widely acknowledged in medical school selection and some researchers also suggest to use personality measures, the use of the factors self-efficacy and self-monitoring is less common. Self-efficacy and self-monitoring belong, together with planning, reflection, evaluation and effort, to a broader concept, named self-regulated learning (Boekaerts, Printrich, & Zeidner, 2000; Ertmer & Newby, 1996; Hong & O'Neil, 2001). Someone is considered to be a self-regulated learner when he or she is behaviourally, metacognitively and motivationally proactive in his or her learning process (Jonker, Elferink-Gemser, Toering, Lyons, & Visscher, 2010; Wolters, 1998; Zimmerman, 1986). Self-regulated learners take active control of their learning activities and their learning is very efficient (Brydges, Nair, Ma, Shanks, & Hatala, 2012; Ertmer & Newby, 1996; Lycke, Grøttum, & Strømsø, 2006). As a result, they update themselves more easily with all developments in their field of expertise, take more initiative to solve problems, and when needed, show more effort and persistency (Bhatti, et al., 2013; Hong & O'Neil, 2001). These qualities are important for all knowledge intensive professions (Bhatti, et al., 2013), and are, since the dynamic and extensive character of the medical field appears to be a significant hurdle to stay up to date for many practitioners (Carrier & Morin, 2014; Mooney, 1993), strongly emphasized in the medical field. The burden of the combination of patient care, administration, research and educational responsibilities often hinders the medical doctors' ability to keep abreast of the many developments in their fields of expertise, while being updated is required to provide high quality patient care (Brydges & Butler, 2012; Greveson & Spencer, 2005). Thus, one's self-efficacy and one's ability to monitor their learning are, together with other self-regulated learning activities, very valuable skills in the medical field. It might therefore be interesting to investigate the possibility to select applicants for medical school based on these characteristics.

However, it is important to note that, in contrast to general cognitive ability and personality, the ability to self-regulate ones learning is not "fixed" but develops gradually (van der Stel & Veenman, 2010, 2014). For example, young children are not able to monitor their learning process, but this skill becomes more and more sophisticated during young adulthood (van der Stel, Veenman, Deelen, & Haenen, 2010). Some researchers argue that these skills do not arise naturally but develop best in an inspiring and goal-oriented environment (Boekaerts, et al., 2000). As the majority of medical school applicants are adolescents, it is likely that their self-regulated learning skills will not be fully developed yet. Hence, before applying these measures to medical school selection procedures, the potential development of these self-regulated learning skills during medical school should be ensured.

Another captivating approach is the expert performance perspective: why do some people perform much better than others do? A considerable amount of research has been done on expert performance in several domains of complex human behaviour. Examples are: music (Ericsson, 1996), arts (Ericsson, 1996), athletics (Côté, Deakin, & Baker, 2005; Ericsson, 2013; Hodges, Kerr, Starkes, Weir, & Nananidou, 2004), chess (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005), typing (Ericsson, 2008), within insurance companies (Sonnentag & Kleine, 2000), and in a few medical areas such as medical diagnosis and surgery (Ericsson, 2004; Schmidt & Rikers, 2007). Interestingly, these studies have shown that not necessarily cognitive abilities or stable characteristics of a person, but a shared set of behaviours led to outstanding performances. Thus, a person will not excel without prolonged commitment to training and performance. This notion is central to the deliberate practice theory (Baker, 2014; Ericsson, Krampe, & Tësch-Romer, 1993; Ericsson, Nandagopal, & Roring, 2009; Roring & Charness, 2007). Deliberate practice refers to a set of highly structured activities specifically designed to improve performance (Ericsson, 2004; Ericsson, et al., 1993). This kind of practice is prolonged, very structured, often not enjoyable nor leading to immediate rewards. Instead, it takes much effort and a person needs to have long-term goals (Ericsson, 2004;Ericsson, 2008; Ericsson, et al., 1993). As a consequence, not everyone is able to engage in deliberate practice and to reach the top level. Research on the role of deliberate practice in the medical field is scarce. The expert approach attempts to describe performance under standardized conditions and representative tasks wherein superior performance can be demonstrated and reproduced (Ericsson & Smith, 1991). But since in the medical field performance is influenced by patient variability and personalized care, it is difficult to identify representative tasks that can be reproduced under standardized conditions (Ericsson, 2004). Nevertheless, in the view of medical school selection, the expert approach is interesting to explore. Whenever knowledge is available on what, or which characteristics or behaviours, makes someone perform well in the field of medicine, the next step is to identify which applicants share these characteristics and behaviours, since these applicants are probably able to reach top positions in the medical field.

Another lesson to be learned from other domains is the necessity of having a predictor-criterion relationship. Thorndike, a renowned psychologist and researcher in personnel selection, already described in 1949 that: *"The key to effective research in personnel selection and classification is an adequate measure of proficiency on the job. Only when proficiency measures can be obtained for the individuals who have been tested is it possible to check the effectiveness of test and selection procedure."* Thorndike's statement remains as true today as it was in 1949: an adequate measure of proficiency on the job is required to truly measure the effectiveness of the used selection methods (Schmitt, 2012). Unfortunately, in many fields, including medicine, no such adequate measure of proficiency exists. That is, we do not have a definition of a good medical doctor. Reasons for the absence of such a definition are that attributes of good doctors are not very well specified or agreed upon (Benbassat & Baumal, 2007; Powis, 1994). In addition, differences in perceptions of good doctors exist (Hurwitz, Kelly, Powis, Smyth, & Lewin, 2013). A possible approach could be to take patients' outcome after treatment as a quality measure. However, many factors influence a patient's outcome, such as the patient's age, gender, medical history and type of disease (McGaghie, 1990). Furthermore, not all medical doctors are involved in direct treatment of patients and since medical doctors mostly work in teams (Mahon, et al., 2013), it is difficult to determine to what extent successful patient outcome can be credited to a single medical doctor. Medical school selection committees thus face a "criterion- problem" (Schmitt, 2012); they aim to select future well-performing medical doctors, but they do not know what constitutes a well-performing medical doctor. Consequently, it is not possible to measure whether there is a relationship between the outcome of the selection methods, and performance as a medical doctor. Therefore, medical schools have chosen a pragmatic approach by selecting potentially good *students* opposed to potentially good *doctors* (Powis, 1994), despite only moderate and weak relations have been found between academic performance and performance on knowledge tests in the early stages of the professional career (Hamdy, et al., 2006) and only predictors for short term professional performance are available (Drenth, 2004; Tamblyn, et al., 2002; Taylor, Price, Richards, & Jacobsen, 1965).

In the personnel assessment and organizational psychology domain, overall job performance is measured and refers to the overall contribution of a single employee to the organization. Overall job performance includes task performance, contextual performance, and avoidance of counterproductive behaviours. As many factors influence performance, overall job performance is most optimally and most frequently assessed via supervisory ratings (Schmitt, 2012). Contrary to many other professions, medical doctors often have no supervisors since the majority of them works in medical groups, is solo practice owner, independent contractor, or has ownership stake in a multi-specialty practice (Hawkinds, 2014; Healthcare, 2015). Assessment by peers is less attractive since previous research shows that peers are always able to judge the performance of their colleagues. While some researchers are positive about the use of peer-reviews (Brundage, et al., 2013; Epstein & Hundert, 2002), others showed that some of the peer-identified experts did not perform better than novices on tasks that were central to their expertise, as for example weather forecasts or treatment of psychotherapy patients (Ericsson, 2009; Ericsson & Lehman, 1996). There is one category of medical doctors who do have supervisors: residents in specialty training. Residents have completed medical school and perform many medical duties while they are in specialty training, but since they are in training, their job performance in continuously monitored. Hence, residency training provides the most optimal opportunity to investigate the relationship between selection measures and job performance.

OUTLINE OF THIS THESIS

Following the previous paragraphs, two aspects will be addressed in this thesis. First, non-cognitive capabilities become more and more important in the medical field, but the predictive value of non-cognitive measures is not well established and further insight is highly required (Balogun, et al., 1986; Morgeson, et al., 2007; Salvatori, 2001; Siu & Reiter, 2009). Second, as selecting the right person is not limited to medical school selection, information from the fields of personnel assessment and organizational psychology will be used to determine what medical student selection can learn from this.

Part 1: Cognitive versus non-cognitive selection
The study represented in *Chapter 2* entered the discussion of the predictive value of cognitive and non-cognitive selection methods. Non-cognitive capabilities are increasingly required in the medical field as it is acknowledged that medical care implies more than medical knowledge, notwithstanding the limited predictive validity of non-cognitive selection methods. However, as no previous studies have measured the effect of cognitive versus non-cognitive selection on academic performance under similar conditions, it is impossible to truly compare the predictive validity of different selection methods. In Chapter 2, a unique experiment is described that enabled the comparison of academic performance of cognitive selected students and non-cognitive selected students, to a control group, i.e. lottery-admitted students.

Part 2: Thinking outside the box
In the second part of this thesis, views from personnel assessment and organizational psychology are investigated in order to determine whether they can be applied to medical school selection. In *Chapters 3 and 4*, the importance of self-regulated learning has been addressed. Firstly, in Chapter 3, the potential development of self-regulated learning

skills during the pre-clinical stage of medical school and the relation between self-regulated learning skills and academic performance was investigated. Students from the first and third year of medical school were asked to complete a questionnaire containing questions about their study behaviour. Subsequently, the relation between their learning behaviour and their study results in the first and third year was assessed.Secondly, in Chapter 4, the influence of the learning environment, i.e. the medical curriculum, on the potential development of self-regulated learning skills, was addressed. Second and sixth semester students from two Brazilian medical schools completed a questionnaire that compared the effect of the curriculum on the development of self-regulated learning skills.

In *Chapter 5*, the role of expert performance during residency training was examined. In many other domains of human endeavour, outstanding performances are reached by those who engage in prolonged, deliberate practice. In order to investigate the relation between aspects of deliberate practice and performance as a resident, supervisor judgments of residents from four different specialty programmes (i.e. anaesthesiology, internal medicine, paediatrics and surgery), were compared to residents' responses regarding their behaviour during residency training.

Chapter 6 has a more specific focus: surgical residency performance. Since no criterion of a good medical doctor is available, medical school are not able to investigate whether the selection methods they use identify those candidates will perform well in the medical profession. However, residents are assessed in the scope of their training. This enables investigating the relation between selection methods and proficiency on the job of residents, who are almost independent medical specialists. In this study, pre-training variables were derived from the files of residents in surgery training and compared to their performance. Two different measures were used as indicators of performance: the overall judgment and scores on knowledge tests.

Finally, in *Chapter 7*, the findings of the different studies are summarized and important considerations for medical school selection are discussed, followed by suggestions for further research.

REFERENCES GENERAL INTRODUCTION

Ackerman, P. L., Chamorro-Premuzic, T., & Furnham, A. (2011). Trait complexes and academic achievement: Old and new ways of examining personality in educational contexts. *British Journal of Educational Psychology, 81*(1), 27-40.

Albanese, M., Snow, M., Skochelak, S., Huggett, K., & Farrell, P. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine, 78*(3), 313-321.

Anderson, W. D. (2011). Outside looking in: observations on medical education since the Flexner Report. *Medical Education, 45*(1), 29-35.

Arthur, W., Woehr, D.J., Graziano, W.G. (2001). Personality testing in employment settings - Problems and issues in the application of typical selection practices. *Personnel Review, 30*(5-6), 657-676.

Baars, G. J. A. (2009). *Factors related to student achievement in medical school.* Thesis: Erasmus University Rotterdam, the Netherlands.

Baker, J. (2014). 20 years later: deliberate practice and the development of expertise in sport. *International review of sport and exercise psychology, 7*(1), 135-137.

Balogun, J. A., Karacoloff, L. A., & Farina, N. T. (1986). Predictors of academic achievement in physical therapy. *Physical Therapy, 66*(6), 976-980.

Basco, W. T., Jr., Lancaster, C. J., Gilbert, G. E., Carey, M. E., & Blue, A. V. (2008). Medical school application interview score has limited predictive validity for performance on a fourth year clinical practice examination. *Advances in Health Sciences Education : theory and practice, 13*(2), 151-162.

Becker, T. E. (2005). “Development and validation of a situational judgment test of employee integrity.” *International Journal of Selection and Assessment,13*(3): 225-232.

Benbassat, J., & Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. *Advances in Health Sciences Education, 12*(4), 509-521.

Bhatti, M. A., Battour, M. M., & Ismail, A. R. (2013). Expatriates adjustment and job performance: an examination of individual and organizational factors. *The international journal of productivity and performance management, 62*(7), 694-717.

Boekaerts, M., Printrich, P. R., & Zeidner, M. (2000). *Handbook of self-regulation* (2 ed.). Burlington, USA: Elsevier Academic Press.

Bore, M. R., Munro, D. & Powis, D. A. (2009). A comprehensive model for the selection of medical students. *Medical Teacher, 31*(12), 1066-1972.

Brundage, M., Foxcroft, S., McGowan. T., Guteirrez, E., Sharpe, M, & Warde, P. (2013). A survey of radiation treatment planning peer-review activities in a provincial radiation oncology programme: current practice and future directions. *BMJ Open 3*(e003241), 1-8.

Brydges, R., & Butler, D. (2012). A reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education, 46*(1), 71-79.

Brydges, R., Nair, P., Ma, I., Shanks, D., & Hatala, R. (2012). Directed self regulated learning versus instructor regulated learning in simulation training. *Medical*

Education, 46(7), 648-656.
Bullimore, D. D. W. (1992). Selection interviewing for medical school admission. *Medical Education, 26*(5), 347-349.
Burch, V. C. (2009). Medical school admissions: where to next? *Advances in Health Sciences Education : theory and practice, 14*(2), 153-157.
Carrier, A., & Morin, C. (2014). Enabling students' self-regulation and teachers' feedback in concept mapping. *Medical Education, 48*(5), 523-524.
Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic per formance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37*(4), 319-338.
Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*(2), 151-165.
Cohen-Schotanus, J., Muijtjens, A. M., Reinders, J. J., Agsteribbe, J., van Rossum, H. J., & van der Vleuten, C. P. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education, 40*(10), 1012-1019.
Côté, J., Deakin, J., & Baker, J. (2005). Expertise in ultra-endurance triathletes early sport involvement, training structure, and the theory of deliberate practice. *Journal of Applied Sport Psychology, 17*(1), 64-78.
Courneya, C. A., Wright, K., Frinton, V., Mak, E., Schulzer, M., & Pachev, G. (2005). Medical student selection: choice of a semi-structured panel interview or an unstructured one-on-one interview. *Medical Teacher, 27*(6), 499-503.
Cremonini, L., van Niekerk, W. & Vossensteyn, H. (2010). Selective admission in higher education: risks and limiting conditions. *Center for Higher Education Policy Studies,* 1-8.
Donnon, T., Paolucci, E. O., & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Academic Medicine, 82*(1), 100-106.
Dore, K. L., Hanson, M., Reiter, H. I., Blanchard, M., Deeth, K., & Eva, K. W. (2006). Medical school admissions: enhancing the reliability and validity of an autobiographical screening tool. *Academic Medicine, 81*(10 Suppl), S70-73.
Dore, K. L., Reiter, H. I., Eva, K. W., Krueger, S., Scriven, E., Siu, E., Hildsen, S., Thomas, J. & Norman, G.R. (2009). Extending the interview to all medical school candidates--Computer-Based Multiple Sample Evaluation of Noncognitive Skills (CMSENS). *Academic Medicine, 84*(10 Suppl), S9-12.
Drenth, P. (2004). Selectie aan de poort van het hoger onderwijs. *Tijdschrift voor Hoger Onderwijs & Management*(4)*,* 48-51.
DUO. (2015). Decentrale selectie: Selectie door universiteit of hogeschool. Retrieved 27-05-2015, from https://www.duo.nl/particulieren/stu dent-hbo-of-universiteit/loten/decentrale-selectie.asp
Epstein, R. M. & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA, 287*(2), 226-235.
Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (Vol. 1). Mahwah, USA: Lawrence Erlbaum Associates.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine Research in Medical Education Proceedings of the Forty-Third Annual Conference November 7-10, 2004, 79*(10), S70-S81.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine, 15*(11), 988-994.

Ericsson, K. A. (2009). *The Cambridge handbook of expertise and expert performance:* edited by K. Anders Ericsson ... [et al.]. New York [etc.]: Cambridge University Press.

Ericsson, K. A. (2013). Training history, deliberate practice and elite sports performance: An analysis in response to Tucker and Collins review-what makes champions? *British journal of sports medicine, 47*(9), 533-535.

Ericsson, K. A., Krampe, R. T., & Tësch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., & Lehman, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273-305.

Ericsson, K. A., Nandagopal, K., & Roring, R. W. (2009). Toward a science of exceptional achievement: attaining superior performance through deliberate practice. *Annual report the New York Academy of Sciences, 1172*, 199-217.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: prospects and limits.* Cambridge [etc.]: Cambridge University Press.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science, 24*(1), 1-24.

Eskander, A., Shandling, M., & Hanson, M. D. (2013). Should the MCAT exam be used for medical school admissions in Canada? *Academic Medicine, 88*(5), 572-580.

Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004a). The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine, 79*(10), S40-S42.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43*(8), 767-775.

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: the multiple mini-interview. *Medical Education, 38*(3), 314-326.

Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *British Medical Journal, 324*(7343), 952-957.

Ferguson, E., James, D., O'Hehir, F., Sanders, A., & McManus, I. C. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal, 326*(7386), 429-432.

Flexner, A. (1909). Adjusting the College to American Life. *Science, 29*(740), 361-372.

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2002). Personality, cognitive ability, and beliefs about intelligence as predictors of academic

performance. *Learning and Individual Differences, 14*(1), 47-64.
Gagne, F., & St Pere, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*(1), 71-100.
Greveson, G. C., & Spencer, J. A. (2005). Self-directed learning – the importance of concepts and contexts. *Medical Education, 39*(4), 348-349.
Haidet, P., Dains, J. E., Paterniti, D. A., Hechtel, L., Chang, T., Tseng, E., Rogers, J.C. (2002). Medical student attitudes toward the doctor–patient relationship. *Medical Education, 36*(6), 568-574.
Hamdy, H., Prasad, K., Anderson, M. B., Scherpbier, A., Williams, R., Zwierstra, R., Cuddihy, H. (2006). BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher, 28*(2), 103-116.
Harasym, P. H., Woloschuk, W., Mandin, H., & Brundin-Mather, R. (1996). Reliability and validity of interviewers' judgments of medical school candidates. *Academic Medicine, 71*(1 Suppl), S40-42.
Harnett, R. T., & Willingham, W. W. (1980). The Criterion Problem: What measure of success in graduate education? *Applied Psychological Measurement, 4*(3), 281-291.
Hawkinds, M. (2014). 2014 Survey of America's physicians practice patterns & perspectives. Retrieved 15/05/2015, 2015, from http://www.physicians foundation.org/uploads/default/2014_Physicians_Foundation_Biennial_ Physician_Survey_Report.pdf
Healthcare, J. (2015). A profile of physicians' practice environments. Retrieved 15/01/2015, 2015, from http://www.jacksonhealthcare.com/ media-room/articles/physician-trends/physician-data-on-practice-evironment -workload-and-patient-access/
Hodges, N. J., Kerr, T., Starkes, J. L., Weir, P. L., & Nananidou, A. (2004). Predicting performance times from deliberate practice hours for triathletes and swimmers: what, when, and where is practice important? *Journal of Experimental Psychology: Applied, 10*(4), 219-237.
Hofmeister, M., Lockyer, J., & Crutcher, R. (2009). The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education, 43*(6), 573-579.
Holton, E. F., Bates, R. A., & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly, 11*(4), 333-360.
Hong, E., & O'Neil, H. F. (2001). Construct validation of a trait self-regulation model. *International Journal of Psychology, 36*(3), 186-194.
Hughes, P. (2002). Can we improve on how we select medical students? *Journal of the Royal Society of Medicine, 95* (1), 18-22.
Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98.
Hurwitz, S., Kelly, B., Powis, D., Smyth, R., & Lewin, T. (2013). The desirable qualities of future doctors - a study of medical student perceptions. *Medical Teacher, 35*(7), e1332-e1339.
Irby, D. M., Cooke, M., & O'Brien, B. C. (2010). Calls for reform of medical education by the Carnegie Foundation for the Advancement of teaching:

1910 and 2010. *Academic Medicine, 85*(2), 220-227.
Jonker, L., Elferink-Gemser, M. T., Toering, T. T., Lyons, J., & Visscher, C. (2010). Academic performance and self-regulatory skills in elite youth soccer players. *Journal of Sports Sciences, 28*(14), 1605-1614.
Kreiter, C., Yin, P., Solow, C., & Brennan, R. (2004). Investigating the reliability of the medical school admissions interview. *Advances in Health Sciences Education, 9*(2), 147-159.
Krupa, C. (2010). Medical education still evolving 100 years after Flexner report. *American Medical News*, p. 22.
Kulatunga-Moruzi, C., & Norman, G. R. (2002). Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and learning in medicine, 14*(1), 34-42.
Lambe, P., & Bristow, D. (2011). Predicting medical student performance from attributes at entry: a latent class analysis. *Medical Education, 45*(3), 308-316.
Leichner, P., Eusebio-Torres, E., & Harper, D. (1981). The validity of reference letters in predicting resident performance. *Journal of Medical Education, 56*(12), 1019-1021.
Lievens, F., Coetsier, P., de Fruyt, F., & de Maeseneer, J. (2002). Medical students' personality characteristics and academic performance: a five-factor model perspective. *Medical Education, 36*(11), 1050-1056.
Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. *Journal of Applied Psychology, 94*(6), 1514- 1535.
Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96*(5), 927-940.
Lycke, K. H., Grøttum, P., & Strømsø, H. I. (2006). Student learning strategies, mental models and learning outcomes in problem-based and traditional curricula in medicine. *Medical Teacher, 28*(8), 717-722.
Mahon, K. E., Henderson, M. K., & Kirch, D. G. (2013). Selecting tomorrow's physicians: The key to the future health care workforce. *Academic Medicine, 88*(12), 1806-1811.
McGaghie, W. C. (1990). Qualitative variables in medical school admission. *Academic Medicine, 65*(3), 145-149.
McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D., & Richards, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal, 331*(7516), 555-559.
McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. R. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *British Medical Journal, 327*(7407), 139-142.
Mitchell, K. J. (1990). Traditional predictors of performance in medical school. *Academic Medicine, 65*(3), 149-158.
Mooney, C. Z. (1993). *Bootstrapping: Christopher Z. Mooney, Robert D. Duval* (Vol. no. 07-095). Newbury Park, [Calif.]: SAGE.
Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel

selection contexts. *Personnel Psychology, 60* (3), 683-729.

Nieuwhof, M. G. H., ten Cate, O., Oosterveld, P. & Soethout, M. B. M. (2004). Measuring Strenght of Motivation for Medical School. *Medical Education Online, 9*(16), 1-7.

Oostrom, J., Born, M., Serlie, A., & van der Molen, H. (2010). Webcam testing: Validation of an innovative open-ended multimedia test. *European Journal of Work and Organizational Psychology, 19*(5), 532-550.

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neill, P. (2012). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education, 46*(9), 850-868.

Powis, D. A. (1994). Selecting medical students. *Medical Education, 28*(5), 443-469.

Powis, D.A. (2015). Selecting medical students: An unresolved challenge. *Medical Teacher, 37*(3), 252-260.

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. & Wilkinson, D. (2011). Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher, 33*(3), 215-223.

Pulvirenti, M., McMillan, J., & Lawn, S. (2014). Empowerment, patient centred care and self-management. *Health Expectations, 17*(3), 303-310.

Razack, S., Maguire, M., Hodges, B., & Steinert, Y. (2012). What might we be saying to potential applicants to medical school? Discourses of excellence, equity, and diversity on the web sites of Canada's 17 medical schools. *Academic Medicine, 87*(10), 1323-1329.

Roberts, C., Zoanetti, N., & Rothnie, I. (2009). Validating a multiple mini-interview question bank assessing entry-level reasoning skills in candidates for graduate-entry medicine and dentistry programmes. *Medical Education, 43*(4), 350-359.

Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging, 22*(2), 291-299.

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in health sciences education: theory and practice, 6*(2), 159-175.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*(1-2), 187-210.

Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education, 41*(12), 1133-1139.

Schmitt, N. (Ed.). (2012). *The Oxford handbook of personnel assessment and selection* (1 ed.). New York: Oxford University Press.

Siu, E., & Reiter, H. I. (2009). Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education : theory and practice, 14*(5), 759-775.

Sonnentag, S., & Kleine, B. M. (2000). Deliberate practice at work: A study with insurance agents. *Journal of Occupational and Organizational Psychology* 73(1), 87-102.

Spina, A. M., Smith, T. A., Marciani, R. D., & Marshall, E. O. (2000). A survey of resident selection procedures in oral and maxillofacial surgery. *Journal of Oral and Maxillofacial Surgery, 58*(6), 660-666.

St-Sauveur, C., Girouard, S., & Goyette, V. (2014). Use of situational judgment tests in personnel selection: Are the different methods for scoring the response options equivalent? *International Journal of Selection and Assessment, 22*(3), 225-239.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240-261.

Stewart, M. (2001). Towards a global definition of patient centred care: The patient should be the judge of patient centred care. *BMJ: British Medical Journal, 322*(7284), 444-445.

Taylor, C. W., Price, P. B., Richards, J. M., Jr., & Jacobsen, T. L. (1965). An investigation of the criterion problem for a group of medical general practitioners. *Journal of Applied Psychology, 49*(6), 399-406.

ten Cate, O. (2007). Medical education in the Netherlands. *Medical Teacher, 29*(8), 752-757.

Tracey, J. B., Sturman, M. C., & Tews, M. J. (2007). Ability versus personality: Factors that predict employee job performance. *Cornell Hotel and Restaurant Ad ministration Quarterly, 48*(3), 313-322.

van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences, 20*(3), 220-224.

van der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: a longitudinal study from a developmental perspective. *European journal of Psychology of Education, 29*(1), 117-137.

van der Stel, M., Veenman, M. V. J., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: a cross-sectional study from a developmental perspective. *Zentralblatt für Didaktik der Mathematik, 42*(2), 219-229.

Violato, C., & Donnon, T. (2005). Does the medical college admission test predict clinical reasoning skills? A longitudinal study employing the medical council of Canada clinical reasoning examination. *Academic Medicine, 80*(10), S1-3.

Whetzel, D. L., McDaniel, M.A. (2009). Situational judgment tests: an overview of current research. *Human resource management review, 19*(3): 188-202.

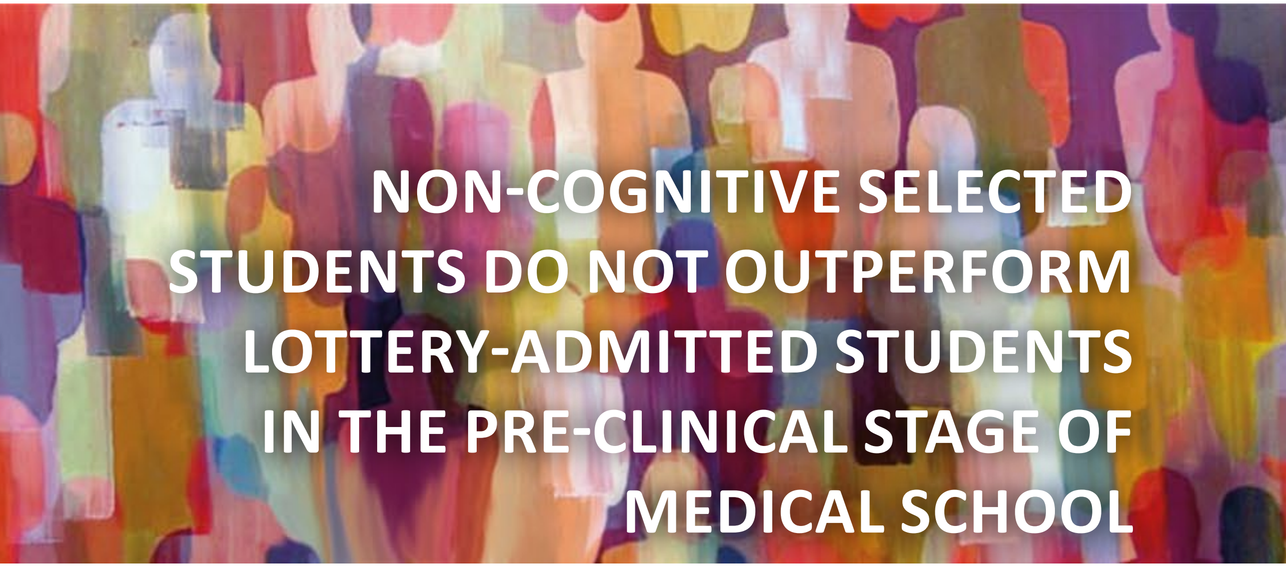
Whitehouse, C. (1997). Pre-medicine and selection of medical students. *Medical Education, 31*(S1), 3-6.

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology June, 90*(2), 224-235.

Zimmerman, B. J. (1986). Becoming a Self-Regulated Learner: Which are the key subprocesses? *Contemporary Educational Psychology,* (11), 307-313.

Susanna M. Lucieer
Karen M. Stegers-Jager
Remy M.J.P. Rikers
Axel P.N. Themmen

Advances in Health Sciences Education (2015)
DOI 10.1007/s10459-015-9610-4



NON-COGNITIVE SELECTED STUDENTS DO NOT OUTPERFORM LOTTERY-ADMITTED STUDENTS IN THE PRE-CLINICAL STAGE OF MEDICAL SCHOOL

Chapter 2

ABSTRACT

Medical schools all over the world select applicants using non-cognitive and cognitive criteria. The predictive value of these different types of selection criteria, however has never been investigated within the same curriculum while using a control group. We therefore set up a study that enabled us to compare the academic performance of three different admission groups, all composed of school leaver entry students, and all enrolled in the same Bachelor programme: students selected on non-cognitive criteria, students selected on cognitive criteria and students admitted by lottery. First-year GPA and number of course credits (ECTS) at 52 weeks after enrolment of non-cognitive selected students (N = 102), cognitive selected students (N = 92) and lottery-admitted students (N = 356) were analysed. In addition, the chances of dropping out, the probability of passing the third-year OSCE, and of completing the Bachelor programme within three years were compared. Although there were no significant differences between the admission groups in first-year GPA, cognitive selected students had obtained significantly more ECTS at 52 weeks and dropped out less often than lottery-admitted students. Probabilities of passing the OSCE and completing the Bachelor programme within three years did not significantly differ between the groups. These findings indicate that the use of only non-cognitive selection criteria is not sufficient to select the best academically performing students, most probably because a minimal cognitive ability is needed to succeed in medical school.

INTRODUCTION

Medical schools all over the world select students as the number of applicants highly exceeds the number of available places. In Canada, the United States, the United Kingdom, and the Netherlands, chances to be admitted to medical school range from six to 33 percent (Razack, et al., 2012; DUO, 2014a). In addition, medical education is expensive for both student and society, and medical schools have the responsibility of educating future medical doctors who are able to provide optimal healthcare (Hughes, 2002). It is therefore important to select those applicants who will be able to successfully complete the programme and will become well-performing medical doctors (Albanese, et al., 2003; Basco, et al., 2008; Burch, 2009; Kulatunga-Moruzi & Norman, 2002; Mitchell, 1990).

A variety of selection instruments has been used to admit the most promising applicants to medical school. These instruments range from cognitive selection tools, such as the pre-university grade point average (pu-GPA; Ferguson, et al., 2002; Siu & Reiter, 2009) and aptitude tests such as the Medical School Admission Test (MCAT; Albanese, et al., 2003; Kelly, et al., 2013) to selection instruments aiming to measure more non-cognitive capabilities of the applicants. Non-cognitive selection tools include regular interviews (Basco, et al., 2008; Burch, 2009), reference letters (Ferguson, et al., 2003; Siu & Reiter, 2009), motivation letters (Prideaux, et al., 2011; Salvatori, 2001), psychometric questionnaires (Arthur, 2001), Multiple Mini Interviews (Eva, Reiter, et al., 2004a; Eva, Rosenfeld, et al., 2004), and Situational Judgment Tests (Lievens, 2013; Patterson, Lievens, Kerrin, Zibarras, & Carette, 2012). The reliability and validity of these selection instruments differ widely, and while research has shown that mainly the pu-GPA and MCAT have, especially in a combined form, a reasonable predictive value for performance during medical school (Albanese, et al., 2003; Dunleavy, Kroopnick, Dowd, Searcy, & Zhao, 2013; Ferguson, et al., 2002), several non-cognitive selection tools appear not to be so reliable and valid (Basco, et al., 2008; Burch, 2009; Ferguson, et al., 2003; Salvatori, 2001). With respect to interviews, a large interviewer variability exists and personal factors such as gender and background of the applicant, as well as expectations of the interviewer, are sources of bias (Eva, Reiter, Rosenfeld, & Norman, 2004b; Salvatori, 2001). Problems with reference and motivation letters are the absence of a scoring system to compare the letters (Ferguson, et al., 2002) and that they rarely distinguish one applicant from another (Spina, et al., 2000). The use of psychometric questionnaires for selection procedures is also criticized, because of the possible lack of reliability and the risk of obtaining desirable answers (Arthur, 2001). Although the first results on the MMI and SJT are promising (Dore, et al., 2010; Eva, Rosenfeld, et al., 2004; Koczwara, et al., 2012), further research is needed to establish the long-term predictive value of these non-cognitive selection tools.

Whereas much research has been conducted on the predictive validity of selection tools, only a few studies included a control group to compare the performance of the selected students to (O'Neill, Hartvigsen, Wallstedt, Korsholm, & Eika, 2011; Urlings-Strop, Stegers-Jager, Stijnen, & Themmen, 2013; Urlings-Strop, Stijnen, Themmen, & Splinter, 2009; Urlings-Strop, Themmen, Stijnen, & Splinter, 2011). In addition, the predictive value of the selection tools might also depend on the curriculum employed by the medical school (Edwards, Friedman, & Pearce, 2013), which limits the generalizability of the findings and the possibility to compare the selection methods. It would thus be valuable to determine the independent contribution of non-cognitive and cognitive selection methods to academic performance in the presence of a control group, allowing a comparison of all groups under similar circumstances.

While many medical schools in the world select 100% of their applicants, the situation in the Netherlands is different, as at the time of the current study a maximum of 50% of the students were selected by the medical schools (Urlings-Strop, et al., 2011), although the percentage of selected students is expected to rise also in the Netherlands in view of recent policy changes phasing out the lottery practice (DUO, 2014b). In the Netherlands, applicants currently have three possibilities to be admitted to medical school: direct access, selection, and lottery. Direct access is provided to those applicants who obtained a pu-GPA of ≥ 8.0, on a scale from 1-10, where students need to score at least a 5.5 to pass. Regarding the students admitted by selection, each medical school employs its own particular, local selection procedure. Finally, non-selected students are admitted via a national lottery system that is weighted for secondary school performance, i.e. applicants are assigned to a lottery batch depending on their pu-GPA, with increasing odds to be admitted with increasing pu-GPA. Applicants who failed to be admitted by the local selection procedure are allowed to participate in the national lottery, which means they still have a chance to enter medical school. This provides a unique opportunity to create a control group of non-selected, but lottery-admitted medical school students.

A few studies have already taken advantage of this situation and studied the effect of selection on academic performance (Hulsman, et al., 2007; ten Cate, 2002; Urlings-Strop, et al., 2013; Urlings-Strop, et al., 2009; Urlings-Strop, et al., 2011). However, some of these studies were conducted during the development of procedures when selection for medical school was first allowed in the Netherlands, resulting in a relatively small number of selected students (1.5-6.2%), which might have influenced their findings (Hulsman, et al., 2007; ten Cate, 2002). These studies showed that selected students were more committed to healthcare during their first two years of medical school but did not perform better (Hulsman, et al., 2007), and that cognitive requirements do have some predictive value for academic performance (ten Cate, 2002). Research conducted at our own medical school showed that a selection procedure existing of a non-cognitive step followed by a cognitive one led to the inclusion of students who received higher grades during clerkships and dropped out less often than students who were admitted by lottery (Urlings-Strop, et al., 2009; Urlings-Strop, et al., 2011). In addition, one of the studies by Urlings-Strop and colleagues showed that success in a cognitive selection step related to a lower dropout rate, while higher grades during clerkships related to success on non-cognitive selection criteria (Urlings-Strop, et al., 2013). However, here, only those applicants who met the non-cognitive requirements were allowed to participate in the following cognitive step of the selection procedure, which made it impossible to measure the effect of cognitive or non-cognitive selection in isolation. A final study is from researchers in Denmark. They investigated the difference in dropout between students who were admitted on either their pu-GPA or on their results in a selection procedure. In this selection procedure, applicants were judged upon motivation, qualifications, general knowledge, and performance in an admission interview. This study showed that students who were selected dropped out less often compared to the students who were admitted on the basis of their pu-GPA (O'Neill, Hartvigsen, Wallstedt, Korsholm, & Eika, 2011). However, this study only focused on dropout and not on other measures of academic performance.

In the current study, we aimed to examine the effect of either non-cognitive or cognitive selection in isolation on academic performance using the lottery-admitted group as a control group. In this manner, the contribution of non-cognitive selection criteria and cognitive selection criteria could be measured independently among students who followed the same curriculum and could be compared to the lottery-admitted students. We expect all selected students to drop out, voluntary or academically dismissed, less often than the students who were admitted by lottery, as previous studies showed that students who had to put effort in their admission, dropped out less often, irrespective of their success in the selection procedure (O'Neill, et al., 2011; Urlings-Strop, et al., 2013). In addition, we expect the non-cognitive selected students to pass the third-year objective structured clinical examination (OSCE) more often than the other admission groups since previous research showed that success on non-cognitive criteria was related to higher clerkship grades (Urlings-Strop, et al., 2011). Next to this, we were interested to investigate whether there were other differences in academic performance, e.g., grade point average and number of course credits, between the non-cognitive selected students, the cognitive selected students and the lottery-admitted students.

METHODS

Context

This study was performed at the Erasmus MC Medical School in Rotterdam, the Netherlands. Everyone who has finished a pre-university secondary school level, with a combination of subjects obligatory for medicine, or who has similar qualifications, is allowed to apply to medical school in the Netherlands (ten Cate, 2007). Normally, in the local selection procedure of the Erasmus MC Medical School, applicants are selected if they succeed in two, consecutive, selection rounds: a non-cognitive (i.e. quality and quantity of their extracurricular activities before application) and a cognitive round (i.e. their scores on a set of five cognitive tests covering a medical subject). The cognitive round was limited to those who met a cut-off value in the first, non-cognitive, round. As regards the extracurricular activities, applicants must have spent at least four hours per week for a minimum of one year on a) voluntary work-related activities in healthcare, b) a managerial position in for example a school board or c) have achieved an outstanding performance in sports, science, literature or art. The rationale behind selecting applicants upon their extracurricular activities is that these applicants are able to distinguish themselves from others, by showing through their behaviour the motivation and ambition to carry out other activities as well as the ability to combine these activities with their secondary education (Urlings-Strop, et al., 2009). In the cognitive step, the applicants' level of cognitive ability and academic study skills was measured by their performance on five different cognitive tests. Test subjects include arithmetic, anatomy, scientific reading, logical thinking, and one test based on a medical subject referring to two lectures applicants have attended.

Once students are admitted to medical school, they follow an integrated and theme-oriented curriculum that comprises a three year Bachelor programme followed by a three year Master programme. The Bachelor of Medicine is divided into thematic blocks of 4–16 weeks, which are organized around pathophysiological systems and cover subjects starting from the basic sciences up to and including clinical practice. Per academic year, students can achieve a maximum of 60 European Credits (ECTS). ECTS are a standard for comparing student workload in the European Union and reflects successfully completed exams and assignments (Kuncel, Kochevar, & Ones, 2014). Students who quit voluntarily within the first two years of enrolment, or who do not meet the requirements set by medical school, i.e. having earned all first-year ECTS by the end of the second-year of enrolment, are considered dropouts.

Participants

In the present study, we included students who started in 2008 and in 2009 at Erasmus MC medical school, and were admitted by selection or lottery. Those who were directly admitted based upon their pu-GPA (≥ 8.0) were excluded from this analysis since these students had not participated in either selection or lottery and their numbers are small (< 10%). In total, 550 students (M_{age} = 19.4, SD = 1.5) were included; 102 non-cognitive selected students, 92 cognitive selected students and 356 lottery-admitted students. No significant differences in age, gender and pu-GPA were found between the admission groups and cohorts (see Table 1).

Table 1: Descriptive statistics per admission group and per year

	N	Mean age (*SD*)	% Female
2008			
- **Non-cognitive selected students**	102	19.3 (*1.5*)	64.7
- **Lottery-admitted students**	190	19.4 (*1.2*)	63.2
- **Total 2008**	292	19.3 (*1.3*)	64.6
2009			
- **Cognitive selected students**	92	19.6 (*2.0*)	51.5
- **Lottery-admitted students**	166	19.5 (*1.4*)	61.4
- **Total 2009**	258	19.5 (*1.6*)	57.8
Total lottery-admitted students	356	19.4 (*1.3*)	62.4
Total	550	19.4 (*1.5*)	60.9

SD = standard deviation

Procedure

For the experiment in this study, the regular local selection procedure of the Erasmus MC Medical School was adapted. While normally applicants are admitted when they succeed in both consecutive rounds, for this experiment, only the score in one round was taken into account. During one year, students were admitted only on the basis of their extracurricular activities (i.e. the non-cognitive selected students). They did participate in the cognitive step, but these results did not influence the decision to admit them to medical school. In the following years' selection, all students were admitted solely on the basis of their test scores (i.e. the cognitive selected students). They still had to hand in information about their extracurricular activities but were all allowed to take the cognitive tests, independently of the quality or quantity of their activities. In addition, in both years, at least 50% of the places were available to students who participated in the national lottery system. From these three different admission groups, dropout and various measures of academic performance were determined.

Dropout and measures of academic performance

Dropout was defined as quitting voluntary within the first two years or failing to acquire all 60 ECTS of the first-year subjects within the first two years of enrolment, as the latter one is a requirement to continue the programme. Academic performance was measured by 1) the mean grade of the exams in the first year at first attempt (first-year GPA), 2) the number of ECTS at 52 weeks of enrolment, 3) passing the third-year OSCE and 4) completing the Bachelor course within three years (i.e. having obtained 180 ECTS in three years).

Statistics

Data were analysed with IBM SPSS AMOS 18.0 (SPSS, Inc., Chicago, IL, USA). Since different cohorts were included in this analysis, first-year GPAs were converted into z-scores. ANOVA with post-hoc Bonferroni tests were used to compare the first-year GPA and ECTS at 52 weeks of the non-cognitive selected students, the cognitive selected students, and the lottery-admitted students. For first-year GPA and the number of ECTS at 52 weeks, *Welch F* was calculated. Effect sizes were determined using eta-squared (η^2) with values of .01, .06, and .14 indicating small, medium and large effects (Cohen, 1988; Lakens, 2013). Odds ratios (OR) with Wald statistics were obtained from logistic regression analyses to determine the chance of passing the OSCE, completing the Bachelor programme within three years and the chance of dropping out, for all three admission groups.

Ethical considerations
In the present study, all data were processed anonymously to make sure that no possible harm to participants could arise from this study. Since medical schools are allowed to set their own selection policies, the methods employed in the present paper did not constitute an experiment with respect to the applicants. The data on academic performance used in this study were collected as part of regular academic activities and obtained from the university administrative system; no individual informed consent was required.

RESULTS

No statistically significant differences were found between the non-cognitive selected, the cognitive selected, and the lottery-admitted students in first-year GPA, but cognitive-selected students had earned significantly more ECTS at 52 weeks than the lottery-admitted students (see Table 2). In addition, cognitive-selected students had a significantly lower probability to drop out than lottery-admitted students (11% versus 21%). The probabilities to pass the third-year OSCE and to obtain 180 ECTS in three years did not significantly differ among the three admission groups (Table 2).

Table 2: Comparison of academic performance of non-cognitive selected, cognitive selected and lottery-admitted students by ANOVA and Odds Ratios

	GPA, mean z- score ± SD	**ECTS at 52 weeks ± SD**	**OSCE OR (95% CI)**	**Drop-out OR (95% CI)**	**Bachelor degree in 3 years OR (95% CI)**
Non-cognitive selected students	.078 ± .87	48.1 ± 18.4	1.22 (.73-2.04)	.97 (.42-2.22)	1.51 (.88-2.57)
Cognitive selected students	.185 ± .86	51.8 ± 14.3	.886 (.50-1.57)	.32 (.14-.77)	1.11 (.66-1.86)
Lottery-admitted students	-.071 ± 1.06	44.8 ± 20.4	1.00 (-)	1.00 (-)	1.00 (-)
Test value	*Welch F =* 1.765	*Welch F =* 7.418**		*Wald = 6.557**	
Effect size		η^2 = .019			

SD = standard deviation, OR = odds ratio, OSCE = objective structured clinical examination, 95% CI = 95% confidence interval
**significant at p =.01, ** significant at p < .01, η^2 = effect size for ANOVA, with .01, .06, and .14 indicating small, medium and large effects*

DISCUSSION

The aim of this study was to gain insight into the effect of different selection criteria on academic performance. We therefore set up an experiment that enabled us to compare the academic performance of non-cognitive selected students to cognitive selected students, and to lottery-admitted students. This study is, to our best knowledge, the first to compare two groups of selected students using different types of criteria, who all take part in the same curriculum, and which also involves a control group (i.e. the lottery-admitted students).

Our hypothesis, that both groups of selected students would drop out less often than the lottery-admitted students, was partly confirmed. Although non-cognitive selected students did not drop out less often than lottery-admitted students, cognitive selected students did. Contrary to our expectations, non-cognitive selected students did not pass the OSCE more often than cognitive selected and lottery-admitted students. We did, however, find a difference in first-year performance between the admission groups: cognitive selected students earned more ECTS in 52 weeks than lottery-admitted students.

It was somewhat surprising that non-cognitive selected students did not drop out less often than lottery-admitted students. Previous research showed that participating in a selection procedure, irrespective of the outcome, prevented students from dropping out. Put differently, students who were willing to put effort in their admission were shown to drop out less often (O'Neill, et al., 2011; Urlings-Strop, et al., 2013), which is contrary to our findings. A possible explanation for this outcome is that motivation and ambition, the characteristics that were aimed to be determined in the non-cognitive round, are by themselves not enough to succeed in medical school. Indeed, research has shown that highly motivated students with inadequate and ineffective study methods have a higher chance of dropping out (Bennett, 2003), and that students also need to be able to adapt to the academic environment in order to perform well (Stage, 1989).

Another explanation for our contrary finding that might be interesting to investigate in future research, is whether the rationale behind the applicants' behaviour has changed over the years. That is, the requirements for admission have been made more transparent since the start of selection for medical school in the Netherlands, and applicants are much more aware of what is expected of them. The original reasoning behind the use of extracurricular activities for selection was that applicants with many extracurricular activities showed this behaviour consistent with their motivation or ambition to employ other activities (Urlings-Strop, et al., 2009). It may be the case that applicants nowadays invest time in extracurricular activities just because they want to enter medical school and are less driven internally. Thus, the quality of the motivation and ambition behind their behaviour may have changed resulting in a diminished positive effect on drop out.

The finding that the non-cognitive selected students do not pass the OSCE more often than the cognitive selected and the lottery-admitted students was also unexpected, since previous research showed that the non-cognitive selection step was related to higher clerkship grades (Urlings-Strop, et al., 2011). It has, however, to be mentioned that only one study examined this relationship so far. In addition, the results on the OSCE are not very discriminative since they are binary; you either pass or fail the test. Other criteria, or even a more discriminating OSCE, may be more relevant in judging performance in different stages of medical education. Future research is needed to investigate whether the non-cognitive selected students in this study do indeed earn higher clerkship grades. Apart from this, one could argue that the non-cognitive selected students did not pass the OSCE more often since a certain level of cognitive ability is required for optimal clinical performance. Several researchers acknowledged that just being able to communicate to a patient is not sufficient, since the conversation is less effective when the necessary knowledge is absent (Eva, et al., 2009; Miller, 1990). Multiple studies therefore encourage the use of a combination of pre-university performance or cognitive measures as well as non-cognitive tests when selecting future medical doctors (Ferguson, et al., 2003; Ranasinghe, Ellawela, & Gunatilake, 2012), and also the findings in this current study indicate that selection on only non-cognitive requirements is not sufficient. Non-cognitive criteria thus should be accompanied by cognitive selection criteria, since someone needs to have a sufficient level of cognitive ability to succeed in medical school.

The final outcome of this study, that cognitive selected students do drop out less often and earn more ECTS in 52 weeks than the lottery-admitted students, is less surprising, since our Bachelor programme is mainly cognitive based. These findings indicate that the cognitive selection criteria set a cognitive standard that students need to have to perform well in medical school. One could argue that, in the Netherlands, a cognitive standard is set by allowing only those applicants in medical school who have completed the highest level of secondary education. Nevertheless, there are differences between secondary school and university that might result in the fact that not everyone who is able to pass secondary education, performs well in medical school. For example, while secondary school exams focus on one course, medical school examinations cover multiple domains, ranging from basic sciences to clinical knowledge. Therefore, different knowledge and study skills are required to pass the exams. Since the cognitive tests used in the selection procedure resemble medical school examinations and aim to access the applicants' academic study skills, their results provide incremental validity on top of the secondary school examinations.

This study is not without limitations. In the Netherlands, a very homogeneous group of students applies to medical school, since different tracks (vocational, pre-higher education) of secondary education exist and students are already at the start of secondary education (approximately twelve years of age) matched to a level that suits their competencies. Only those students who have finished the highest level of secondary school, a pre-university secondary school level with a combination of subjects adequate for medicine, are allowed to apply to medical school (ten Cate, 2007). Thus, almost all applicants for medical school share many characteristics; they all graduated at the highest level of secondary education, all followed the same subjects and they all have the ambition to study medicine. Selection in homogeneous groups is difficult to achieve since clear cut-offs between who is selected and who is not are hard to define, resulting in little variation in outcome scores. The differences in academic performance between those who are, and those who are not selected, may be more apparent in a more heterogeneous group where more variation in selection scores is visible and better defined cut-offs can be applied. Another limitation concerns the use of the OSCE in year three; the strength of this outcome measure is diminished by its dichotomous nature and by student dropout, limiting student completion. A final limitation is that the Dutch lottery system uses pu-GPA weighted odds, and consequently the lottery-admitted students do not constitute an ideal, random control group. Nevertheless, their age, gender and pu-GPA did not differ from the admission groups and between the cohorts.

This study provides an important practical implication for medical schools. Most medical schools have to select, since the number of applicants is exceeding the number of places available (Razack, et al., 2012), and even though the importance of non-cognitive selection is widely acknowledged (Hughes, 2002; Hurwitz, et al., 2013), the outcomes of this study indicate that medical schools should include at least some cognitive requirements, besides non-cognitive measures, in their selection procedure, in order to admit the most promising students to medical school. Such cognitive tests should ideally measure knowledge and academic study skills required for medical school, as, for example, the ability to obtain deep understanding of complex subjects by integrating information from various sources such as lectures, scientific papers and self-study. This way, the cognitive tests provide incremental validity over secondary school. In addition, our findings show that selection of non-cognitive characteristics alone has no predictive value for success in medical school, suggesting that non-cognitive selection should be accompanied by some type of cognitive measure, such as pu-GPA, aptitude tests or study skill determination.

CONCLUSION

This study showed that students who were selected on non-cognitive criteria do not outperform lottery-admitted students during the Bachelor years of medical education, but students who were selected on cognitive criteria did earn more ECTS in the first 52 weeks and dropped out less often than lottery-admitted students. Our findings suggest that non-cognitive selection is not sufficient by itself but should be accompanied by cognitive selection, since a certain cognitive basis is required to succeed in medical school.

REFERENCES CHAPTER 2

Albanese, M., Snow, M., Skochelak, S., Huggett, K., & Farrell, P. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine, 78*(3), 313-321.

Arthur, W., Woehr, D. J., Graziano, W. G. (2001). Personality testing in employment settings - Problems and issues in the application of typical selection practices. *Personnel Review, 30*(5-6), 657-676.

Basco, W. T., Jr., Lancaster, C. J., Gilbert, G. E., Carey, M. E., & Blue, A. V. (2008). Medical school application interview score has limited predictive validity for performance on a fourth year clinical practice examination. *Advances in Health Sciences Education : theory and practice, 13*(2), 151-162.

Bennett, R. (2003). Determinants of undergraduate student drop out rates in a university business studies department. *Journal of Further and Higher Education, 27*(2), 123-141.

Burch, V. C. (2009). Medical school admissions: where to next? *Advances in Health Sciences Education : theory and practice, 14*(2), 153-157.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: NY: Routledge Academic.

Dore, K. L., Kreuger, S., Ladhani, M., Rolfson, D., Kurtz, D., Kulasegaram, K., Cullimore, A. J., Norman, G. R., Eva, K. W., Bates, S. & Reiter, H. I. (2010). The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions. *Academic Medicine, 85*(10 Suppl), S60-63.

Dunleavy, D. M., Kroopnick, M. H., Dowd, K. W., Searcy, C. A., & Zhao, X. (2013). The predictive validity of the MCAT exam in relation to academic performance through medical school: A lof 2001-2004 matriculants. *Academic Medicine, 88*(5), 666-671.

DUO. (2014a). Decentrale selectie. Selectie door universiteit of hogeschool. Retrieved 12-09-2014, from: https://www.duo.nl/particulieren/student-hbo-of-universiteit/loten/decentrale-selectie.asp

DUO. (2014b). Centrale loting afgeschaft. Retrieved 12-09-2014, from: http://www.rijksoverheid.nl/nieuws/2014/08/30/centrale-loting-afgeschaft.html

Edwards, D., Friedman, T., & Pearce, J. (2013). Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Medical Education, 13*(1), 173.

Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004a). The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine Research in Medical Education Proceedings of the Forty-Third Annual Conference November 7-10, 2004, 79*(10), S40-S42.

Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004b). The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Academic Medicine, 79*(6), 602-609.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Pre dictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43*(8), 767-775.

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE:

the multiple mini-interview. *Medical Education, 38*(3), 314-326.

Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *British Medical Journal, 324*(7343), 952-957.

Ferguson, E., James, D., O'Hehir, F., Sanders, A., & McManus, I. C. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal, 326*(7386), 429-432.

Hughes, P. (2002). Can we improve on how we select medical students? *Journal of the Royal Society of Medicine, 95*(1), 18-22.

Hulsman, R. L., van der Ende, J. S. J., Oort, F. J., Michels, R. P. J., Casteelen, G., & Griffioen, F. M. M. (2007). Effectiveness of selection in medical school admissions: evaluation of the outcomes among freshmen. *Medical Education, 41*(4), 369-377.

Hurwitz, S., Kelly, B., Powis, D., Smyth, R., & Lewin, T. (2013). The desirable qualities of future doctors-a study of medical student perceptions. *Medical Teacher, 35*(7), e1332-e1339.

Kelly, M. E., Regan, D., Dunne, F., Henn, P., Newell, J., & O'Flynn, S. (2013). To what extent does the Health Professions Admission Test-Ireland predict performance in early undergraduate tests of communication and clinical skills? - An observational cohort study. *BMC Medical Education, 13*(68)*, 1-11*.

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education, 46*(4), 399-408.

Kulatunga-Moruzi, C., & Norman, G. R. (2002). Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and learning in medicine, 14*(1), 34-42.

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22*(1), 101-107.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, (863),* 1-12.

Lievens, F. (2013). Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Medical Education, 47*(2), 182-189.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine 65*(9), S63-S67.

Mitchell, K. J. (1990). Traditional predictors of performance in medical school. *Academic Medicine, 65*(3), 149-158.

O'Neill, L., Hartvigsen, J., Wallstedt, B., Korsholm, L., & Eika, B. (2011). Medical school dropout - testing at admission versus selection by highest grades as predictors. *Medical Education, 45*(11), 1111-1120.

Patterson, F., Lievens, F., Kerrin, M., Zibarras, L., & Carette, B. (2012). Designing selection systems for medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. *International Journal* *of Selection and Assessment, 20*(4), 486-496.

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. & Wilkinson, D. (2011). Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher, 33*(3), 215-223.

Ranasinghe, P., Ellawela, A., & Gunatilake, S. B. (2012). Non-cognitive characteristics predicting academic success among medical students in Sri Lanka. *BMC Medical Education, 12*(1), 66.

Razack, S., Maguire, M., Hodges, B., & Steinert, Y. (2012). What might we be saying to potential applicants to medical school? Discourses of excellence, equity, and diversity on the web sites of Canada's 17 medical schools. *Academic Medicine, 87*(10), 1323-1329.

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in health sciences education : theory and practice, 6*(2), 159-175.

Siu, E., & Reiter, H. I. (2009). Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. *Advances in Health Sciences Education : theory and practice, 14*(5), 759-775.

Spina, A. M., Smith, T. A., Marciani, R. D., & Marshall, E. O. (2000). A survey of resident selection procedures in oral and maxillofacial surgery. *Journal of Oral and Maxillofacial Surgery, 58*(6), 660-666.

Stage, F. K. (1989). Motivation, academic and social integration, and the early dropout. *American Educational Research Journal, 26*(3), 385-402.

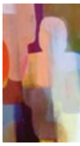
ten Cate, O. (2007). Medical education in the Netherlands. *Medical Teacher, 29*(8), 752-757.

ten Cate, O., Hendrix, H. L., de Fockert Koefoed, K. J. J. & Rietveld, W. J. (2002). Studieresultaten van toegelatenen binnen en buiten de loting. *Tijdschrift voor Medisch Onderwijs, 21*(6), 253-258.

Urlings-Strop, L. C., Stegers-Jager, K. M., Stijnen, T., & Themmen, A. P. N. (2013). Academic and non-academic selection criteria in predicting medical school performance. *Medical Teacher, 35*(6), 497-502.

Urlings-Strop, L. C., Stijnen, T., Themmen, A. P. N., & Splinter, T. A. (2009). Selection of medical students: a controlled experiment. *Medical Education, 43*(2), 175-183.

Urlings-Strop, L. C., Themmen, A. P. N., Stijnen, T., & Splinter, T. A. (2011). Selected medical students achieve better than lottery-admitted students during clerkships. *Medical Education, 45*(10), 1032-1040.



Susanna M. Lucieer
Laura Jonker
Chris Visscher
Remy M.J.P. Rikers
Axel P.N. Themmen

Medical Teachter (2015)
DOI 10.3109/0142159x.2015.1073240

SELF-REGULATED LEARNING AND ACADEMIC PERFORMANCE IN MEDICAL EDUCATION

Chapter 3

ABSTRACT

Content
Medical schools aim to graduate medical doctors who are able to self-regulate their learning. It is therefore important to investigate whether medical students' self-regulated learning skills change during medical school. In addition, since these skills are expected to be helpful to learn more effectively, it is of interest to investigate whether these skills are related to academic performance.

Methods
In a cross-sectional design, the Self-Regulation of Learning Self-Report Scale (SRL-SRS) was used to investigate the change in students' self-regulated learning skills. First and third-year students (N = 949, 81.7%) SRL-SRS scores were compared with ANOVA. The relation with academic performance was investigated with multinomial regression analysis.

Results
Only one of the six skills, reflection, significantly changed during medical school. The third year students' reflection scores were higher than the first year students' reflection scores. Additionally, a small, but positive relation of monitoring, reflection and effort with first-year GPA was found, while only effort was related to third-year GPA.

Conclusion
The change in self-regulated learning skills is minor as only the level of reflection differs between the first and third year. In addition, the relation between self-regulated learning skills and academic performance is limited. Medical schools are therefore encouraged to re-examine the curriculum and methods they use to enhance their students' self-regulated learning skills. Future research is required to understand the limited impact on performance.

INTRODUCTION

The medical profession has to ensure that high standards in providing patient care are repeatedly being met in the context of a rapidly and constantly changing medical world (Bjork, Dunlosky, & Kornell, 2013; Brydges & Butler, 2012). This means that medical doctors have to stay updated on the developments in their field of expertise and have to maintain their competencies (Artino, et al., 2012; Brydges & Butler, 2012; Greveson & Spencer, 2005; Premkumar, et al., 2013). To be able to benefit and choose from the many opportunities of continuous medical education, medical doctors have to define their own learning needs, set personal goals and engage in the most appropriate learning activities (Brydges, et al., 2012; Lycke, et al., 2006; Premkumar, et al., 2013). In short, medical doctors have to be self-regulated learners, which means that they have to be behaviourally, metacognitively and motivationally proactive in their learning process (Jonker, et al., 2010; Wolters, 1998; Zimmerman, 1986).

According to Ertmer and Newby (1996), self-regulated learners are individuals who are able to plan their study behaviour, monitor their progress, reflect upon and evaluate the entire learning process. Other researchers also highlighted the importance of motivational components in self-regulated learning (Hong & O'Neil, 2001; Sitzmann & Ely, 2011). They argued that one may be able to plan, monitor, reflect upon, and evaluate his or her learning behaviour, but that these competencies are of little value when one is not motivated to employ them. Therefore, they added two subcomponents of motivation to the concept of self-regulated learning, i.e. effort and self-efficacy. Effort is crucial to reach the goals self-regulated leaners have set, and self-efficacy is important since one needs to have trust in his or her own potential in order to complete a task (Hong & O'Neil, 2001; Sitzmann & Ely, 2011).

Since medical doctors should be self-regulated learners, medical schools aim to graduate medical doctors who have acquired these competencies (Greveson & Spencer, 2005; Premkumar, et al., 2013). Students can be helped to become self-regulated learners by providing them effective learning skills and appropriate and timely feedback (Hong & O'Neil, 2001; Zimmerman, 1989). When a task is correctly illustrated, a student can observe and imitate the performance afterwards. After first heavily relying on the observations, the process can become more and more internalized until it eventually becomes "self-regulated" (Brydges, et al., 2015). Unfortunately, self-regulated learning skills are not always emphasized during medical school (Artino, et al., 2012). While studies showed that students do develop self-regulated learning skills during medical school (Loyens, Magda, & Rikers, 2008), some graduates feel uncertain and unprepared to do so (Artino, et al., 2012). Therefore, it is important to investigate to what extent medical students' self-regulated learning skills change during their education.

In addition, it has been shown that self-regulated learning is one of the best predictors of academic performance (Pintrich & de Groot, 1990). Self-regulated learning is viewed as a proactive learning process that is used to set learning goals and develop effective strategies for learning (Zimmerman, 2008). This process helps people to transform mental abilities in academic skills, such as setting goals, developing learning strategies, and monitoring the progress and effectiveness of their learning (Zimmerman, 2002, 2008). Knowing how to monitor the progress of your learning and how to control and adapt your learning behaviour is seen as a requirement for being a truly effective learner (Bjork, et al., 2013; Ertmer & Newby, 1996). Although research suggests that it is not necessary to use self-regulated learning skills for high achievement (Ablard & Lipschultz, 1998), it has been shown that self-regulated learners are more effective learners (Nota, 2004; Toering, Elferink-Gemser, Jordet, & Visscher, 2009) who get more out of their potential (Zimmerman, 1986) and attain higher grades during secondary school (Nota, 2004) and in college (Ablard & Lipschultz, 1998).

In the present study, the Self-Regulation of Learning Self-Report Scale is used. This questionnaire contains six subscales: planning, monitoring, evaluation, reflection, effort and self-efficacy, following the theories of Ertmer and Newby (1996) and Hong and O'Neill (2001). The questionnaire has been shown to be a valid and reliable measure of self-regulated learning (Lucieer, et al., 2015; Toering, Elferink-Gemser, Jonker, van Heuvelen, & Visscher, 2012) and addresses two important questions on self-regulated learning. First, do students' self-regulated learning skills change during medical school? We hypothesize that they do, since medical education is based upon the premise that students should develop self-regulated learning skills (Greveson & Spencer, 2005). Second, is there a positive relation between self-regulated learning skills and grade point average during medical school? We hypothesize that there is, since it is expected that self-regulated learners are more efficient learners (Bjork, et al., 2013) and are academically more successful (Ablard & Lipschultz, 1998; Nota, 2004; Toering, et al., 2009; Turan & Konan, 2012; Zimmerman, 1986).

METHODS

Setting

The present study was performed at Erasmus MC Medical School, Rotterdam, the Netherlands. The integrated and theme-oriented curriculum at this school comprises a three year Bachelor programme followed by a three year Master programme. The Bachelor of Medicine is divided into thematic blocks of 4–16 weeks, which are organized around pathophysiological systems and each theme covers subjects ranging from the basic sciences up to and including clinical practice. The Bachelor includes both lectures and small-group learning. While the lectures are voluntary, participation in about a quarter of the small-group sessions is compulsory. Skills such as planning, monitoring, evaluation, and reflection are specifically addressed during education, but students are not assessed on these skills.

Participants

Two cohorts of first-year students and one cohort of third-year students from the Erasmus MC Medical School were included in the present study. In total, 1161 students were approached to participate, of whom 949 completed the questionnaire, giving an overall response rate of 81.7%. The mean age of the 949 respondents was 20.2 years (*SD* = 2.2 years); 582 respondents were female (61.3%) and 367 were male, a division similar to the total medical school student population at Erasmus MC Medical School. Of the two cohorts of first-year students, 595 out of 803 responded (73.9%) and of the third-year students, 354 out of 358 responded (98.9%). The mean age of the first-year students was 19.1 years (*SD* = 1.9) and 360 of them were female (60.5%) and 235 were male. The mean age of the third-year students was 21.5 years (*SD* = 2.0), 222 of them were female (62.7%) and 132 were male.

Questionnaire: Self-Regulation of Learning – Self-Report Scale

The Self-Regulation of Learning – Self-Report Scale (SRL-SRS) was used to investigate the students' level of self-regulated learning. The SRL-SRS contains 50 items on a 4 or 5-point Likert scale, depending on the subsection of the questionnaire. Following the theory described by Ertmer and Newby (1996) and Hong and O'Neill (2001), the questionnaire comprises six subscales of original English-language questionnaires: planning, monitoring, evaluation, reflection, effort, and self-efficacy. An example of a question in the subscale monitoring is: "*While making an assignment, I check my progress,*" and an example from the subscale effort is: "*I keep trying to finish my assignment, even when I find the assignment extremely difficult*." The questionnaire has been compiled and validated in a Dutch study (Toering, et al., 2012). The questionnaire was originally created for secondary school students. Therefore, in the present study, minor changes were made in a few questions, e.g., the term homework was replaced by study assignments.

Measurements of academic performance

First-year and third-year grade point average (GPA) were used to investigate the correlation between students' level of self-regulated learning skills and their academic performance. Grades were given on a 10-point scale (1 = poor, 10 = excellent) where 5.5 was the cut-off for passing the course. First-year GPA was calculated from all first-year exams taken by the students. Scores obtained on resits were not taken into account, including the ones for those students whose resit was their first attempt. For this study, students were divided in quartiles based on their GPA; the 25% students with the lowest GPA, the 25% students with a slightly higher GPA and so on. The first-year quartiles were Q1: GPA < 5.4; Q2: GPA 5.4-5.9; Q3: GPA 6.0-6.5 and; Q4: GPA > 6.5. Third-year GPA was calculated from all obligatory third-year exams taken, since here 25% of the curriculum existed of chosen courses, i.e. the minors. The third-year GPAs were slightly higher and were therefore divided as follows: Q1: < 5.8; Q2: 5.8-6.3; Q3: 6.4-6.9 and; Q4: > 6.9.

Procedure

The questionnaire was integrated as an assignment in both a first-year and in a third-year course. Students were notified in advance about the use of their data for this study. All students received an e-mail and at most two reminders with a personal link plus a deadline to complete the questionnaire online in LimeSurvey 1.19+ (Schmitz, 2012). Students had to complete the questionnaire at home, and those who failed to complete it before the deadline were excluded from this study. The questionnaire took approximately 20 minutes to complete. After completion, the students received a personal report which included background information on self-regulated learning, their personal score on each subscale and the scores of their fellow students divided into quantiles to allow comparison of their score to that of their fellow students. The reported personal scores were discussed in small-group meetings under the supervision of a tutor to provide the students insight into their own study behaviour. The scores were used to investigate the development of students' self-regulated learning skills between the first and third year of medical school and to determine the association between self-regulated learning and academic performance. Data on academic performance was obtained from the university administrative system. All data were made anonymous.

Data analysis

Data were analysed with the use of IBM SPSS AMOS 18.0 (SPSS, Inc., Chicago, IL, USA) and IBM SPSS Statistics 21.0 (SPSS, Inc., Chicago, IL, USA). Confirmatory Factor Analysis and Cronbach's alpha were used to investigate whether the constructs of the questionnaire fitted the model and to measure the internal consistency of the factors. A one-way ANOVA was performed to compare the level of self-regulated learning skills of the first and third-year medical students, a p-value of $< .05$ was considered significant. For the subscale reflection, *Welch F* was calculated since equal variances could not be assumed. Effect sizes, eta-squared, were converted where .01, .06 and .14 indicate a small, medium and large effect (Cohen, 1988; Lakens, 2013). The correlation between the self-regulated learning skills and the measures of academic performance were calculated with Pearson correlations and multinomial logistic regression analysis. Here, given the multiple comparisons, a more conservative p-value of $< .01$ was considered significant.

RESULTS

Validation of the questionnaire

To validate the questionnaire, a Confirmatory Factor Analysis was performed. The original six-factor model, developed by Toering et al (2012) showed a mediocre fit in the current study; CFI was .86 (Byrne, 2010), CMIN/d.f.-ratio was, with a score of 3.85, too high since this ratio is required to be less than 3.0, but the value of RSMEA, .055, was reasonable, since here a value less than .06 is required (Byrne, 2010). In the first model, factor loadings of items 5, 13, 32, 29, 28 and 27 were low. These six items belonged to the subscales planning (2), effort (1) and self-efficacy (3). By removing these items in the order of increasing factor loading, an adjusted model was obtained, which showed a good fit; a CFI of .93, a CMIN/d.f.-ratio of 2.94 and a RSMEA of .045. A summary of the χ^2 values, χ^2 differences and degrees of freedom of the adjusted model described above, compared to a model with all factor loadings constrained equal is provided in Table 1. Since two groups of students were included in the study, factor invariance had to be tested between these groups. The χ^2 difference was not significant which indicates that the same constructs were measured in the first-year students and the in third-year students.

The internal consistency of the adjusted factors was strong (see Table 2) and did not improve noteworthy when any other item within the factors was removed. Thus, the adjusted model was chosen to analyse the data.

Table 1a: Model summary of goodness of fit statistics for tests for invariance

Model description	***χ^2***	***Df***	***$\Delta\chi^2$***	***ΔDf***	***Statistical significance***
Adjusted model	*2130.065*	*725*			
Model with factor loadings constrained equal	*2960.930*	*1450*	*830.865*	*725*	*NS*

χ^2 = Chi-square; Df = degrees of freedom; $\Delta\chi^2$ = difference in chi-square between the models; ΔDf = difference in number of degrees of freedom between the two models; NS = Not significant at .05 level

Table 2: Descriptive Statistics of the 6 factors of the Self-Regulation of Learning Self-Report Scale, Reliability Coefficients and Pearson Correlations[a]

	No of items	Mean	Minimum	Maximum	1	2	3	4	5	6
1. Planning	6	16.1	6	24	(.86)					
2. Monitoring	6	17.8	9	24	.417**	(.79)				
3. Evaluation	8	29.9	17	40	.484**	.614**	(.79)			
4. Reflection	5	17.4	5	25	.033	.067*	.104**	(.96)		
5. Effort	8	22.7	8	32	.355**	.385**	.446**	.031	(.83)	
6. Self-efficacy	7	21.0	12	28	.245**	.227**	.186**	.075*	.197**	(.80)
7. GPA year 1[¤]		2.6	1	4	.070	.144**	.090*	.209**	.148**	.048
8. GPA year 3[¤]		2.5	1	4	.133*	.168**	.180**	-.080	.247**	.037

[a]Numbers in parentheses are Cronbach's alpha
[¤]GPA; 1 = lowest 25%, 4 = highest 25% of the students included in this study
$^*p < .05$, $^{**}p < .01$

Table 3: Only the level of reflection is significantly different between the first and third year

	Planning	Monitoring	Evaluation	Reflection	Effort	Self-efficacy
Year 1 Mean ± SD	16.0 ± 3.4	17.9 ± 3.1	29.9 ± 3.7	15.7 ± 6.3	22.7 ± 3.9	20.9 ± 3.1
Year 3 Mean ± SD	16.1 ± 3.3	17.6 ± 3.0	29.9 ± 4.0	20.4 ± 3.5	22.7 ± 3.9	21.2 ± 3.2
Test value	*F* = .001	*F* = 2.272	*F* = .001	*F** = 221.918**	*F* = .002	*F* = 1.621
η^2				.152		

SD = Standard Deviation
*Test value = F-ratio with (1, 947) degrees of freedom, * = Welch F with (1,944.417) degrees of freedom, ** =significant at .05 level*
η^2 = eta-squared, with .01, .06 & .14 indicating small, medium and large effects.

Change of self-regulated learning skills

Table 3 shows the results of the comparison of the level of self-regulated learning skills of the first and the third-year medical students. On reflection, third-year students (M = 20.4) scored significantly higher than first year students (M = 15.7); *Welch F* (1, 944.417) = 221.918, $p < .001$, $\eta^2 = .152$, with the effect size indicating a large effect. No differences in level of self-regulated learning skills were found on the other subscales between first and third-year medical students.

Correlation with academic performance

Pearson correlation showed multiple significant relations between measures of academic performance in both year one and year three and the self-regulated learning skills, as reported in Table 2. To gain more insight into the direction and strength of the relations, a multinomial regression analysis was performed. Multinomial logistic regression analyses showed that self-regulated learning skills explained a small proportion of the variance in GPA among first-year medical students; $R^2 = .086$, Model χ^2 (18) = 1592.612, $p < .001$ as well as some of the variance of the third-year students; $R^2 = .105$, Model χ^2 (18) = 38.735, $p = .003$ (see Table 4a and Figure 1a for first-year and Table 4b and Figure 1b for third-year). Among the first-year students, the students with the lowest GPA (< 5.4) reported significantly less reflection (B = -.099, $p < .001$) and effort (B = -.106, $p = .003$) than students with the highest GPA (> 6.5). Students with a GPA of 5.4-5.9 reported significantly less monitoring (B = -.130, $p = .01$) than students with the highest GPA. In addition, students with a GPA of 5.5-5.9 reported significantly more reflection (B = .052, $p = .01$) than students with the lowest GPA (< 5.4). Among the third-year students, the students with the highest GPA (> 6.9) reported significantly more effort than students with the lowest and second highest GPA (Q1: B = -.190, $p < .001$; Q3: B = -.130, $p = .009$). No other differences in reported use of self-regulated learning skills were found between third-year students when focusing on their GPA.

Table 4b: The importance of effort for academic performance in year 3

Third-year GPA	B (SE)	95% CI for Odds Ratio		
		Lower	Odds Ratio	Upper
Group 1 vs 4				
Intercept	3.483(1.727)			
Planning	.004(.061)	.891	.891	1.131
Monitoring	-.032(.075)	.835	.835	1.123
Evaluation	-.051(.061)	.844	.951	1.071
Reflection	.095(.050)	.997	1.099	1.212
Effort	-.190(.052)*	.747	.827	.915
Self-Efficacy	.048(.056)	.940	1.049	1.170
Group 2 vs 4				
Intercept	3.502(.1608)			
Planning	-.033(.056)	.867	.967	1.079
Monitoring	-.019(.070)	.856	.981	1.125
Evaluation	-.089(.056)	.821	.915	1.021
Reflection	.102(.046)	1.011	1.107	1.212
Effort	-.100(.048)	.824	.905	.994
Self-Efficacy	.029(.052)	.929	1.029	1.140
Group 3 vs 4				
Intercept	1.454(1.645)			
Planning	.028(.058)	.918	1.029	1.152
Monitoring	.022(.074)	.885	1.022	1.181
Evaluation	-.051(.059)	.848	.951	1.066
Reflection	.050(.042)	.969	1.052	1.141
Effort	-.130(.050)*	.848	.878	.969
Self-Efficacy	.062(.054)	.957	1.064	1.184

Note: R^2 =.105 (Cox & Snell), .112 (Nagelkerke). Model χ^2 (18) = 38.735, p =.003
B = slope, SD = Standard deviation, CI = confidence interval
**p < .01*

Table 4a: The importance of reflection on academic performance in year 1

First-year GPA	B (SE)	95% CI for Odds Ratio		
		Lower	Odds Ratio	Upper
Group 1 vs 4				
Intercept	3.638 (1.256)*			
Planning	-.023 (.041)	.901	.977	1.059
Monitoring	-.100 (.051)	.818	.905	1.001
Evaluation	.048 (.044)	.962	1.050	1.144
Reflection	-.099 (.020)*	.871	.906	.942
Effort	-.106 (.036)*	.839	.900	.965
Self-Efficacy	.015 (.041)	.936	1.015	1.101
Group 2 vs 4				
Intercept	2.255 (1.190)			
Planning	.039 (.041)	.960	1.040	1.127
Monitoring	-.130 (.050)*	.796	.878	.970
Evaluation	.010 (.043)	.928	1.010	1.100
Reflection	-.047 (.019)	.919	.954	.991
Effort	-.047 (.034)	.892	.954	1.020
Self-Efficacy	.015 (.040)	.939	1.015	1.098
Group 3 vs 4				
Intercept	1.387 (1.142)			
Planning	.014 (.134)	.940	1.014	1.095
Monitoring	-.095 (.049)	.826	.909	1.001
Evaluation	.022 (.042)	.942	1.022	1.110
Reflection	-.028 (.019)	.937	.972	1.008
Effort	-.047 (.033)	.894	.954	1.018
Self-Efficacy	.039 (.039)	.963	1.039	1.122

*Note: R^2 =.086 (Cox & Snell), .091 (Nagelkerke). Model χ^2 (18) = 1592.612, p < .001 B = slope, SD = standard deviation, CI = confidence interval *p < .01 GPA; 1 = lowest quartile, 4 = highest quartile of the first-year students included in this study*

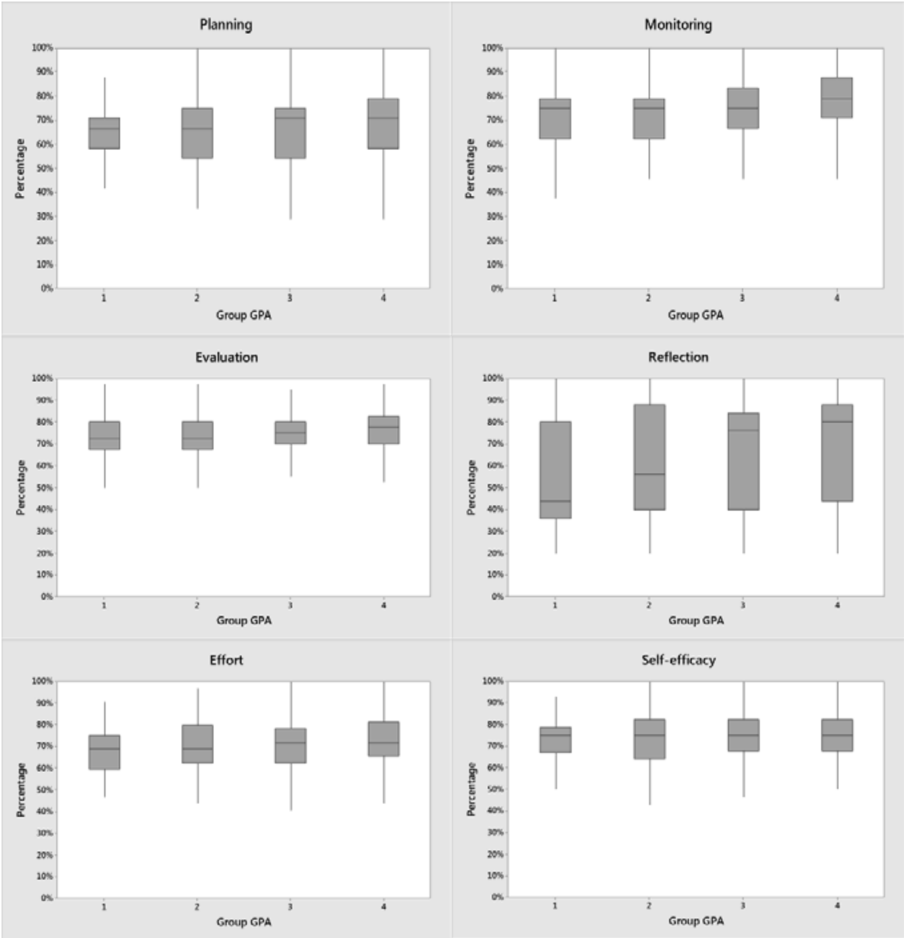


Figure 1a: Distribution of self-regulated learning skills in different first-year GPA groups

Scores on the self-regulated learning subscales are provided in percentages, where 100% equals the maximum score on the subscale The numbers 1 till 4 of GPA represent the first-year GPA of the students: Q1: GPA < 5.4; Q2: GPA 5.4-5.9; Q3: GPA 6.0-6.5 and; Q4: GPA > 6.5.

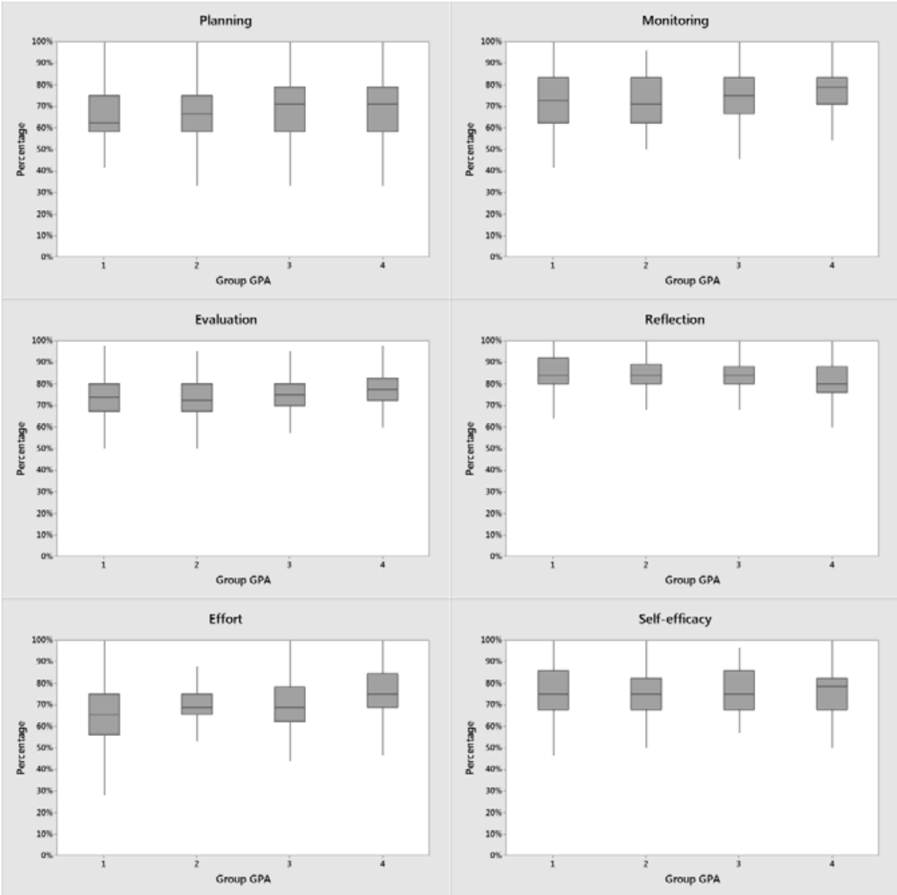


Figure 1b: Distribution of self-Regulated learning skills in different third-year GPA groups

Scores on the self-regulated learning subscales are provided in percentages, where 100% equals the maximum score on the subscale The numbers 1 till 4 of GPA represent the third-year GPA of the students: Q1: GPA < 5.8; Q2: GPA 5.8-6.3; Q3: GPA 6.4-6.9 and; Q4: GPA > 6.9.

DISCUSSION

In the present study, we attempted to answer two questions. Firstly, do students' self-regulated learning skills change between the first and third year of medical school? Secondly, is there a relation between students' self-regulated learning skills and their academic performance? Concerning the first question, we hypothesized that students' self-regulated learning skills would change during medical school. However, we found that the levels of most self-regulated learning skills did not differ between the first and third year at medical school, except reflection, which was higher in the third year. Although in the curriculum self-regulated learning skills are emphasized, the overly structured character of the curriculum leaves little room for the students to develop and apply these skills (Premkumar, et al., 2013). Nevertheless, this finding was unexpected. Not only have other studies reported that more mature students, as our third-year students are compared to our first-year students, showed higher levels of self-regulated learning (Kell & van Deursen, 2003; Premkumar, et al., 2013; Reio & Davis, 2005), some researchers also showed a positive development of the use of self-regulated learning skills within just 15 months after enrolment in university (Downing, Kwong, Chan, Lam, & Downing, 2009). It is, however, possible that, since only the best students are accepted for medical school (Razack, Maguire, Hodges, & Steinert, 2012) these students already score relatively high at entrance, and therefore show little development of self-regulated learning during medical school itself (i.e. ceiling effect). It would be interesting to measure self-regulated learning skills in a more heterogeneous student population, where entry requirements are not as high as for medical school, in order to investigate if their level of self-regulated learning skills at start of university leaves more room for improvement.

Another explanation could be that most people have a strong assumption that children and adults do not need to be taught how to learn and how to manage their learning behaviour (Bjork, et al., 2013). It is often expected that everyone will gradually acquire learning skills during school, at home and in other situations (Bjork, et al., 2013). Research showed that self-regulated learning skills can be taught, but they have to be specifically emphasized (Hong & O'Neil, 2001; Zimmerman, 1989) and students need to be provided with appropriate instructions (Brydges, etal., 2015). It might be that medical schools too easily assume that students develop these self-regulated learning skills themselves, and therefore not explicitly teach their students how to do so. Further, people often have a flawed mental model of how they learn and remember (Bjork, et al., 2013) and tend to overestimate their self-regulated learning skills (Zimmerman, 2008), especially when they do not have knowledge of the criteria and standards of good performance (Kostons, van Gog, & Paas, 2012). It is possible that first-year students overestimated their use of self-regulated learning skills more than third-year students, and thus, reported a higher use of self-regulated learning skills. The question, however, remains why there is one skill that does change, i.e. reflection. Future research with a longitudinal design is required to gain more insight into the thoughts of students on self-regulated learning.

Our second hypothesis was that high level self-regulated learners would achieve higher grades during medical school than low level self-regulated learners, since many studies have shown that self-regulated learners are more effective learners (Nota, 2004; Toering, et al., 2009), who attain higher grades in secondary school (Nota, 2004) and university (Ablard & Lipschultz, 1998). The present study confirmed that some variation in performance could be explained by the students' self-regulated learning skills, both in the first-year and in the third-year, but a large part of the variation remained unexplained. In the first-year, a relation was found between academic performance and reflection, effort and monitoring. The finding that the first self-regulated learning skill, reflection, is important for academic performance is not surprising, since reflection is the key to transform knowledge about learning into behaviour (Ertmer & Newby, 1996). Previous research showed that learners who display greater control of their learning are academically more successful (Wolters, 1998; Zimmerman, 1986). Reflection allows learners to make changes in their future learning behaviour. Reflection is seen as the centre of learning, and the more people reflect the more automatic and efficient the entire learning process becomes.

The second self-regulated learning skills that was related to first-year academic performance was effort. According to Hong and O'Neill (Hong & O'Neil, 2001), effort is necessary to actually use the other self-regulated learning skills one possesses. Effort is crucial to reach the goals a learner has set (Hong & O'Neil, 2001) and is required to persist on difficult tasks (Hong & O'Neil, 2001; Pintrich & de Groot, 1990). It is therefore not surprising that first-year students with higher levels of effort obtain higher grades. Regards to monitoring, the third skill, not the students with the lowest GPA reported the lowest level, but those with the second lowest GPA did. This could be the result of the so called Kruger-Dunning effect; poorly performing learners rarely monitor their learning and consequently are unlikely to notice that they are not doing so (Ertmer & Newby, 1996; Kostons, et al., 2012; Kruger & Dunning, 1999; Langendyk, 2006). This latter finding could explain why we did not find a lower level of monitoring in the poorest performing students, but we did find it in those students who did perform a little better. These students were perhaps more aware of the fact that they did not monitor their learning very well.

In the third year of medical school, only effort was to some extent related to performance differences. As described before, effort can be seen as the perseverance to reach goals and accomplish tasks and assignments. Perseverance is especially required when examinations and assignments get more difficult, and when students need to make a bigger effort to successfully complete them. Perseverance, or self-discipline, has also been put forward by other studies as the main predictor of academic performance, even surpassing IQ and previous performance (Duckworth & Seligman, 2005). It is therefore not surprising that effort stays important throughout medical school. Reflection and monitoring were not related to third-year academic performance. The third-year students all showed higher levels of reflection than the first-year students. Further research is needed to clarify why these skills are less important in the third year of medical school than in the first year. However, a large proportion of the variance in academic performance, both in the first as in the third year, was not explained by the self-regulated learning skills. Previous research already indicated that participation and attendance of lectures and skills training had the largest impact on first-year academic performance (Stegers-Jager, Cohen-Schotanus, & Themmen, 2012). In addition, the study of Stegers-Jager and colleagues (2012) showed that deep learning negatively influenced the first-year grades, and they suggested that deep-learning strategies were only valuable in combination with attendance. Future research is required to investigate whether attendance would explain the difference in performance among the students in the current study.

Although this study has identified several interesting findings regarding the change of self-regulated learning skills during medical school and their relation to academic performance, some limitations are worth mentioning. One notable limitation of this study is the use of a cross-sectional design, while a longitudinal design would have been more appropriate. Still, a cross-sectional design is deemed acceptable since the groups are comparable in age and gender, the sample size is large, the response rate is comparable, and all students attended the same medical school and the same curriculum. It is therefore appropriate to assume that all students will change in a similar way (William & Darity, 2008). A second limitation is that we investigated the difference in self-regulated learning between the first and third year at medical school, while some students' skills might change in later years once the connection with their future professional life becomes more apparent, such as during clerkships. We, however, deliberately chose to measure the change in the pre-clerkship stage since the impact of the medical school curriculum is more visible in these years, while the hospital setting might influence the students more during their clerkships, and the various hospitals might influence students' learning behaviour in different ways. A final concern is the tool we used to measure the students' self-regulated learning skills. Some students may have overestimated their self-regulated learning skills, which may have influenced the findings. Other methods previously described to measure self-regulated learning are computer tasks, thinking-aloud protocols, observations, interviews and structured diaries (Zimmerman, 2008). However, when using computer tasks, the focus is on measuring changes in self-regulated learning during performance, and not over a longer time. Moreover, the other methods are less suited for a study in a large student population such as the present, and are less effective in a cross-sectional approach. Since the SRL-SRS has been found to be a valid and reliable measure of self-regulated learning (Lucieer, et al., 2015; Toering, et al., 2012) we decided that this questionnaire was the best option to use in the present study.

CONCLUSION

Although medical schools aim to graduate medical doctors who are lifelong learners, we found that most self-regulated learning skills did not change during medical school, except for the skill reflection. Although the first-year students reported already high levels of self-regulated learning skills, many factors can negatively influence these skills. Students need to be taught how to regulate their learning behaviour, need to receive sufficient instructions and need to be supported by teachers. Medical schools should evaluate their curriculum to see to what extent they truly stimulate their students to develop self-regulated learning skills, and which aspects can hinder this development. In addition, effort, but also reflection and monitoring, explain a small part of the variance in academic performance during medical school. Future research is required to gain understanding of this limited role, and to understand which other factors are related to academic performance in medical school.

REFERENCES CHAPTER 3

Ablard, K. E., & Lipschultz, R. E. (1998). Self-regulated learning in high-achieving students: Relations to advanced reasoning, achievement goals, and gender. *Journal of Educational Psychology, 90*(1), 94-101.

Artino, A. R. J., Dong, T., de Zee, K. J., Gilliland, W. R., Waechter, D. M., Cruess, D. & Durning, S. J. (2012). Achievement goal structures and self-regulated learning: Relationships and changes in medical school. *Academic Medicine, 87* (10), 1375-1381.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Brydges, R., & Butler, D. (2012). A reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education, 46*(1), 71-79.

Brydges, R., Manzone, J., Shanks, D., Hatala, R., Hamstra, S. J., Zendejas, B., Cook, D.A. (2015). Self-regulated learning in simulation-based training: a systematic review and meta-analysis. *Medical Education, 49*(4), 368-378.

Brydges, R., Nair, P., Ma, I., Shanks, D., & Hatala, R. (2012). Directed self regulated learning versus instructor regulated learning in simulation training. *Medical Education, 46*(7), 648-656.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New York Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: NY: Routledge Academic.

Downing, K., Kwong, T., Chan, S. W., Lam, T. F., & Downing, W. K. (2009). Problem-based learning and the development of metacognition. *Higher Education, 57*(5), 609-621.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*(12), 939-944.

Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science, 24*(1), 1-24.

Greveson, G. C., & Spencer, J. A. (2005). Self-directed learning – the importance of concepts and contexts. *Medical Education, 39*(4), 348-349.

Hong, E., & O'Neil, H. F. (2001). Construct validation of a trait self-regulation model. *International Journal of Psychology, 36*(3), 186-194.

Jonker, L., Elferink-Gemser, M. T., Toering, T. T., Lyons, J., & Visscher, C. (2010). Academic performance and self-regulatory skills in elite youth soccer players. *Journal of Sports Sciences, 28*(14), 1605-1614.

Kell, C., & van Deursen, R. (2003). Does a problem-solving based curriculum develop life-long learning skills in undergraduate students? *Physiotherapy, 89*(9), 523-530.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121-132.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments.

Journal of Personality and Social Psychology, 77(6), 1121-1134.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4* (863)*,* 1-12.

Langendyk, V. (2006). Not knowing that they do not know: self-assessment accuracy of third-year medical students. *Medical Education, 40*(2), 173-179.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in pro blem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20*(4), 411-427.

Lucieer, S. M., van der Geest, J. N., Elói-Santos, S. M., Delobone de Faria, R. M., Jonker, L., Visscher, C., Rikers, R. M. J. P. & Themmen, A. P. N. (2015). The development of self-regulated learning during the pre-clinical stage of medical school: a comparison between a lecture-based and a problem-based curriculum. *Advances in Health Sciences Education*, DOI: 10.1007/s10459-015-9613-1.

Lycke, K. H., Grøttum, P., & Strømsø, H. I. (2006). Student learning strategies, mental models and learning outcomes in problem-based and traditional curricula in medicine. *Medical Teacher, 28*(8), 717-722.

Nota, L., Soresi, S. & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. International Journal of *Educational Research, 41*(3), 198-215.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic-performance. Journal of Educational *Psychology, 82*(1), 33-40.

Premkumar, K., Pahwa, P., Banerjee, A., Baptiste, K., Bhatt, H., & Lim, H. J. (2013). Does medical training promote or deter self-directed learning? A longitudinal mixed-methods study. *Academic Medicine, 88*(11), 1754-1764.

Razack, S., Maguire, M., Hodges, B. & Steinert. Y. (2012). "What might we be saying to potential applicants to medical school? Discourses of excellence, equity, and diversity on the web sites of Canada's 17 medical schools." *Academic Medicine, 87*(10): 1323-1329.

Reio, T. G., & Davis, W. (2005). Age and gender differences in self-directed learning readiness: A developmental perspective. *Age and Gender Differences, 2*(1), 40-49.

Schmitz, L. P. T. C. (2012). *LimeSurvey: An open source survey tool* (Version 19.1+). Hamburg, Germany: LimeSurvey Project.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421-442.

Stegers-Jager, K. M., Cohen-Schotanus, J., & Themmen, A. P. N. (2012). Motivation, learning strategies, participation and medical school performance. *Medical Education, 46*(7), 678-688.

Toering, T. T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J. G., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology, 10*(1), 24.

Toering, T. T., Elferink-Gemser, M. T., Jordet, G., & Visscher, C. (2009). Self-regulation and performance level of elite and non-elite youth soccer players. *Journal of Sports Sciences, 27*(14), 1509-1517.

Turan, S., & Konan, A. (2012). Self-regulated learning strategies used in surgical clerkship and the relationship with clinical achievement. *Journal of Surgical Education, 69*(2), 218-225.

William , E., & Darity, A. J. (2008). *International encyclopedia of the social sciences* (Vol. 7). Detroit: Macmillan Reference USA.

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology, 90*(2), 224-235.

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*(4), 307-313.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329-339.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64-70.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical back ground, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Susanna M. Lucieer
Jos N. van der Geest
Silvana M. Elói-Santos
Rosa M. Delbone de Faria
Laura Jonker
Chris Visscher
Remy M.J.P. Rikers
Axel P.N. Themmen

Advances in Health Sciences Education (2015)
DOI 10.1007/s10459-015-9613-1

THE DEVELOPMENT OF SELF-REGULATED LEARNING IN THE PRE-CLINICAL STAGE OF MEDICAL SCHOOL

A COMPARISON BETWEEN A LECTURE-BASED AND A PROBLEM-BASED CURRICULUM

Chapter 4

ABSTRACT

Society expects physicians to continuingly improve their competencies and to be up to date on developments in their field. Therefore, an important aim of medical schools is to educate future medical doctors to become self-regulated, lifelong learners. However, it is unclear if medical students become better self-regulated learners during the pre-clinical stage of medical school, and whether students develop self-regulated learning skills differently, dependent on the educational approach of their medical school. In a cross-sectional design, we investigated the development of 384 medical students' self-regulated learning skills with the use of theSelf-Regulation of Learning Self-Report Scale (SRL-SRS). In addition, we compared this development in students who enrolled in two distinct medical curricula: a lecture-based curriculum and a problem-based curriculum. Analysis showed that more skills decreased rather than increased during the pre-clinical stage of medical school, and that the difference between the curricula was mainly caused by a decrease in the skill evaluation in the lecture-based curriculum. These findings seem to suggest that, irrespective of the curriculum, self-regulated learning skills do not develop during medical school.

INTRODUCTION

It is widely recognized that medical doctors are expected to stay updated with developments in their field (Brydges & Butler, 2012; Greveson & Spencer, 2005), maintain and improve their competencies (Artino, et al., 2012; Brydges & Butler, 2012; Dannefer & Prayson, 2013). To be a lifelong learner, it is important to manage one's learning by actively taking control of one's learning activities (Brydges, et al., 2012; Ertmer & Newby, 1996; Lycke, et al., 2006), also known as self-regulated learning (Wolters, 1998; Zimmerman, 1989). According to Ertmer and Newby (1996), self-regulated learners are individuals who plan and organize their learning activities, set goals, monitor their progress, reflect upon, and evaluate their learning process. Hong and O'Neill (2001) added two motivational components that play an important role in self-regulated learning: effort and self-efficacy. Self-regulated learners show high levels of effort, they persist and reach the goals that they have set. Hong and O'Neill also consider self-efficacy as a key factor in the learning process, since it refers to the amount of trust someone has in his/her own potential to complete the task. Learners need to believe in their own abilities to successfully complete a task and to be motivated to use self-regulated learning skills (Hong & O'Neil, 2001; Sungur & Tekkaya, 2006). Since the need for highly self-regulated, lifelong learning medical doctors is widely recognized, medical education is based upon the premise that students should be encouraged to develop their self-regulated learning skills, since this will enable them to continue learning in their professional careers (Greveson & Spencer, 2005).

Curricula can take different forms, based on the educational approach that is used for the students' development. One approach with a long-standing history is the teacher-centred, lecture-based (LB) curriculum. Lectures are known to be an effective method to transfer a large quantity of information from instructor to student (Liker, Evans, Ulin, & Joseph, 1990). One of the prevailing alternatives to a LB curriculum is a problem-based learning (PBL) curriculum. Although variation in implementation of PBL exists, all implementations share core characteristics. That is, the education is student-centred, students study in small groups that are guided by a tutor, a problem is discussed to start the learning process and activate the students' prior knowledge, and after the discussion, students spent considerable time on self-study while only a few complementary lectures are provided (Barrows, 2002; Dolmans, Wolfhagen, & Ginns, 2010; Evensen, Salisbury-Glennon, & Glenn, 2001; Wijnia, Loyens, & Derous, 2011). PBL curricula thereby emphasize self-regulated learning and thus lifelong learning skills (Dolmans & Schmidt, 1994; Lloyd-Jones & Hak, 2004; Loyens, Rikers, & Schmidt, 2006). Several studies have investigated the development of self-regulated learning skills in PBL settings, in pre-clinical and clinical medical education but also in other disciplines, and concluded that PBL students become better self-regulated learners during their education, due to the active participation and discussions required in the PBL curriculum (Dolmans & Schmidt, 1994; Downing, et al., 2009; Lycke, et al., 2006; Vernon & Blake, 1993; Wijnia, et al., 2011). However, it has also been argued that all students will adopt some self-regulated learning skills, independent of the educational approach (Loyens, et al., 2008).

The present study was devised to answer two questions. First, do self-regulated learning skills develop in the pre-clinical stage of medical school irrespective of the curriculum? We hypothesize that they do, since medical schools aim to graduate medical doctors who are able to be lifelong learners (Greveson & Spencer, 2005). Second, do self-regulated learning skills develop more in students who are enrolled in the pre-clinical stage of a PBL curriculum than in students who are enrolled in the pre-clinical stage of a LB curriculum? Since previous studies reported contradictory findings (Dolmans & Schmidt, 1994; Loyens, et al., 2008; Lycke, et al., 2006; Vernon & Blake, 1993) it is interesting to investigate this assumption, as it would indicate that the development of self-regulated learning is better supported in PBL. Our hypothesis is that students in a PBL curriculum will show a stronger development in self-regulated learning skills during the pre-clinical stage of medical school, given that these skills are more explicitly incorporated in a PBL curriculum than in a LB curriculum (Lloyd-Jones & Hak, 2004; Loyens, et al., 2006) since PBL focuses on active engagement of students in their learning process (Dolmans & Schmidt, 1994; Evensen, et al., 2001).

METHODS

Setting

This quasi-experimental cross-sectional study took place at two different medical schools in Belo Horizonte, Brazil. The first medical school selected for this study was a medical school where education is lecture-based. The second medical school selected was one where a PBL approach has been implemented nine years ago. At both medical schools, the total duration to obtain the MD degree was twelve semesters (six years). The students admitted are of a comparable level, since medicine is a very competitive field in Brazil and only those with very high qualifications are accepted to enter medical school (Castro, 2012). In addition, both medical schools have been recently evaluated by the governmental institution in charge of accreditation of Brazilian medical schools (The Brazilian National Institute for Educational Studies and Research - Ministry of Education (INEP-MEC)). The evaluation scores are based, among others, on on-site evaluations of the courses, the quality of courses, the structure of the institution, and the student's grades on a national academic examination. Both medical schools received an average grade of 4 out of 5 (http://emec.mec.gov.br).

Participants

Four different groups were included in this study: the second and sixth-semester students of the lecture-based medical school (LB 2 & LB 6) and the second and sixth-semesters students of the problem-based medical school (PBL 2 & PBL 6). The second and sixth-semester students were chosen for this study to reduce the effect of pre-university influences and to maximize the possible influence of the medical school itself on the development of self-regulated learning during the pre-clinical stages of medical school. In total, 478 students were approached to participate in this research, of whom 384 completed the questionnaire, resulting in an overall response rate of 80.3%. The mean age of all respondents was 22.4 years (*SD* = 4.0 years, range 18-46); 59.9% was female. The response rate in LB 2 was 70.7% (111/157), with a mean age of 20.6 years

(*SD* = 3.3, range 18-36) and 54% of the respondents was female. In LB 6 the response rate was 82.7% (139/168), their mean age was 23.6 (*SD* = 3.6, range 19-40) and 60.1% of the respondents was female. In PBL 2 the response rate was 85.7% (66/77), the mean age was 21.7 (*SD*= 4.6, range 18-46) and 63.6% of the respondents was female and in PBL 6, the response rate was 89.5% (68/76) and in here, the mean age was 23.5 (*SD* = 4.3, range 20-43) with 63.2% of the respondents was female. The groups did not significantly differ in percentage of females, and the second semester and sixth-semester students were comparable in age, respectively. The lower number of students from the PBL curriculum was the result of the lower number of students enrolled in this medical school.

Instrument: Self-Regulation of Learning Self-Report Scale

The students' level of self-regulated learning skills was investigated using the Self-Regulation of Learning Self-Report Scale (SRL-SRS), which contains 50 items, each with a 4 or 5-point Likert scale, depending on the subsection of the questionnaire. The subscales are planning, monitoring, evaluation, reflection, effort, and self-efficacy, based on the definitions of self-regulated learning stated by Ertmer and Newby (1996), and Hong and O'Neill (2001). For example, a question in the subscale monitoring is: "*While making an assignment, I check my progress,*" and an example from the subscale effort is: "*I keep trying to finish my assignment, even when I find the assignment extremely difficult*." The questionnaire has been composed and validated in a Dutch study. In this study, two confirmatory factor analyses were conducted involving 601 and 600 adolescents aged 11 to 17 years to test and cross-validate the six-factor model. These analyses, plus a relative and an absolute test-retest reliability, showed the questionnaire to be a valid and reliable measure of self-regulated learning (Toering, et al., 2012). The guidelines described by Beaton and colleagues (Beaton, Bombardier, Guillemin, & Ferraz, 2000) were used to translate the questionnaire to Brazilian-Portuguese.

Procedure

Within a five week period, all second and sixth-semester students of both universities were asked to voluntarily participate in the study. The students were informed that the questionnaire was aimed at gaining more insight into their study behaviour. The students provided their e-mail address to receive a link of the online version of the questionnaire that they could immediately complete in a computer room or they received the questionnaire on paper in their classroom. The questionnaire took approximately 20 minutes to complete and a small gift was provided after completion.

Data analysis

The data were analysed with IBM SPSS AMOS 20.0 (SPSS, Inc., Chicago, IL, USA) and IBM Statistics SPSS 20.0 (SPSS, Inc., Chicago, IL, USA). Confirmatory Factor Analysis (CFA) was used to investigate whether the constructs of the questionnaire still fitted the model, since the questionnaire was translated and completed by students instead of adolescents. Cronbach's alphas were calculated to measure the internal consistency of the factors. Analysis of variance (ANOVA) was used to address the first question; to compare the level of self-regulated learning skills between the second-semester students and sixth-semester students and to analyse the over all development of self-regulated learning skills. A *p*-value of < .05 was considered significant. In addition, effect sizes, eta-squared (η^2), were converted. Values of .01, .06 and .14 indicate small, medium or large effects, respectively (Cohen, 1988; Lakens, 2013). For the second question, whether there was a difference in development in SRL skills between the two curricula, the interaction effect between semester and curriculum was analysed with an ANOVA General Linear Model. Here, also a p-value of < .05 was considered significant and eta-squared were converted. In order to gain more understanding of the results of these analyses, the difference between the two groups of second-semester students was analysed to determine whether the baseline measures differed, and both the differences within the LB curriculum and within the PBL curriculum were assessed, all with ANOVA. Again, a p-value of < .05 was considered significant and eta-squared (η^2) were converted.

Ethics

Ethical approval for this study in both medical schools was received from the ethical committee COEP/UFMG (case number CAAE - 0469.1.203.203-11). Students were informed about the study and had the choice to participate following the rules of informed consent.

RESULTS

Validation of the questionnaire

To determine the goodness of fit of the model, a confirmatory factor analysis (CFI) was performed. Missing values (< 1% of the data) were replaced by mean scores (Ludbrook, 2008). The original six-factor model showed a moderate fit; the CFI was .86 (Byrne, 2010). The values of CMIN/d.f. and RSMEA were reasonable, since a CMIN/d.f. score below 3.0 is required and a RMSEA score less than .06 (Byrne, 2010). In the original model, factor loadings of one item of the subscale planning, one of effort and two of the subscale self-efficacy were low. By removing these items in this order, we came to an adjusted model that showed a good fit; a CFI of .902, a CMIN/d.f. of 1.8 and a RSMEA of .045. Table 1a provides summary of the χ^2 values and χ^2 differences of the adjusted model described above compared to a model where all factor loading constrained equal. The χ^2 difference between these two models was not significant, indicating that the same constructs were measured across the four student groups in the present study: the second and sixth-semester students of the LB curriculum and the second and sixth-semester students of the PBL curriculum. Therefore, the adjusted model was chosen to analyse the data. The internal consistency of the adjusted factors was strong, both for the entire group as for all four separate groups (see Table 1b) and did not improve in a noteworthy way when any other item within the factors was removed.

Table 1a: The model summary of goodness of fit statistics for tests for invariance indicated that the same constructs were measured across the four student groups

Model description	χ^2	Df	$\Delta\chi^2$	ΔDf	Statistical significance
Adjusted model	1435.547	804			
Model with factor loadings constrained equal	4889.067	3216	3453.480	2408	NS

χ^2 = chi-square; $\Delta\chi^2$ = difference in chi-square between the models; ΔDf = difference in number of degrees of freedom between the two models; NS = Not significant at .05 level

Table 1b: Descriptive statistics and alpha reliabilities for the adjusted 6 factors of the Self-Regulated Learning-Self Report Scale

Factor	No. of items	Minimum	Maximum	Mean	Alpha reliability Entire group	Alpha reliability LB 2	Alpha reliability LB 6	Alpha reliability PBL 2	Alpha reliability PBL 6
Planning	7	8	28	21,3	.84	.85	.82	.80	.82
Monitoring	6	8	24	19,2	.83	.85	.82	.80	.82
Evaluation	8	17	40	33,0	.84	.87	.89	.89	.88
Reflection	5	13	25	22,5	.85	.83	.75	.80	.79
Effort	8	11	32	24,6	.83	.83	.81	.82	.86
Self-efficacy	8	13	32	23,8	.80	.82	.82	.72	.82

LB = lecture-based curriculum, PBL = problem-based curriculum 2 = semester 2, 6 = semester 6

Comparison of factor scores between and within universities and semesters

Table 2 shows the results of the comparisons of the level of self-regulated learning skills between all second and sixth-semester students, irrespective of their curriculum. The analysis shows that three of the six self-regulated learning skills developed negatively between the second and sixth semester of medical school. Thisdecrease is seen for the skills planning; $F(1,383) = 4.743$, $p = .030$, $\eta^2 = .012$; monitoring; $F(1,383) = 7.987$, $p = .005$, $\eta^2 = .021$; and evaluation; $F(1,383) = 9.250$, $p = .003$, $\eta^2 = .024$, although the sizes of the effects are small. The level of the other self-regulated learning skills, reflection, effort and self-efficacy, did not change between the second and sixth semester.

Table 2: The self-regulated learning skills planning, monitoring and evaluation slightly decrease between the second semester and sixth semester

	N	Planning	Monitoring	Evaluation	Reflection	Effort	Self-Efficacy
Mean semester 2	177	21.7	19.7	33.8	22.3	25.0	23.6
Mean semester 6	207	20.9	18.8	32.4	22.7	24.3	23.9
Test value (*F*)		4.743	7.987	9.250	2.196	3.236	.434
***p*-value**		.030	.005	.003	.139	.073	.510
η^2		.012	.021	.024			

η^2 = effect size, where .01, .06 and .14 indicate small, medium and large effect

A p-value < .05 is considered significant

Differences in development of self-regulated learning skills between curricula

To determine the difference in development of self-regulated learning skills between the different curricula, the interaction effect of semester and curriculum on each of the six subscales was assessed with an ANOVA. The results are displayed in Table 3a. The only, large, significant interaction effect was found on the skill evaluation; $F(1,382) = 6.718$, $p = .010$, $\eta^2 = .256$. Analyses of the difference between second-semester students of the LB curriculum and the PBL curriculum, as well as of the difference between the second and sixth semester students within each curriculum, show that this effect is caused by a decrease in the subscale evaluation in the LB curriculum; $F(1,249) = 15.506$, $p < .001$, $\eta^2 = .059$ (see Table 3b).

Table 3a: The self-regulated learning skill evaluation develops differently in the two curricula

	Planning	Monitoring	Evaluation	Reflection	Effort	Self-Efficacy
Test value (F)	.399	2.118	6.718	3.350	1.849	1.179
***p*-value**	.528	.146	.010	.068	.175	.278
η^2			.256			

η^2 = effect size, where .01, .06 and .14 indicate small, medium and large effects
A p-value < .05 is considered significant

Table 3b: The difference in the development of the skill evaluation between the curricula is caused by a negative development of this skill in the LB curriculum

	N	Planning	Monitoring	Evaluation	Reflection	Effort	Self-Efficacy
Mean LB 2	111	21.6	19.7	33.7	22.3	24.2	23.5
Mean PBL 2	66	22.0	19.8	33.9	22.4	26.3	23.8
Test value (F)		.439	.029	.085	.053	12.112	.200
***p*-value**		.509	.865	.771	.819	.001	.656
η^2						.065	
Mean LB 2	111	21.6	19.7	33.7	22.3	24.2	23.5
Mean LB 6	139	20.6	18.5	31.5	22.4	23.9	23.5
Test value (F)		4.415	9.664	15.506	.039	.285	.002
***p*-value**		.037	.002	< .001	.845	.594	.964
η^2		.018	.038	.059			
Mean PBL 2	66	22.0	19.8	33.9	22.4	26.3	23.8
Mean PBL 6	68	21.5	19.5	34.1	23.3	24.9	24.7
Test value (F)		.516	.205	.111	6.529	4.272	1.883
***p*-value**		.474	.651	.740	.012	.041	.172
η^2					.047	.031	

η^2 = effect size, where .01, .06 and .14 indicate small, medium and large effects, a p-value < .05 is considered significant

DISCUSSION

The first aim of the present study was to investigate the development of medical students' self-regulated learning skills during the pre-clinical stage of medical school. Instead of the expected increase in skills, we found that three self-regulated learning skills, e.g. planning, monitoring and evaluation, slightly decreased between the second and sixth semesters, while other skills did not change. However, since previous research showed a positive development of students' use of self-regulated learning skills within just 15 months of enrolment (Downing, et al., 2009), we were surprised to find no or even a negative development in two years. The scores on the subscale reflection were already quite high during the second semester and could therefore not improve much anymore, but the scores for other skills showed room for development during medical school.

Studies have suggested that the development of self-regulated learning skills can be distorted by various factors (Frambach, Driessen, Chan, & van der Vleuten, 2012; Moust, Berkel, & Schmidt, 2005; Schmidt, Rotgans, & Yew, 2011). Students need to have time to learn how to self-regulate their learning (Bjork, et al., 2013). Busy time schedules could obstruct this process since they severely constrain the amount of time students can invest in the development of self-regulated learning skills (Schmidt, et al., 2011). It has also been shown that there is a negative correlation between the use of self-regulated learning skills and the dependence on a teacher or tutor; when students receive more structured education, they use less self-regulated learning skills (Premkumar, et al., 2013). Finally, research has indicated that there are mixed findings on how accurate students' reports of self-regulated learning are. That is, they tend to overestimate their self-regulated learning skills (Zimmerman, 2008), and it might be the case that the second-semester students overestimated their skills more than the sixth-semester students, which could be an explanation why we did not find a development in self-regulated learning skills. These issues require more attention in future research. Another suggestion for further research is to include fourth-semester students, in addition to second and sixth-semester students, when investigating the development of self-regulated learning skills in medical education. The inclusion of fourth-semester students would provide more accurate information on how self-regulated learning evolves during medical school. For instance, when does the negative development of some self-regulated learning skills begin?

For the second aim of this study, we compared the development of self-regulated learning skills between the LB and a PBL curriculum. Here, we found a difference in the development of the skill evaluation between the two curricula. A closer look showed that this difference was caused by a negative development of evaluation in the LB curriculum. Next to this, in the LB curriculum, sixth-semester students scored lower on planning and monitoring than second-semester students. In the PBL curriculum, second-semester students started with a relative high level of effort but this level dropped a little. Interestingly, the level of reflection did slightly develop in the PBL curriculum. Thus, although some skills developed negatively and the skill reflection increased a little in the PBL curriculum, we found, unlike previous studies (Dolmans & Schmidt, 2006; Lycke, et al., 2006; Vernon & Blake, 1993), only a significant difference in the development of the self-regulated learning skill evaluation between the curricula, and this difference is not caused by a positive development but by a negative development in the LB curriculum. An explanation as to why the self-regulated learning skills did not develop more in the PBL curriculum, may be that there are differences between the educational psychology principles of PBL and their implementation in a curriculum (Moust, et al., 2005). In fact, there are some differences in definitions and implementations of PBL (Ates & Eryilmaz, 2010; Charlin, Mann, & Hansen, 1998) and the implementation can even change over time within universities (Baroffio, Vu, & Gerbase, 2013). As a result, emphasis on self-regulated learning can differ between PBL curricula. In addition, the organization of the education, such as the quality of the learning material (Azer, Guerrero, & Walsh, 2013; Baroffio, et al., 2013; Peterson, 2004) and the role of the tutor, are crucial for the successful implementation of PBL (Ates & Eryilmaz, 2010; Azer, et al., 2013; Baroffio, et al., 2013; Peterson, 2004). In addition, PBL can evoke feelings of uncertainty among students. They may respond to this uncertainty by relying upon other students' advice for planning and studying and consequently do not enough use and develop their own self-regulated learning skills (Baroffio, et al., 2013).

Culture also seems to influence the effect of PBL (Frambach, et al., 2012). Since it has been argued that self-regulated learning relies on Western principles, Frambach and colleagues investigated the cultural effect on the implementation of PBL in the Netherlands, Hong Kong, and the Middle East (Frambach, et al., 2012). They showed that students from the Middle East were more uncertain than students from the Netherlands and Hong Kong, and that these students felt lost in the PBL system. In addition, a strong teacher-centred type of secondary education, as in Brazil, may hinder students' development of self-regulated learning (Frambach, et al., 2012; Moust, et al., 2005). However, in Frambach's study, all students appeared to have internalized the concept of self-regulated learning between the first and third year (Frambach, et al., 2012), while we were not able to conclude this for the Brazilian students. Future research should therefore examine the quality of the implementation of PBL and the cultural effects in other countries as well.

This study is not without limitations. One of these limitations is the use of a cross-sectional design, as a longitudinal design might have been more informative. However, since the groups are comparable in age and gender, the sample sizes are large, and response rates are comparable, we consider a cross-sectional design to be adequate. Another limitation is that we could not completely control for differences between the two universities. Although both medical school received the same evaluation score from The Brazilian National Institute for Educational Studies and Research – Ministry of education, it is possible that the student groups were different, since one university was publicly funded (LB) and the other was privately funded (PBL). However, this difference was most likely small. Medicine is one of the most competitive educational areas in Brazil with 150 applicants competing for one position. As a consequence, only those applicants with the highest scores on a national admission exam (ENEM/SISU; exam and scores are provided by the Brazilian Ministry of Education) are accepted for medical school (Castro, 2012). In addition, except for the skill effort, the students' self-regulated learning skills were comparable in the second semester, which indicates that the level of self-regulated learning of the second semester students was comparable. A third limitation is that some students completed the questionnaire on paper while others completed the questionnaire online. However, this limitation is relatively minor since the circumstances under which the questionnaire was completed were equal; they were all completed in a classroom at the medical school, and research indicated that paper and online versions of questionnaires can be taken as equivalent (Vallejo, Jordán, Diaz, Comeche, & Ortega, 2007). Another limitation is that we investigated the development in self-regulated learning skills between second and sixth-semester students, and not between first and sixth-semester students. One could argue that excluding the first semester may have influenced our findings. The decision to not include the first, but the second semester students was deliberate, since we wanted the students to report based on their experience in medical school when answering the questions, and not on the very intensive study behaviour they must have shown to pass their entry exams. In addition, several studies have shown that many PBL students initially have difficulties in adjusting to the programme. They lack confidence in active learning and rely heavily on their tutors, instead of defining their own learning goals (Lee, Mann, & Frank, 2010; Miin, Campbell, & Price, 1999; Schmidt, van der Arend, Moust, Kokx, & Boon, 1993), which might have been less in the second semester. Apart from this, only the subscale effort and none of the other subscales were different between the second-semester students, which indicated that the baseline was equal. We thus do not expect that our findings are biased due to the inclusion of second-semester students. A final limitation is that our group sizes are different. This is however a result of the difference in places available to study medicine in both universities and did most likely not influence the findings in this study.

CONCLUSION

Although medical schools, and in particular those with a PBL curriculum, are based upon the premise that their graduates should become self-regulated learners, our study shows that most self-regulated learning skills of medical students do not develop during the pre-clinical stage of medical school, while some skills even decrease. This can be the result of a lack of stimulation of self-regulated learning due to the structure of the pre-clinical curriculum. Especially in a PBL setting, the quality of the implementation of the curriculum and the cultural background of the students may have influenced the development of the self-regulated learning skills. In sum, medical schools should carefully consider the actions they take to improve the development of self-regulated learning skills in their students.

REFERENCES CHAPTER 4

Artino, A. R. J., Dong, T., de Zee, K. J., Gilliland, W. R., Waechter, D. M., Cruess, D. & Durning, S. J. (2012). Achievement goal structures and self-regulated learning: Relationships and changes in medical achool. *Academic Medicine, 87*(10), 1375-1381.

Ates, O., & Eryilmaz, A. (2010). Factors affecting performance of tutors during problem-based learning implementations. *Procedia - Social and Behavioral Sciences, 2*(2), 2325-2329.

Azer, S. A., Guerrero, A. P. S., & Walsh, A. (2013). Enhancing learning approaches: Practical tips for students and teachers. *Medical Teacher, 35*(6), 433-443.

Baroffio, A., Vu, N., & Gerbase, M. (2013). Evolutionary trends of problem-based learning practices throughout a two-year preclinical program: a comparison of students' and teachers' perceptions. *Advances in Health Sciences Education, 18*(4), 673-685.

Barrows, H. (2002). Is it truly possible to have such a thing as dPBL? *Distance Education, 23*(1), 119-122.

Beaton, D. E., Bombardier, C. M. D., Guillemin, F. M. D., & Ferraz, M. B. M. D. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Brydges, R., & Butler, D. (2012). A reflective analysis of medical education research on self-regulation in learning and practice. *Medical Education, 46*(1), 71-79.

Brydges, R., Nair, P., Ma, I., Shanks, D., & Hatala, R. (2012). Directed self regulated learning versus instructor regulated learning in simulation training. *Medical Education, 46*(7), 648-656.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. New York Routledge.

Castro, C. D. M. (2012). *Busting myths of Brazilian higher education*: Presidency of the Positivi Group.

Charlin, B., Mann, K., & Hansen, P. (1998). The many faces of problem-based learning: a framework for understanding and comparison. *Medical Teacher, 20*(4), 323-330.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: NY: Routledge Academic.

Dannefer, E. F., & Prayson, R. A. (2013). Supporting students in self-regulation: Use of formative feedback and portfolios in a problem-based learning setting. *Medical Teacher, 35*(8), 655-660.

Dolmans, D. H. J. M., & Schmidt, H. G. (1994). What drives the students in problem-based learning? *Medical Education, 28*(5), 372-380.

Dolmans, D. H. J. M., & Schmidt, H. G. (2006). What do we know about cognitive and motivational effects of small group tutorials in problem-based learning? *Advances in Health Sciences Education, 11*(4), 321-336.

Dolmans, D. H. J. M., Wolfhagen, I. H., & Ginns, P. (2010). Measuring approaches to learning in a problem based learning context. *International Journal of Medical*

Education, 1, 55-60.
Downing, K., Kwong, T., Chan, S. W., Lam, T. F., & Downing, W. K. (2009). Problem-based learning and the development of metacognition. *Higher Education, 57*(5), 609-621.
Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science, 24*(1), 1-24.
Evensen, D. H., Salisbury-Glennon, J. D., & Glenn, J. (2001). A qualitative study of six medical students in a problem-based curriculum: Toward a situated model of self-regulation. *Journal of Educational Psychology, 93*(4), 659-676.
Frambach, J. M., Driessen, E. W., Chan, L.C., & van der Vleuten, C. P. M. (2012). Rethinking the globalisation of problem-based learning: how culture challenges self-directed learning. *Medical Education, 46*(8), 738-747.
Greveson, G. C., & Spencer, J. A. (2005). Self-directed learning – the importance of concepts and contexts. *Medical Education, 39*(4), 348-349.
Hong, E., & O'Neil, H. F. (2001). Construct validation of a trait self-regulation model. *International Journal of Psychology, 36*(3), 186-194.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863)*,* 1-12.
Lee, Y. M., Mann, K. V., & Frank, B. W. (2010). What drives students' self-directed learning in a hybrid PBL curriculum. *Advances in Health Sciences Education, 15*(3), 425-437.
Liker, J. K., Evans, S. M., Ulin, S. S., & Joseph, B. S. (1990). The strengths and limitations of lecture-based training in the acquisition of ergonomics knowledge and skill. *International Journal of Industrial Ergonomics, 5*(2), 147-159.
Lloyd-Jones, G., & Hak, T. (2004). Self-directed learning and student pragmatism. *Advances in Health Sciences Education, 9*(1), 61-73.
Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review, 20*(4), 411-427.
Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2006). Students' conceptions of constructivist learning: A comparison between a traditional and a problem-based learning curriculum. *Advances in Health Sciences Education, 11*(4), 365-379.
Ludbrook, J. (2008). Outlying observations and missing values: How should they be handled? *Clinical and Experimental Pharmacology and Physiology, 35*(5-6), 670-678.
Lycke, K. H., Grøttum, P., & Strømsø, H. I. (2006). Student learning strategies, mental models and learning outcomes in problem-based and traditional curricula in medicine. *Medical Teacher, 28*(8), 717-722.
Miin, B. M., Campbell, C. B. & Price, D. A. (1999). A lesson from the introduction of a problem-based, graduate entry course: the effects of different views of self-direction. *Medical Education, 33*(11), 801-807.
Moust, J. H. C., Berkel, H. J. M. V., & Schmidt, H. G. (2005). Signs of erosion: Reflections on three decades of problem-based learning at Maastricht university. *Higher Education 50*(4), 665-683.

Peterson, T. O. (2004). So you're thinking of trying problem based learning?: Three critical success factors for implementation. *Journal of Management Education, 28*(5), 630-647.

Premkumar, K., Pahwa, P., Banerjee, A., Baptiste, K., Bhatt, H., & Lim, H. J. (2013). Does medical training promote or deter self-directed learning? A longitudinal mixed-methods study. *Academic Medicine, 88*(11), 1754-1764.

Schmidt, H. G., Rotgans, J. I., & Yew, E. H. J. (2011). The process of problem-based learning: what works and why. *Medical Education, 45*(8), 792-806.

Schmidt, H. G., van der Arend, A., Moust, J. H. C., Kokx, I., & Boon, L. (1993). Influence of tutors' subject-matter expertise on student effort and achievement in problem-based learning. *Academic Medicine, 68*(10), 784-791.

Sungur, S., & Tekkaya, C. (2006). Effects of problem-based learning and traditional instruction on self-regulated learning. *Journal of Educational Research, 99*(5), 307-317.

Toering, T. T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J. G., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology, 10*(1), 24.

Vallejo, M. A., Jordán, C. M., Diaz, M. I., Comeche, M. I., & Ortega, J. (2007). Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the general health questionnaire-28 (GHQ-28) and the symptoms check-list-90-revised (SCL-90-R). *Journal of Medical Internet Research, 9*(1:e2).

Vernon, D. T., & Blake, R. L. (1993). Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine, 68*(7), 550-563.

Wijnia, L., Loyens, S. M. M., & Derous, E. (2011). Investigating effects of problem-based versus lecture-based learning environments on student motivation. *Contemporary Educational Psychology, 36*(2), 101-113.

Wolters, C. A. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology, 90*(2), 224-235.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329-339.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Susanna M. Lucieer
Axel P.N. Themmen
Robert Jan Stolker
Matthijs de Hoog
Yvo M. Smulders
Abel Thijs
Jan N.M. IJzermans
Remy M.J.P. Rikers

Sumitted for publication

INDENTIFYING DELIBERATE PRACTICE ACTIVITIES IN RESIDENCY TRAINING

Chapter 5

ABSTRACT

Introduction

The concept of deliberate practice has been defined as a set of behavioural characteristics leading to the acquisition of expertise. In the medical domain, studies on the role of deliberate practice in attaining clinical expertise have largely been limited to surgery and medical diagnosis. One of the limitations in these studies has been the difficulty to obtain a clear outcome measure, i.e. a measure of expertise. In the present study, the principles of deliberate practice have been used to identify behavioural characteristics, of residents in four different specialty programmes that are related to the attainment of clinical expertise as assessed by residency supervisors.

Methods

Residents from anaesthesiology, internal medicine, paediatrics and surgery training programmes were asked to complete a survey with questions measuring aspects of deliberate practice (N = 90). To obtain an expertise outcome measure, training supervisors were asked to assess the residents' performance on an scale from 0-100. Based on supervisor score, the residents were divided into tertiles. Correlations between survey items and supervisor assessment were calculated and differences in scores on survey items of the tertiles were analysed.

Results

Positive correlations were found between supervisors' assessment and survey items relating to feedback and repetition. Comparison of the tertiles revealed that residents in the top two tertiles more often acknowledged that they made mistakes, asked for eedback and introduced repetitions in their practice than residents who were in the bottom tertile.

Discussion/Conclusion

Our study identified several differences in behavioural characteristics relating to deliberate practice, between tertiles of global resident performance, as assessed by supervisors.

INTRODUCTION

Postgraduate clinical residency programmes aim to train specialists to a level of expertise that allows them to practice medicine as an independent professional (Long, 2000; Teunissen, et al., 2007). Some residents in the programmes outperform others and as to date it is unclear as to how these differences in performance may be explained (Ericsson, 2004). One prominent theory that aims to explain why some people excel is the theory of deliberate practice.

In the early nineties, Ericsson and colleagues showed that age and experience do not predict expert performance (Charness, 1981; Ericsson, 2008; Jastrzembski, Charness, & Vasyukova, 2006), and neither does IQ (Ericsson & Lehman, 1996). According to Ericsson, prolonged engagement in deliberate practice is required for the attainment of expertise. Deliberate practice refers to a set of highly structured activities aimed to improve performance. These activities are characterized by, 1) repetition of knowledge, skills, principles and practice, 2) seeking feedback on performance, 3) creating situations in which mistakes can be made and corrected and, 4) perform practice at a challenging but attainable level (Ericsson, 2004). Deliberate practice distinguishes itself from regular practice in the sense that the activities are challenging, require a high level of concentration and aim to improve performance or a particular skill (Charness, et al., 2005). Deliberate practice is often less enjoyable as it requires a large effort and often does not lead to immediate rewards, but is done with the goal of attaining long-term goals, improving and maintaining expert performance (Ericsson, 2004; Ericsson, 2008; Ericsson, et al., 1993).

The role of deliberate practice has been investigated in many different, complex domains of human endeavour (Ericsson, 1996). Initial studies of deliberate practice focused on expert performance in chess and music (Charness, 1981; Charness, et al., 2005; Ericsson, et al., 1993), but its importance has now also been established in other domains such as athletics, typing, surgery, and medical diagnosis setting (Cote, Ericsson, & Law, 2005; Ericsson, 1996; Gifford & Fall, 2014). The most common approach to investigate the role of deliberate practice in a domain, is to start by identifying reproducible superior performance and, subsequently, by investigating preceding activities necessary to perform at this level in a retrospective manner. Interviews and diaries can provide insight into the type of behaviour the expert performers have previously shown (Ericsson, 2009). For example, in a study on elite violinists, participants were interviewed during three sessions and asked to keep a structured diary to obtain biographical information to gain insight in the time spent on certain activities, and the relevance of these activities according to the violinists. This information was related to the level of the violinists to determine which activities related to the performance differences (Ericsson, et al., 1993). Another option is creating situations in a laboratory setting where performance can be reproduced under controlled conditions, often combined with thinking aloud protocols (Ericsson, 2004; Ericsson, 2009). In a laboratory, situations can be presented in which individuals from different performance levels can generate their best actions (Ericsson, 2007). This way, differences in performance and cognitive processes underlying these performances can be mapped out (Ericsson, 2008).

These approaches to investigate expert performance depend on carefully measured criteria or controlled conditions in order to differentiate levels of performance.However, in the clinical domain official rankings are absent and in view of patient variability and personalized care, controlled situations are very difficult, if not impossible, to create (Ericsson, 2004). These conditions make it difficult to determine which physician performs best. Nevertheless, the role of deliberate practice has been established in the clinical domain in areas where it is possible to compare performance, such as in surgery and medical diagnosis. Usually, expert performance in the clinical field means mastering broad and extensive knowledge in addition to other cognitive, motor and interpersonal skills, which makes it difficult to measure the level of performance (Ericsson, 2009). In contrast, in surgery, motor skills are very important and some researchers have used measures such as blood loss and complications after surgery as indicators of treatment success (Ericsson, 2004; Ericsson, 2009). In addition, performance can be videotaped followed by blinded assessment of the surgeon (Ericsson, 2011). Researchers were therefore able to show that there is a large variability in performance among surgeons, indicating that experience by itself is not sufficient for expert performance (Ericsson, 2004). In addition, studies on clinical diagnosis have shown that years of experience were negatively related to performance, especially in those conditions where feedback was missing, whereas continued training improved performance, i.e. led to more accurate diagnoses (Ericsson, 2004; Ericsson, 2009).

In the present study, clinical performance is investigated within the framework of deliberate practice through supervisor assessments of the residents. While residents are still in clinical training, they perform many clinical duties. In addition, since deliberate practice refers to a long-term engagement in highly structured practice, residency training presents a setting where elements of deliberate practice may already have a clear relation to the level of performance. This approach allows studying the role of deliberate practice in multiple clinical domains. Although the clinical setting is a high-stake environment where patient care is the main responsibility, we did expect to find differences between residents' performance levels and the extent to which behavioural characteristics related to deliberate practice. To investigate this hypothesis, residents were asked to complete a survey on behavioural characteristics relating to deliberate practice, and the outcome was related to their supervisor assessments.

METHODS

Participants

Participants in this study were recruited from four different specialty resident programmes in the Netherlands: anaesthesiology, internal medicine, paediatrics, and surgery. The decision to include residents from different training programs was made to create a representative sample of specialty types. In total, 211 residents were approached to complete a survey of which 141 responded. Supervisor assessments were available of 90 residents (42.8%). Consequently, all results in this study are based on these 90 residents. Of these 90 residents, 29 were in anaesthesiology training, 21 in internal medicine training, 12 in paediatric training and 28 in surgery training. Their mean age was 31.0 years (*SD* = 3.4) and 41 residents (45.6%) were male. Descriptive statistics per specialty are displayed in Table 1. Residents were in different stages of their training, ranging from year one to the final year, with a mean of 2.8 years (*SD* = 1.7). They worked on average 49.5 hours (*SD* = 6.6) per week and spent an additional 5.5 hours (*SD* = 3.8) per week on training-related activities. With respect to the use of learning resources, residents owned on average 18.9 (*SD*= 32.5) books on a clinical subject and read 3.1 (*SD* = 2.4) scientific articles per week. Regarding other learning activities, residents visited on average 2.0 (*SD* = 1.3) conventions per year and took part in 3.3 (*SD* = 2.1) courses per year.

Table 1: Descriptive statistics of the different subgroups

	N	Mean age (*SD*)	% Males
Anaesthesiology	29	30.2 (*4.3*)	62.1
Internal medicine	21	31.1 (*3.9*)	28.6
Paediatrics	12	31.6 (*2.7*)	16.7
Surgery	28	31.4 (*2.0*)	53.6

SD = standard deviation

Materials

To create questions that measure aspects of deliberate practice in the clinical field, interviews were conducted with seven experienced medical doctors. These medical doctors have been in charge of the residency training, or contributed substantially to the residency training. Their background were in general practice, internal medicine and surgery. They were asked to describe behaviour of residents who, in their opinion, performed at different levels in several work-related situations. This information was used by three investigators (SML, APT & RMR) to develop a survey with questions on behavioural activities relating to different deliberate practice activities, including repetition of knowledge and skills, principles and practice, seeking feedback on performance, creating situations in which mistakes can be made and corrected, and practice at a challenging but attainable level. Medical doctors and supervisors in anaesthesiology, internal medicine, paediatrics and surgery checked the survey to ascertain that the described situations are applicable to their specialty afterwards. Participants completed the survey consisting of questions on: (1) demographics and experience in the clinical field, including year and type of specialty training, the number of learning resources and other educational activities undertaken, and (2) behavioural activities related to the main characteristics of deliberate practice. The different questions are displayed in Table 2. Questions were evaluated on a 5-point Likert scale (1 = completely disagree, 5 = completely agree) and the entire survey took a maximum of 20 minutes to complete.

Outcome variable: global performance score

Participants were assessed on overall performance by three to six supervisors (*M* = 4.56) involved in their specialty training to create a global performance score. The interrater reliability ranged from .564 - .691 (ICC; one-way random effects model, average measures, 95% CI) which indicated that there was a moderate agreement between the supervisors on the performance levels of the residents. While checklists and existing assessment tools focus on specific skills or measure theoretical knowledge, global assessments focus on clinical performance (Bennett, Gordon, & Williamson Shaffer, 2004). To our knowledge, no global assessment tool for the clinical field has been developed yet and, according to the literature on personnel assessment and selection, overall job performance is most optimally and most frequently assessed via supervisor ratings (Schmitt, 2012). Therefore, supervisors were asked to indicate on a scale from 0-100 how much confidence they had in the resident's skills to take care of the supervisors' close relatives. Including the reference to close relatives in the instructions increased the importance that supervisors attached to the assessments. A score of 0 indicated no confidence and a score of 100 indicated complete confidence in the resident. Since residents handle increasingly complex procedures during their training (Hopmans, et al., 2014), supervisors were instructed to take into account the number of years spent in residency training, information that was readily available to them. As some supervisors used the entire scale to score their residents while others only used a small portion of this scale, assessment scores were standardized on supervisor level to correct for the natural variation in scoring range between supervisors (Kastner, Gore, & Novack, 1984; McGill, van der Vleuten, & Clarke, 2011, 2013). Consequently, scores provided by supervisors who assessed less than three residents, were removed (Kastner, et al., 1984). The mean z-score per resident was used as an indicator of global performance.

Procedure

The residents were approached during educational or work-related activities and either competed the survey immediately on paper, or afterwards online. Participation was voluntary, and the participating residents had a chance to win one of two e-readers available per residency programme. Residents were divided in tertiles (low, average, high performance) based on the mean standardized supervisor assessment score (c.f. Kerdijk, et al., 2013; Hammond, et al., 2007). By dividing the residents into tertiles, it was possible to compare the low performing residents with the average and high performing residents, providing insight into which specific behavioural characteristics are shown to a greater or lesser extent by the low, average, and high performing residents.

Statistics

The data were analysed with IBM SPSS Statistics 20.0 (SPSS, Inc., Chicago, IL, USA). Pearson correlations were calculated between the items of the survey, the behavioural characteristics, and mean standardized supervisors' assessment. One-way analysis of variance (ANOVA) with Bonferroni correction was used to identify differences between the tertiles on the different items of the survey. For items of which equal variances could not be assumed, *Welch F* was calculated. Statistical significance level was set $p < .05$. Effect sizes were determined using eta-squared (η^2) with values of .01, .06, and .14 indicating small, medium and large effects, respectively (Cohen, 1988; Lakens, 2013).

Ethics

All residents were allowed to ask questions and had to sign an informed consent form before they were able to complete the questionnaire. After participation, the residents were informed about the outcome measure, i.e. the performance scores provided by the supervisors, and were explained that the aim of the study was to measure behavioural characteristics relating to deliberate practice. The residents were given the opportunity to withdraw once all information was provided.

RESULTS

Descriptive statistics

The means and standard deviations of the scores on the survey items are presented in Table 2. Note that items 1, 3, 5, 7, 9, 13, 14 and 16 were negatively stated and reversed scored. Thus, on these items, a lower score indicates that more relevant behaviour related to the principles of deliberate practice is displayed. No statistically significant differences were found in demographics and experience in the clinical field between the low, average and high-performing residents (see Table 3a).

Correlations between characteristics of behaviour and supervisors' assessment

Pearson correlations were calculated to estimate the relation between the survey items and the mean standardized supervisors' assessment (see Table 2). Age and scores on item 14, a reversed item, were negatively correlated with supervisors' assessment, while scores on items 8, 19 and 20 were positively correlated with supervisors' assessment. The four correlating items relate to feedback and repetition.

Comparison between resident tertiles

Differences on item scores between the three tertiles were analysed using one-way ANOVA (see Table 3a-3e). Significant differences between groups were found on four items belonging to three subscales of the survey. The first difference was found on an item belonging to "repetition of basic knowledge and principles", item 20: "*To refresh my knowledge I frequently consult the clinical literature*"; *Welch* $F(2, 55.069) = 8.756$, $p < .001$, $\eta^2 = .14$ indicating a large effect. The lowest tertile ($M = 3.47$) scores significantly different from the highest tertile ($M = 4.17$), $p = .001$, and the average tertile ($M = 3.67$) significantly different from the high tertile ($M = 4.17$), $p = .033$. Two significant differences were found on items referring to feedback-seeking behaviour: items 14 and 19. On reversed item 14: "*I rarely use the internet or literature to check my medical actions*", the difference was; $F(2, 87) = 4.819$, $p = .010$, $\eta^2 = .10$, indicating a medium to large effect. Differences were found between the low ($M = 1.83$) and average ($M = 1.37$) tertile, $p = .015$, as well as between the low ($M = 1.83$) and high ($M = 1.43$) tertile, $p = .048$. On item 19: "*I do not avoid colleagues who provide me with feedback on my performance*", the differences was; $F(2, 87) = 3.923$, $p = .023$, $\eta^2 = .08$, indicating a medium to large effect. Here, significant differences were found between the low ($M = 4.07$) and average ($M = 4.43$) tertile, $p = .024$. Furthermore, a significant difference was found on an item belonging to the subscale "creating situations to make and correct mistakes". On item 9: "*When I have done something incorrectly, I try to hide it as good as possible*", the difference between the tertiles was; $F(2.87) = 3.320$, $p = .041$, $\eta^2 = .07$, indicating a medium to large effect. This item was reversed, and the average ($M = 2.03$) tertile scored significantly different from the high ($M = 1.60$) tertile, $p = .048$.

Table 2: Means and standard deviations on the items of the questionnaire and correlation with supervisor assessment, divided per survey subsection

	Mean	SD	Supervisor assessment®
Demographics and experience in the clinical field			
Age	30.1	3.4	-.214*
Year of training	2.8	1.7	-.019
Working hours per week	49.5	6.6	.082
Hours available for study per week	5.5	3.8	-.075
Number of books on clinical subject owned	18.9	32.5	.050
Articles read per week	3.1	2.4	.110
Conferences visited per year	2.0	1.3	.031
Participation in courses per year	3.3	2.1	-.029
Repetition of basic knowledge and principles			
1. When I have performed certain clinical procedures n times, I do not think it is sensible to keep doing them myself. ¥	1.86	.758	-.030
3. When, during morning handover, patients with clinical presentations that I am familiar with are discussed, I lose my focus. ¥	1.37	.814	-.006
4. Refresher courses provide a good opportunity to maintain skills that I rarely use.	3.86	.801	.001
6. Patient rounds are always interesting, despite the fact that I repeatedly see the same types of clinical case.	3.42	.924	.006
20. To refresh my knowledge I frequently consult clinical literature.	3.77	.794	.278**
Seeking feedback on performance			
8. It is a good thing to discuss my patients with colleagues, so that I can hear their views as well.	4.47	.580	.251*
13. I find it difficult to ask for help when things are complicated. ¥	1.78	.700	-.100
14. I rarely use the internet or literature to check my medical actions. ¥	1.54	.650	-.294**
17. I find it important that an experienced colleague is present when I treat a patient.	3.57	.750	-.186
19. I do not avoid colleagues who provide me with feedback on my performance.	4.28	.541	.264*
Creating situations to make and correct mistakes			
2. It is not a problem to make mistakes, I can only learn from them.	3.08	.939	.005
5. It is better to avoid doing things than to make mistakes while doing them. ¥	1.53	.997	.001
9. When I have done something incorrectly, I try to hide it as good as possible. ¥	1.78	.700	-.055
15. I deliberately look for situations where I can practice and where mistakes are allowed	3.43	.862	.034
18. When I have done a procedure incorrectly, I want to try it again.	4.22	.514	.110

Table 3a: Differences in demographics and experience in the clinical field between low, average and high-performing residents

		Age	Educational year	Working hours per week	Hours available for study per week	No of books on clinical subject owned	No articles read per week	No subscriptions to scientific journals	No of conventions visited per year	No of courses participated per year
Low	**Mean *(SD)***	31.3 *(4.0)*	2.7 *(1.6)*	49.4 *(7.1)*	5.8 *(4.7)*	13.5 *(10.5)*	2.9 *(2.5)*	1.2 *(.9)*	2.2 *(1.8)*	3.4 *(2.6)*
Average	**Mean *(SD)***	31.0 *(2.6)*	2.8 *(1.8)*	49.2 *(6.1)*	5.1 *(3.4)*	27.2 *(53.6)*	2.9 *(2.4)*	1.4 *(1.0)*	1.8 *(.8)*	3.1 *(1.6)*
High	**Mean *(SD)***	30.5 *(3.4)*	2.8 *(1.8)*	49.8 *(6.6)*	5.7 *(3.3)*	16.1 *(12.0)*	3.5 *(2.2)*	1.2 *(.9)*	2.1 *(1.2)*	3.4 *(1.9)*
Test value		*F* = .386	*F* = .011	*F* = .070	*Welch F* = .269	*F* = 1.529	*F* = .539	*F* = .422	*F* = .713	*F* = .190
η^2										

SD = Standard Deviation
Test value = F-ratio with (2,87) degrees of freedom, Welch F with (2,56.878) degrees of freedom
η^2 = effect size, with .01, .06 & .14 indicating small, medium and large effect

Table 3b: Differences in repetition of basic knowledge and principles between low, average and high-performing residents

		1.(R)	3.(R)	4.	6.	20.
Low	**Mean *(SD)***	1.83 *(.87)*	2.33 *(.66)*	3.90 *(.80)*	3.43 *(.82)*	3.47 *(.86)*
Average	**Mean *(SD)***	1.90 *(.61)*	2.50 *(1.01)*	3.73 *(.91)*	3.43 *(.94)*	3.67 *(.80)*
High	**Mean *(SD)***	1.83 (.79)	2.37 *(.74)*	3.93 *(.69)*	3.40 *(1.04)*	4.17 *(.53)*
Test value		*F* = .076	*Welch F* = .520	*F* = .531	*F* = .013	*Welch F* = 8.756**
η^2						.14

SD = Standard Deviation, 1,3,4,6 & 10 = survey items, R = reversed item
Test value = F-ratio with (2,87) degrees of freedom, Welch F question 3 with (2, 56.568) degrees of freedom and Welch F question 20 (2, 55.069) degrees of freedom
** = significant at .05 level, ** = significant at .01 level*
η^2 = effect size, with .01, .06 & .14 indicating small, medium and large effects

Table 3c: Differences in seeking feedback on performance between low, average and high-performing residents

		8.	13.(R)	14.(R)	17.	19.
Low	**Mean *(SD)***	4.27 *(.64)*	1.83 *(.46)*	1.83 *(.75)*	3.77 *(.57)*	4.07 *(.58)*
Average	**Mean *(SD)***	4.57 *(.50)*	1.73 *(.83)*	1.37 *(.49)*	3.53 *(.78)*	4.43 *(.50)*
High	**Mean *(SD)***	4.57 *(.57)*	*1.73 (.77)*	1.43 *(.63)*	3.40 *(.86)*	4.33 *(.48)*
Test value		*F* = 2.738	*Welch F* = .201	*F* = 4.819**	*Welch F* = 2.153	*F* = 3.923*
η²				.10		.08

SD = Standard Deviation, 8, 13, 14, 17 & 19 = survey items, R = reversed item
Test value = F-ratio with (2,87) degrees of freedom, Welch F question 13 with (2,53.678) degrees of freedom and Welch F question 17 with (2.56.081) degrees of freedom
** = significant at .05 level, ** = significant at .01 level*
η² = effect size, with .01, .06 & .14 indicating small, medium and large effects

Table 3d: Differences in situations to make and correct mistakes between low, average and high-performing residents

		2.	5.(R)	9.(R)	15.	18.
Low	**Mean *(SD)***	3.07 *(.98)*	2.53 *(.86)*	1.70 *(.54)*	3.30 *(.84)*	4.17 *(.46)*
Average	**Mean *(SD)***	3.10 *(.89)*	2.53 *(1.04)*	2.03 *(.89)*	3.63 *(.77)*	4.23 *(.57)*
High	**Mean *(SD)***	3.07 *(.98)*	2.53 *(.99)*	1.60 *(.65)*	3.37 *(.94)*	*4.17* (.52)
Test value		*F* = .012	*F* = .000	*F* = 3.320*	*F* = 1.264	*F* = .289
η²				.07		

SD = Standard Deviation, 2, 5, 9, 15 & 18 = survey items, R = reversed item
*Test value = F-ratio with (2,87) degrees of freedom, * = significant at .05 level*
η² = effect size, with .01, .06 & .14 indicating small, medium and large effects

Table 3e: Differences in performing practice at a challenging but attainable level between low, average and high-performing residents

		7.(R)	10.	11.	12.	16.(R)
Low	**Mean *(SD)***	3.30 *(.99)*	3.77 *(.68)*	4.40 *(.81)*	3.80 *(.89)*	3.07 *(.74)*
Average	**Mean *(SD)***	3.27 *(1.14)*	3.87 *(.73)*	4.70 *(.47)*	3.73 *(.87)*	2.93 *(.82)*
High	**Mean *(SD)***	3.29 *(1.07)*	4.07 *(.52)*	4.70 *(.54)*	3.87 *(1.01)*	3.03 *(.93)*
Test value		*F* = .009	*F* = 1.659	*F* = 2.317	*F* = .156	*F* = .207
η²						

SD = Standard Deviation, 7, 10, 11, 12 & 16 = survey items, R = reversed item
Test value = F-ratio with (2,87) degrees of freedom
η² = effect size, with .01, .06 & .14 indicating small, medium and large effects

DISCUSSION

To our knowledge, this is the first study investigating whether behavioural characteristics, fitting within the concept of deliberate practice, relate to performance differences between residents in the clinical domain. Twenty questions on behavioural activities referring to repetition of knowledge and skills, principles and practices, seeking feedback on performance, creating situations in which mistakes can be made and corrected, and practice at a challenging but attainable level, were answered by residents from four different specialties, including anaesthesiology, internal medicine, paediatrics, and surgery. The answers revealed that residents from the top two tertiles searched for more feedback, reread the literature more often, checked their actions more often on the internet, and did not try to hide the mistakes they made as much as residents from the bottom tertile.

Several studies have indicated that the characteristics of deliberate practice are difficult to utilize in the field of medicine, especially in work situations (Hoffman & Donaldson, 2004; van der Wiel, 2013, 2011). Here, task demands and goals vary in difficulty and are often beyond one's control (Charness, et al., 2005). Furthermore, ethical considerations limit the possibilities for practice and the work environment is designed to prevent mistakes being made to a maximum. In addition, while for example chess players can decide by themselves which chess problem they want to practice (Charness, et al., 2005), in the clinical field, the majority of learning activities are embedded in every day patient care (van der Wiel, 2011). The activities in the clinical field are mostly guided by practical experiences and not deliberately inducing the time needed for learning activities that aim at performance improvement (van der Wiel, 2013). Residency training, however, is a combination of work and training, and therefore the supervisor can adapt the degree of difficulty and demands to the level of the resident. Thus even though patient care is extremely important and time pressure is common, this study showed that those who are willing to improve their performance, are able to find possibilities and situations that are fruitful to their development.

One of the options residents used to improve their performance was searching for feedback. Without feedback, residents may not know how well they performed and what actions they need to undertake in order to improve their performance (Delva, et al., 2013; Ibrahim, MacPhail, Chadwick, & Jeffcott, 2014). Previous research in the clinical field has shown that feedback leads to more accurate diagnoses (Ericsson, 2004; Ericsson, 2009). Although feedback is seen as one of the most important requirements for learning in the workplace (Crommelinck & Anseel, 2013; Delva, et al., 2013; Ericsson, 2007), it is not always easily available since it takes time to confirm or reject a diagnosis, and during surgery or in other emergency situations feedback may be delayed (Ericsson, 2007). Residents, and other professionals, need to take time to receive and provide feedback (Ramani & Krackov, 2012) and they need to be able to establish a relationship in which they feel comfortable enough to ask for feedback (Delva, et al., 2013). In addition, since feedback serves the goal to improve performance (Delva, et al., 2013; Hattie & Timperley, 2007), residents must be willing to be open about the level of their skills to allow improvement. One of the other findings in this study was that high-performing residents were more willing to be honest about the mistakes they made. Thus, these residents create the opportunity to receive feedback that helps them to improve their performance. However, it should be noted that, to some extent, residents who ask for more feedback are observed better. Further research is required to gain more insight into the relation between residents' feedback-seeking behaviour and supervisor appraisal. This study also indicated that there are differences in repetition between residents. Residents who were assessed as high performers reread the literature more often. According to the theory of deliberate practice, it is also required to repeat practice (Ericsson, 2004). The latter, however, appears to be less common in the clinical field, probably because it is not possible to practice on patients just for practice purposes. Future research is needed to investigate this issue in more detail.

Although this study enabled us to identify some aspects of deliberate practice relating to performance differences in residents, other aspects were not identified. As stated above, repetition of knowledge is feasible, but repetition of practice appears to be more challenging. Apparently, it is difficult to fully employ deliberate practice in the clinical field as patient care is the most important task and proceedings are adjusted to the individual patient. Since every patient is different and needs personal care, many practices cannot be frequently repeated. It is therefore expected that in healthcare units where specialized care is provided and more similar proceedings are performed, more deliberate practice can be realized. Alternatively, an increased focus on practice in skills labs and on simulation trainings as well as portfolios in which tangible learning goals are described and that require longitudinal assessment, could help to improve the residents' performance (Donato & George, 2012).

In addition to the challenge in repeating proceedings, it is also difficult to create settings in which mistakes can be made and corrected, and to practice at a challenging but attainable level. Obviously, patient safety is the mainstay of patients' healthcare. However, an environment in which medical errors can be discussed is essential to learning. Regarding the level of practice, though supervisors can adapt practice partially to the residents' level, and portfolios can be used to structure the practice, the level still depends on the required care. Activities can only be labelled as deliberate practice when they are performed on a regular basis with the deliberate intention to improve performance (Sonnentag & Kleine, 2000), which often is not explicitly the case in a clinical setting. Although some deliberate practice can be performed during residency training, the setting does seem to limit the full amount of deliberate practice possible. Next to the differences in behavioural activities that were found between the different performance groups, age was negatively correlated with supervisors' assessment score, while the training year was not. It seems that those residents who were older when they started their residency training, performed less well. More research is required to gain insight into this negative relation.

Our findings also seem to indicate that the different activities aimed at improving performance are not autonomous, but depend on and influence each other. So far, deliberate practice has been described as a "set of highly structured activities", which indicates that several activities can be assigned to concept deliberate practice. Ericsson clearly distinguished "several activities" (Ericsson, et al., 1993) and Sonnentag (2000) even specifies that people may vary with respect to the extensiveness of different deliberate practice activities (Sonnentag & Kleine, 2000). These activities can nonetheless be very effective individually since repeating of basic principles is required to improve performance levels and when you increase your level, the possibility of making mistakes increases as well. When someone is open about the mistakes they make, feedback can help to overcome such weaknesses. These elements seem to reinforce each other; the most optimal outcome may be found when all elements are applied together: practice basics to attain higher levels, dare to make mistakes and be open about them, ask for feedback and practice again.

Although this study was able to identify relevant differences in behaviour between the best and least performing residents, several limitations are worth mentioning. Since in medicine little research has been performed to establish a measurement for excellent performance, and no valid and reliable methods exist (Ericsson, 2004), some kind of gold standard for performance had to be created. We therefore asked supervisors to assess the performance of their residents, which makes them set the standard for performance. Supervisor assessments are the most frequently used measures of overall job performance (Schmitt, 2012), but supervisors do naturally vary in opinion on the level of the residents, as people have the tendency to allow overall impressions of an individual to influence the judgment of a person's job performance (King, Hunter, & Schmidt, 1980). Our mediocre interrater reliability (ICC) confirmed the expected judgment variation, which strengthened our decision to include multiple supervisor opinions per resident in the analysis. However, as a consequence, 54 residents who had completed the questionnaire could not be included in this study as they were assessed by less than three supervisors. In addition, it should be noted that there is no valid questionnaire on deliberate practice in the clinical field available; such an instrument would have made our findings more solid. A strength of this study is that we have created a representative sample of specialties by including residents from four different residency training programmes. Several studies make the division technical-oriented versus personal-oriented specialties (Borges & Gibson, 2005; Lieu, Schroeder, & Altman, 1989; Zeldow, Devens, & Daugherty, 1990) or primary care versus specialty care versus supporting services (Babbott, Baldwin, Jolly, & Williams, 1988), and all these subgroups were covered by the four included specialties. A final strength is the collaborative research in this study (Gruppen, 2007), as residents and supervisors from different hospitals and even different training regions were included.

This study provides information that may be translated to practical implications. Since some of the principles of deliberate practice appear to be related to the performance levels of the residents in the various specialty programmes, inclusion of these elements in residency programmes will be beneficial. Thus, supervisors should be aware of the importance of practice, while the value of repetition must be emphasized by residents. A residency training that employs a competency-based approach could help to increase the focus on skills training. In addition, the use of portfolios including prescribed training could facilitate repetition, and feedback could be scheduled as part of the training (Holmboe, et. al., 2011). Nevertheless, both supervisors and residents should invest in a culture that facilitates openness, allowing errors, level dependent practice and constructive feedback.

CONCLUSION

This study enabled us to identify differences in behavioural characteristics relating to deliberate practice, between low, average and high-performing residents. High-performing residents were found to search for more feedback, to reread the literature more often and to be more open about mistakes they make. However, the clinical setting appears to limit the possibilities for deliberate practice, as it would be beneficial to residents to have more situations in which they are allowed to make and discuss mistakes and where they have time and feel comfortable to ask for feedback.

REFERENCES CHAPTER 5

Babbott, D., Baldwin, D. C., Jolly, P., & Williams, D. J. (1988). The stability of early specialty preferences among us medical school graduates in 1983. *JAMA, 259*(13), 1970-1975.

Bennett, N. L., Gordon, J. A., & Williamson Shaffer, D. (2004). Learning, testing, and the evaluation of learning environments in medicine: Global performance assessment in medical education. *Interactive Learning Environments, 12*(3), 167-178.

Borges, N. J., & Gibson, D. D. (2005). Personality patterns of physicians in person-oriented and technique-oriented specialties. *Journal of vocational behavior, 67*(1), 4-20.

Charness, N. (1981). Visual short-term memory and aging in chess players. *Journals of Gerontology, 36*(5), 615-619.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*(2), 151-165.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York: NY: Routledge Academic.

Cote, J., Ericsson, K. A., & Law, M. P. (2005). Tracing the development of athletes using retrospective interview methods: A proposed interview and validation procedure for reported information. *Journal of Applied Sport Psychology, 17*(1), 1-19.

Crommelinck, M., & Anseel, F. (2013). Understanding and encouraging feedback-seeking behaviour: a literature review. *Medical Education, 47*(3), 232-241.

Delva, D., Sargeant, J., Miller, S., Holland, J., Alexiadis Brown, P., Leblanc, C., Lightfoot, K. & Mann, K. (2013). Encouraging residents to seek feedback. *Medical Teacher, 35*(12), e1625-e1631.

Donato, A. A., & George, D. L. (2012). A blueprint for implementation of a structured portfolio in an internal medicine residency. *Academic Medicine, 87*(2), 185-191.

Ericsson, K. A. (1996). *The Road To Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games* (Vol. 1). Mahwah, USA: Lawrence Erlbaum Associates.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine Research in Medical Education Proceedings of the Forty-Third Annual Conference November 7-10, 2004, 79*(10), S70-S81.

Ericsson, K. A. (2007). An expert-performance perspective of research on medical expertise: the study of clinical performance. *Medical Education, 41(*12), 1124-1130.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: a general overview. *Academic Emergency Medicine, 15*(11), 988-994.

Ericsson, K. A. (2009). *The Cambridge handbook of expertise and expert performance:* edited by K. Anders Ericsson [et al.]. New York [etc.]:

Cambridge University Press.
Ericsson, K. A. (2011). *The surgeon's expertise* (Vol. 2, pp. 107-121). Dordrecht: Springer Netherlands.
Ericsson, K. A., Krampe, R. T., & Tësch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.
Ericsson, K. A., & Lehman, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47,* 273-305.
Gifford, K. A., & Fall, L. H. (2014). Doctor coach: a deliberate practice approach to teaching and learning clinical skills. *Academic medicine : journal of the Association of American Medical Colleges, 89*(2), 272-276.
Gruppen, L. D. (2007). Improving medical education research. *Teaching and learning in medicine, 19*(4), 331-335.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.
Hoffman, K. G., & Donaldson, J. F. (2004). Contextual tensions of the clinical environment and their influence on teaching and learning. *Medical Education, 38*(4), 448-454.
Holmboe, E. S., Ward, D. S., Reznick, R. K., Katsufrakis, P. J., Leslie, K. M., Patel, V. L., Ray, D. D. & Nelson, E. A. (2011). Faculty development in assessment: The missing link in competency-based medical education. *Academic Medicine, 86*(4), 460-467.
Hammond, K. L., Harmon, H. A. & Webster, R. L. (2007). University performance and strategic marketing: an extended study. *Marketing Intelligence & Planning, 25*(5), 436-459.
Hopmans, C. J., den Hoed, P. T., van der Laan, L., van der Harst, E., van der Elst, M., Mannaerts, G. H., Dawson, I., Timman, R., Wijnhoven, B. P. & IJzermans, J. N. M. (2014). Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery, 156*(6), 1078-1088.
Ibrahim, J., MacPhail, A., Chadwick, L., & Jeffcott, S. (2014). Interns' perceptions of performance feedback. *Medical Education, 48*(4), 417-429.
Jastrzembski, T. S., Charness, N., & Vasyukova, C. (2006). Expertise and age effects on knowledge activation in chess. *Psychology and Aging, 21*(2), 401-405.
Kastner, L., Gore, E., & Novack, A. H. (1984). Pediatric residents' attitudes and cognitive knowledge, and faculty ratings. *The Journal of pediatrics, 104*(6), 814-818.
Kerdijk, W., Tio, R. A., Mulder, B. F. & Cohen-Schotanus, J. (2013). Cumulative assessment: strategic choices to influence students' study effort. *BMC Medical Education, 13*(172), 1-7.
King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology, 65*(5), 507-516.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology,*

4(863), 1-12.

Lieu, T. A., Schroeder, S. A., & Altman, D. F. (1989). Specialty choices at one medical school: Recent trends and analysis of predictive factors. *Academic Medicine, 64*(10), 622-629.

Long, D. M. (2000). Competency-based residency training: The next advance in graduate medical education. *Academic Medicine, 75*(12), 1178-1183.

McGill, D. A., van der Vleuten, C. P. M., & Clarke, M. J. (2011). Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Advances in Health Sciences Education, 16*(3), 405-425.

McGill, D. A., van der Vleuten, C. P. M., & Clarke, M. J. (2013). A critical evaluation of the validity and the reliability of global competency constructs for supervisor assessment of junior medical trainees. *Advances in Health Sciences Education, 18*(4), 701-725.

Ramani, S., & Krackov, S. K. (2012). Twelve tips for giving feedback effectively in the clinical environment. *Medical Teacher, 34*(10), 787-791.

Schmitt, N. (2012). *The Oxford handbook of personnel assessment and selection* (1 ed.). New York: Oxford University Press.

Sonnentag, S., & Kleine, B. M. (2000). Deliberate practice at work: A study with insurance agents. *Journal of Occupational and Organizational Psychology* 73 (1), 87-102.

Teunissen, P. W., Scheele, F., Scherpbier, A. J. J. A., van der Vleuten, C. P. M., Boor, K., van Luijk, S. J. & van Diemen-Steenvoorde, J. A. A. M. (2007). How residents learn: qualitative evidence for the pivotal role of clinical activities. *Medical Education, 41*(8), 763-770.

van der Wiel, M. W. J., van den Bossche, P. (2013). Deliberate practice in medicine: The motivation to engage in work-related learning and its contribution to expertise. *Vocations and Learning, 6*(1), 135-158.

van der Wiel, M. W. J., van den Bossche, P., Janssen, S. & Jossberger, H. (2011). Exploring deliberate practice in medicine: how do physicianse learn in the workplace? *Advances in Health Sciences Education 16*(1), 81-95.

Zeldow, P. B., Devens, M., & Daugherty, S. R. (1990). Do person-oriented medical students choose person-oriented specialties? Do technology-oriented medical students avoid person-oriented specialties? *Academic Medicine, 65*(9), S45-S46.

Susanna M. Lucieer
Cornelis J. Hopmans
Pieter T. den Hoed
Reinier Timman
Rosemarijn T. Steensma
Remy M.J.P. Rikers
Axel P.N. Themmen
Jan N.M. IJzermans

Sumitted for publication

INDENTIFICATION OF FACTORS ASSOCIATED WITH PERFORMANCE DURING SURGICAL RESIDENCY TRAINING

Chapter 6

ABSTRACT

Background

Selection of surgical residents is a high-stake process entailing the great responsibility to recruit candidates who will be successful during training and in future practice. However, information on factors that may affect performance during surgical residency training is limited. This study aimed to determine the relation between pre-training variables and indicators for performance during surgical training.

Methods

In 2014, application files of a cohort of surgical residents enrolled in a regionally organized training programme between 2007 and 2013 in the Netherlands, were retrospectively analysed. Pre-training variables included demographic data, secondary and medical school grades, academic credentials, and clinical work experience. Using regression analyses, these variables were related to scores for in-training examinations completed between 2011 and 2013 and global performance scores obtained by telephone interviews in 2013.

Results

The mean ±SD age at time of application of the 43 surgical residents (24 men) was 28.4 ±1.4 years. Twenty-seven percent (adjusted R^2 = .266, F = 8.418, p = .001) of the variance for the in-training examinations scores was explained by gender (B = -8.130, p = .005), and mean pre-clinical grade in medical school (B = 10.475, p = .002). For the global performance scores, 39% (adjusted R^2 = .388, F = 7.661, p < .001) of the variance was assigned to: type of secondary school (B = -5.046, p = .032), graduation grade for mathematics (B = 3.186, p = .004), mean pre-clinical grade in medical school (B = 5.327, p = .044), and length of clinical work experience (B = -.292, p = .025).

Conclusion

Several pre-training variables were associated with indicators for performance during training. These variables may be helpful to optimize the recruitment of promising surgical residents.

INTRODUCTION

The aim of surgical training is to adequately prepare the next generation of surgeons to independently perform in professional practice, and to guarantee the continued delivery of high-quality patient care. In order to meet contemporary healthcare demands, such as the progressive sub specialization in surgery, an increased emphasis on lifestyle issues, and a rising awareness for accountability, transparency and quality in medicine by an increasingly well-informed public, it is critical to train surgeons that match today's practice (Frenk, et al., 2010). Therefore, the goal for selection committees is to ensure that those candidates, who are able to acquire skills quickly and will work safe and competent in their future careers, are accepted into surgical training programmes, which stresses the importance of valid selection methods.

The total costs of surgical training programmes are high, ranging from around \$650,000 in the United States, about £413,000 in the United Kingdom, to €840,000 in the Netherlands per full-time equivalent. These large investments are either financed with public money, funded by insurance companies, or paid personally, often with private loans (The Cost of Surgical Training. Position statement by the Association of Surgeons in Training, 2007 ; Kostprijsonderzoek tot Medisch Specialist door Samenwerkende Topklinische Ziekenhuizen, 2012; Wynn, 2013). Consequently, the recruitment process should be directed to select surgical residents with a low risk of failure to prevent financial loss. In addition, the use of public money stresses the importance of appropriately functioning training programmes in order to meet societal demands and bear social responsibility. Taken together, selection is a high-stake and challenging process entailing the great responsibility to correctly identify and recruit those candidates who will be successful during training, as well as in future practice (Kenny, McInnes, & Singh, 2013; Thordarson, Ebramzadeh, Sangiorgio, Schnall, & Patzakis, 2007). However, dropout rates in surgical training programmes are high, with approximately 20% of the surgical residents not completing their training and yearly attrition rates up to 7% have been reported (Dodson & Webb, 2005; Longo, Seashore, Duffy & Udelsman, 2009; Yeo, et al., 2009). To minimize the attrition, more evidence-based information about validity of selection methods is needed.

A typical selection procedure for surgical residency training starts with the screening of cover letters, personal statements, letters of recommendation, academic credentials and curriculum vitae provided by the candidates. Due to the overwhelming demand for training positions in most Western countries (Carr, Munsch, Buggle, & Hamilton, 2011; Margreiter, 2011; Veldkamp, 2007), it is not uncommon for aspiring surgeons to undertake various activities in a bid to improve their eligibility. These activities include participation in scientific research, writing a PhD thesis, and acquiring clinical experience and technical skills as a resident not in training (Veldkamp, 2007). It has been reported that in the US United States Medical Licensing Examination (USMLE) scores and honour society memberships play a pivotal role (Makdisi, et al., 2011; Melendez, Xu, Sexton, Shapiro, & Mohan, 2008; Sharp, et al., 2014; Thordarson, et al., 2007). However, the predictive validity of these selection criteria is not supported by scientific research (McGaghie, Cohen, & Wayne, 2011).

After screening of all documents, the following step in the selection procedure is to invite those who are considered promising for a personal interview (Makdisi, Takeuchi, Rodriquez, Rucinski, & Wise, 2011). However, the value of the personal interview, which is generally considered to be the decisive factor for selection committees in assigning a training post (Makdisi, et al., 2011; Melendez, et al., 2008; Swanson, et al., 2005) has also been called into question (Kenny, et al., 2013).

Summarized, there is a paucity of information about the factors known at the time of application that can be related to performance during surgical residency training or beyond (Harfmann & Zirwas, 2011; Kenny, et al., 2013). Therefore, more information about the predictive value of the factors known at the time of application is highly requested in order to ensure that the best candidates are attracted into surgical residency training programmes. This study aimed to determine the relation between pre-training variables that were derived from structured curriculum vitae, as well as secondary school and medical school diplomas with indicators for performance.

METHODS

Setting

In the Netherlands, surgical residency training lasts six years and is organized in eight training regions, each consisting of one university hospital and several affiliated district hospitals. After graduating from medical school, consisting of a pre-clinical stage followed by a clinical stage of rotational clerkships in various medical disciplines, those having the ambition to pursue a career in surgery are allowed to apply for a training post. Biannually, a nationwide selection round takes place that is overseen and coordinated by the Association of Surgeons of the Netherlands. Candidates have to specify two training regions of preference and are required to submit a motivation letter together with structured curriculum vitae, as well as supporting documents such as copies of secondary school and medical school diplomas. The selection committees of the preferred training regions, consisting of the local programme directors of the hospitals involved in surgical residency training complemented by one resident-representative, eventually decide which candidates are invited for a personal interview and finally, which are selected (Bonjer & Bruining, 1999; Borel-Rinkes, Gouma, & Hamming, 2008).

Study design

This study was conducted in a surgical training region in the Netherlands, consisting of one university hospital and six district hospitals. From April to May 2014, two of the authors (PdH & RS) independently retrospectively analysed the submitted application files of a cohort of surgical residents enrolled in the training programme between 2007 and 2013. The pre-training variables derived from these files were related to indicators for performance during training, including scores for in-training examinations completed between 2011 and 2013 and global performance scores obtained by telephone interviews by the programme director (JIJ) in August 2013. Ethical approval for this study was obtained from the local ethical research committee (MEC-2014-167).

Pre-training variables

The variables that were entered into an anonymized database consisted of demographic details, performance during secondary school and medical school, academic credentials and clinical work experience. The demographic details included gender and age. Performance during secondary school, which generally lasts six years, included the graduation grades obtained at the national exam for the courses obligatory for medical school: biology, chemistry, mathematics, physics and the languages Dutch and English. In the Netherlands, these courses are all prerequisites for entrance into medical school. In addition, since in the Netherlands two types of secondary school allow admission to university, the influence of both these different school types was measured. Briefly, type I secondary schools, also called Athenaeum, differ from the type II schools, which are called Gymnasium or Lycée (conditional Europe equivalent to North American university-preperatory schools and British grammar schools), in a way that in a type II school students are educated in Greek and/or Latin as an additional course. Medical school performance was assessed by calculating the mean grade obtained for written knowledge tests during the pre-clinical stage, the mean score for the clinical stage of rotational clerkships, the awarded score for the surgery clerkship in particular, and finally the grade obtained for graduation research. It should be noted that grades and test scores in the Netherlands range from 1-10, with 10 being the highest available score and a score of 5.5 being the pass/fail threshold. Academic credentials were reflected by the number of publications, both published and accepted, and the number of oral or poster presentations at national and international conferences. In addition, a completed PhD thesis at the time of application, or within one year after starting the surgical residency training, was considered an academic achievement. Finally, the number of months of clinical work experience as a resident not in training was evaluated.

Indicators for performance

The first indicator of performance during surgical residency training was the mean score for in-training examinations completed between 2011 and 2013. These computer-based examinations take place annually and consist of 90 multiple-choice questions that address medical knowledge and the management of surgical patients regarding diagnosis and treatment. Scores are reported as the percentage of the number of questions answered correctly. Participation in these exams is highly recommended but is not obligatory. The second indicator of performance was a score of global performance. Different faculty surgeons were asked to indicate their trust in the ability for each of the surgical residents in their hospital to treat/operate on their relatives. Global performance scores were obtained by telephone interview in August 2013 with scores ranging from 0, indicating zero confidence, to 100, indicating full confidence.

Statistics

The data were analysed with IBM SPSS Statistics version 20.0 (SPSS, Inc., Chicago, IL, USA). Descriptive statistics were calculated for all pre-training variables. Missing data (< 5%) were replaced with substituted values using the imputation method. Scores for the in-training examinations were corrected for level of difficulty, by calculating z-scores for each of the analysed years separately. Subsequently, these z-scores were further corrected by adjusting for postgraduate year of training, thereby taking into account that surgical residents' level of knowledge develops gradually during training.

For the global performance scores, the interrater agreement was computed using Intra Class Correlation (ICC). A One-way random effects model of the ICC was used as not all residents were evaluated by all supervisors, and average measures were used since this type of assessment is expected to be done by multiple supervisors (Weir, 2005; Dankbaar, 2014). To determine the relation between both the indicators of performance used in this study, Pearson correlation coefficient was calculated. A stepwise regression analysis with a backward elimination approach was carried out to identify pre-training variables that were related to scores for the in-training examinations and the faculty evaluations. The level for statistical significance was set at $p < .05$.

RESULTS

Demographics

In total, application files of 44 surgical residents were analysed. As a consequence of missing data, one resident was excluded from further analysis. Of the 43 remaining residents, 24 (56%) were men and 19 (44%) were women. The mean ±SD age at the time of application was 28.4 ±1.4 years. Twenty-nine of these residents (67%) had a type I secondary school degree. An overview of the other pre-training variables can be found in Table 1.

Correlation between indicators for performance

The scores for the in-training examinations correlated modestly to the global performance scores by the faculty surgeons ($r = .309$, $p = .047$), indicating that the global performance scores most probably refer to other qualities than to medical knowledge.

Variables related to in-training examination scores

The mean ±SD score for the annual in-training examinations was 48% ±7.9 (range 32-62) in 2011, 55% ±6.8 (range 44-67) in 2012, and 49% ±6.2 (range 37-62) in 2013. The median number of in-training examinations performed per individual resident was 2 (range 1-3). Regression analyses revealed that 27% (adjusted $R^2 = .266$, $F = 8.418$, $p = .001$) of the variance of the in-training examination scores was explained by gender ($B = -8.130$, $p = .005$), with men outperforming women, and by the mean pre-clinical grade in medical school ($B = 10.475$, $p = .002$).

Variables related to global performance scores

The mean ±SD assigned score for global performance by the faculty surgeons was 76 ±8.5 (range 57-93). Per individual resident, 5 evaluations were completed. The ICC of the global performance score was .738, indicating substantial agreement between the different faculty surgeons. Thirty-nine percent (adjusted $R^2 = .388$, $F = 7.661$, $p < .001$) of the variance for the global performance scores was assigned to type of secondary school ($B = -5.046$, $p = .032$), with type I outperforming type II. In addition, the graduation grade for mathematics in secondary school ($B = 3.186$, $p = .004$), the mean pre-clinical grade ($B = 5.327$, $p = .044$), and the length of clinical work experience ($B = -.292$, $p = .025$) were found to be related.

Table 1: Descriptive statistics of the pre-training variables derived from the files submitted at the time of application

		Mean	SD	Range
Secondary school	Biology	7.04	0.74	6.0 - 9.0
	Chemistry	6.98	1.05	5.0 - 9.0
	Mathematics	7.13	1.07	5.0 - 9.0
	Physics	6.96	1.07	5.0 - 9.0
	Dutch	7.12	0.73	6.0 - 9.0
	English	6.91	0.97	5.0 - 9.0
Medical school	Pre-clinical grade	6.92	0.44	6.2 - 8.5
	Clerkship grade	8.08	0.35	7.0 - 8.7
	Surgery clerkship	8.21	0.41	7.0 - 9.0
	Graduation research	8.40	0.78	6.9 – 10.0
Academic credentials	Published and accepted articles *(median)*	8		1 -27
	Oral and/or poster presentations *(median)*	8		0-20
	Completed PhD thesis* *(number and percentage)*	15 (35%)		
Clinical work experience in months		11.8	8.6	0 - 36

Data are presented as means and standard deviations (SD), unless otherwise indicated
**completed at the time of application or within one year after starting surgical residency training*

DISCUSSION

Admission into surgical residency training is not based on tangible indicators that are scientifically proven. Historically, selection is based on passing examinations, academic credentials, or letters from acknowledged academicians, followed by evaluation of personal, professional and communication skills involving a personal interview (Makdisi, et al., 2011; Melendez, et al., 2008). Findings of the studies previously examining the predictive value of selection criteria on surgical residency performance revealed conflicting results (Sutton, et al., 2014; Thordarson, et al., 2007) and were of limited quality (Fryer, Corconan, George, Wange, & Darosa, 2012; Grewal, Yeung, & Brandes, 2013; McGaghie, et al., 2012; Oldfield, Beasley, Smith, Anthony, & Watt, 2013). Using data from application files of a cohort of surgical residents, this study aimed to determine whether pre-training variables were associated with indicators for performance during surgical training. Indicators for performance were both objectively and subjectively measured. The objective indicator consisted of scores for in-training examinations in which surgical residents' medical knowledge and the management of surgical patients regarding diagnosis and treatment were assessed. The subjective measure existed of global performance scores provided by different faculty surgeons regarding their trust to let their relatives be taken care of by a surgical resident employed in their hospital.

Of all the pre-training variables, only the mean pre-clinical grade was found to be related to both the in-training examination scores and the global performance scores. During the pre-clinical years of medical school, basic science courses, studies on human organ systems and pathophysiological processes are integrated in thematic blocks, providing a fundamental basis for clinical medicine. Accordingly, it may be assumed that grades obtained for written knowledge tests during this stage relate to performance in later years. This finding corresponds to evidence that medical students' examination scores appear to be the best predictive markers of performance during the later years of training (Harfmann & Zirwas, 2011; Kenny, et al., 2013).

Interestingly, gender was observed as a pre-training variable related to the score for the in-training examinations. Although it has been shown that women perform better during medical school (Haist, Wilson, Elam, Blue, & Fosson, 2000; Haq, Higham, Morris, & Dacre, 2005), results of this study demonstrate that this finding cannot be translated to surgical residency training, with men outperforming their female colleagues. One explanation could be that women have a different view on their careers. Some studies have reported that women tend to be more focused on a good balance between work and private life, may prefer more flexible hours, and may be less committed or attracted to seek leadership or management roles, to do research or to teach (Bleakley, 2013; Sanfey, et al., 2014). In addition, it may be that female surgical residents differently focus their energy and are less competitive to score highly on the in-training examinations. It may be speculated that the overlap between the years in surgical training and the years of childbearing and parenting may play a role as well. However, it is of interest to note that that the global performance scores did not show a gender association. Nonetheless, although still underrepresented, women are increasingly entering the surgical workforce (Roupret, Maggiori, & Lefevre, 2011). Also in the present study this demographic change is observed, with 44% of the surgical residents being women.

A variable that was related to the mean score for global performance was the type of secondary school. It was found that surgical residents with a type I secondary school degree received higher global performance scores than those with a type II secondary school degree. This was contrary to our expectations since it seems conceivable that familiarity with Greek or Latin (= type II secondary school) is beneficial to medical performance. However, both this study and previous research show no support for this hypothesis (Pampush & Petto, 2011). One could argue that students graduating from a type II secondary school are more linguistically, culturally and philosophically oriented; skills that may be less relevant to technical-oriented specialties such as surgery. From this perspective, it would be interesting to investigate whether in more person-oriented specialties, such as internal medicine and neurology, residents with a type II secondary school degree perform better.

Furthermore, the graduation grade for mathematics in secondary school was also found to be associated with the clinical performance of surgical residents, as reflected by the global performance scores. Mathematics relates to the degree one is able to think across different levels of abstraction, while abstraction is fundamental to cognitive development (Mitchelmore & White, 2007). Many studies have found strong relations between secondary school grades and pre-clinical performance (Ferguson, James, & Madeley, 2002; Lambe & Bristow, 2011; McManus, et al., 2005), but these relations appear to decrease during the clerkships (McManus, Smithers, Partridge, Keeling, & Fleming, 2003; Mitchell, 1990; Salvatori, 2001). However, no relation could be found with the in-training examination scores, probably because these scores are more directed to assess declarative knowledge as opposed to procedural knowledge.

Surprisingly, the length of clinical work experience as a resident not in training was negatively related to the global performance. At first glance this is unexpected as more experience is expected to relate to better performance. However, it may be that candidates who are motivated to become a surgeon and considered "ready for practice" by his/her local programme director are encouraged to apply for a training position at an early stage. As a result, these better candidates are also the ones who are selected early compared to those who do not have everything planned and organized.

Some pre-training variables were unexpectedly not related to performance during surgical residency training. These variables included the mean score for the clinical stage of rotational clerkships and more specifically the surgery clerkship grade. Further analysis of the data revealed that these variables are probably less informative as a serious restriction of range was observed, with mean scores > 8.0 and small standard deviations. This finding indicates that clerkship scores, which are generally awarded after an oral examination, have a weak discriminative ability and are prone to subjectivity that might be tainted by the relationship between student and assessor. Indeed, a previous study showed that clerkship scores acquired at different US medical schools are difficult to compare and are not a reliable indicator of performance (Takayama, et al., 2006). In contrast to findings in a systematic review examining the predictive value of selection methods for surgical residency training (Maan, Maan, Darzi, & Aggarwal, 2012), also academic credentials, such as the number of both publications and presentations, and a completed PhD thesis were not associated with performance during surgical residency training in this study.

A limitation of this study is that the data of candidates that were selected to enter the surgical training program could not be compared to those that were not. Unfortunately, in the Netherlands, it is mandatory by law that application files of candidates who have not been selected are destructed following a selection round. Other limitations include the relatively small number of surgical residents that could be analysed and the fact that this study was limited to one regionally organised training programme. Therefore, a nationwide study with a larger sample size is needed to confirm the generalizability of the results found in this study. Moreover, a greater number of study objects would also allow to analyse the data of surgical residents that dropped out of training, potentially providing selection committees an even better insight in which pre-training variables should be taken into account when recruiting candidates to enter a surgical training programme.

In conclusion, this study showed that the pre-training variables gender, type of secondary school, mathematics grade, average pre-clinical grade and length of clinical work experience were associated with performance during surgical residency training. Using these variables may be helpful to optimize recruitment of promising surgical residents and to reduce dropout rates. However, in addition to the lack of widely used measures of success in surgical training and a limited body of literature reporting about validated long-term performance outcomes in professional practice, more work is needed to determine whether the wide range of selection criteria commonly used are truly predictive in an evidence-based manner for successful performance during surgical residency training.

REFERENCES CHAPTER 6

Bleakley, A. (2013). Gender matters in medical education. *Medical Education, 47(*1), 59-70.

Bonjer, H. J., & Bruining, H. A. (1999). Surgery in the Netherlands. *Archives of Surgery, 134*(1), 92-98.

Borel-Rinkes, I. H., Gouma, D. J., & Hamming, J. F. (2008). Surgical training in the Netherlands. *World Journal of Surgery, 32*(10), 2172-2177.

Carr, A. S., Munsch, C., Buggle, S., & Hamilton, P. (2011). Core surgical training and progression into specialty surgical training: how do we get the balance right? *Bulletin of The Royal College of Surgeons of England, 93*(7), 244-248.

Dankbaar, M. E. W., Stegers-Jager, K. M., Baarveld, F., van Merrienboer, J. J. G., Norman, G R., Rutten, F. L., van Saase, J. L. C. M., & Schuit, S. C. E. (2014). Assessing the assessment in emergency care training. *Plos One, 9*(12), 1-13.

Dodson, T. F., & Webb, A. L. (2005). Why do residents leave general surgery? The hidden problem in today's programs. *Current Surgery, 62*(1), 128-131.

Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *British Medical Journal, 324*(7343), 952-957.

Frenk, J., Chen, L., Bhutta, Z. A., Cohen, J., Crisp, N., Evans, T., Fineberg, H., Garcia, P.J., Ke, Y., Kelley, P., Kistnasamy, B., Meleis, A., Naylor, D., Pablos-Mendez, A., Reddy, S., Scrimshaw, A., Sepulveda, J., Serwadda, D. & Zurayk, H. (2010). Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet, 376*(9756), 1923-1958.

Freyer, J. P., Corcoran, N., George, B., Wang, E., & Darosa, D. (2012). Does resident ranking during recruitment accurately predict subsequent performance as a surgical resident? *Journal of Surgical Education, 70*(1), 138-143.

Grewal, S. G., Yeung, L. S., & Brandes, S. B. (2013). Predictors of success in a urology residency program. *Journal of Surgical Education, 70*(1), 138-143.

Haist, S. A., Wilson, J. F., Elam, C. L., Blue, A. V., & Fosson, S. E. (2000). The effect of gender and age on medical school performance: An important interaction. *Advances in Health Sciences Education, 5*(3), 197-205.

Haq, I., Higham, J., Morris, R., & Dacre, J. (2005). Effect of ethnicity and gender on performance in undergraduate medical examinations. *Medical Education, 39*(11), 1126-1128.

Harfmann, K. L., & Zirwas, M. J. (2011). Can performance in medical school predict performance in residency? A compilation and review of correlative studies. *Journal of the American Academy of Dermatology, 65*(5), 1010-1022.e1012.

Kenny, S., McInnes, M., & Singh, V. (2013). Associations between residency selection strategies and doctor performance: a meta-analysis. *Medical Education, 47*(8), 790-800.

Kostprijsonderzoek tot medisch specialist door samenwerkende topklinische ziekenhuizen. (2012). November 2012, from www.stz.nl/download/f61e145f90ee8e4b94c496d40445eb541471.pdf

Lambe, P., & Bristow, D. (2011). Predicting medical student performance from attributes at entry: a latent class analysis. *Medical Education, 45*(3), 308-316.

Longo, W. E., Seashore, J., Duffy, A., & Udelsman, R. (2009). Attrition of categoric surgery residents: results of a 20-year audit. *American Journal of Surgery, 197*(6), 774-780.

Maan, Z. N., Maan, I. N., Darzi, A. W., & Aggarwal, R. (2012). Systematic review of predictors of surgical performance. *Britisch Journal of Surgery, 99*(12), 1610-1621.

Makdisi, G., Takeuchi, T., Rodriguez, J., Rucinski, J., & Wise, L. (2011). How we select our residents - A survey of selection criteria in general surgery residents. *Journal of Surgical Education, 68*(1), 67-72.

Margreiter, R. (2011). To be or not to be a general surgeon! *Annals of Surgery, 254*(5), 679-683.

McGaghie, W. C., Cohen, E. R., & Wayne, D. B. (2011). Are United States Medical Licensing Exam step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Academic Medicine : Journal of the Association of American Medical Colleges, 86*(1), 48-52.

McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D., & Richards, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal, 331*(7516), 555-559.

McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. R. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *Brithish Medical Journal, 327*(7407), 139-142.

Melendez, M. M., Xu, X., Sexton, T. R., Shapiro, M. J., & Mohan, E. P. (2008). The Importance of basic science and clinical research as a selection criterion for general surgery residency programs. *Journal of Surgical Education, 65*(2), 151-154.

Mitchell, K. J. (1990). Traditional predictors of performance in medical school. *Academic Medicine, 65*(3), 149-158.

Mitchelmore, M., & White, P. (2007). Abstraction in mathematics learning. *Mathematics Education Research Journal, 19*(2), 1-9.

Oldfield, Z., Beasley, S. W., Smith, J., Anthony, A., & Watt, A. (2013). Correlation of selection scores with subsequent assessment scores during surgical training. *ANZ Journal of Surgery, 83*(6), 412-416.

Pampush, J. D., & Petto, A. J. (2011). Familiarity with Latin and Greek anatomical terms and course performance in undergraduates. *Anatomical Sciences Education, 4*(1), 9-15.

Roupret, M., Maggiori, L., & Lefevre, J. H. (2011). Upcoming female preponderance within surgery residents and the association of sex with the surgical career choice in the new millennium: results from a national survey in France. *American Journal of Surgery, 202*(2), 237-242.

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education : theory and practice, 6*(2), 159-175.

Sanfey, H. A., Fromson, J. A., Mellinger, J. D., Rakinic, J., Williams, M., & Williams, B.

(2014). Surgeons in difficulty: an exploration of behavior differences among male and female surgeons. *Journal of the American College of Surgeons, 219*(4), e141.

Sharp, C., Plank, A., Dove, J., Woll, N., Hunsinger, M., Morgan, A., Blansfield, J. & Shabahang, M. (2014). The predictive value of application variables on the global rating of applicants to a general surgery residency program. *Journal of Surgical Education, ,72(*1), 148-155.

Sutton, E., Richardson, J. D., Ziegler, C., Bond, J., Burke-Poole, M., & McMasters, K. M. (2014). Is USMLE Step 1 score a valid predictor of success in surgical residency? *American Journal of Surgery, 208*(6), 1029-1034.

Swanson, W. S., Harris, C. M., Master, C., Gallagher, P. R., Mauro, A. E., & Ludwig, S. (2005). The impact of the interview in pediatric residency selection. *Ambulatory Pediatrics, 5*(4), 216-220.

Takayama, H., Grinsell, R., Brock, D., Foy, H., Pellegrini, C., & Horvath, K. (2006). Is it appropriate to use core clerkship grades in the selection of residents? *Current Surgery, 63*(6), 391-396.

The Cost of Surgical Training. (2007). Position statement by the association of surgeons in training. May 2007, from http://www.asit.org/assets/documents/ASiT_Cost_of_Surgical_Training_final.pdf

Thordarson, D. B., Ebramzadeh, E., Sangiorgio, S. N., Schnall, S. B., & Patzakis, M. J. (2007). Resident selection: How we are doing and why? *Clinical orthopaedics and related research,* (459), 255-259.

Veldkamp, J. (2007). General surgical training in the Netherlands. *Bulletin of The Royal College of Surgeons of England, 89*(3), 92-93.

Weir, J. P. (2005). Quantifying test-retest relaibility using the intraclass correlation coefficient and the SEM. *Journal of Strenght and Conditioning Research, 19*(1), 231-240.

Wynn, B. O., Smalley, R. & Cordasco, K. M. (2013) *Does It cost more to train residents or to replace them? A look at the costs and benefits of operating graduate medical education programs*. Santa Monica, CA: RAND Corporation, 2013.

Yeo, H., Viola, K., Berg, D., Lin, Z., Nunez-Smith, M., Camman, C., Bell, R. H. Jr., Sosa, J. A., Krumholz, H. M., & Curry, L. A. (2009). Attitudes, training experiences, and professional expectations of US general surgery residents: a national survey. *JAMA, 302*(12), 3101-1308.

GENERAL DISCUSSION

Chapter 7

Chapter 7 General discussion

The aims of this thesis were to provide further insight into the predictive validity of non-cognitive selection methods and to investigate whether insights from domains outside medical education are of interest to medical school selection. In total, four different themes were addressed in this thesis: cognitive and non-cognitive selection methods, self-regulated learning, the expert performance approach, and the criterion-predictor relationship. In this general discussion, the main findings of the different studies and considerations for medical school selection are discussed, followed by suggestions for further research.

REVIEW OF THE MAIN FINDINGS AND CONSIDERATIONS FOR MEDICAL SCHOOL SELECTION

Part 1: Cognitive versus non-cognitive selection

The study described in *Chapter 2* contributed to the discussion on the predictive validity for academic performance of cognitive and non-cognitive selection methods. A unique experiment was set up that enabled the comparison of academic performance of students selected on cognitive capabilities to those selected on non-cognitive capabilities and to students admitted by lottery (i.e., a control group). The results revealed that selection on cognitive capabilities has the strongest predictive power for academic performance, and that selection on non-cognitive capabilities on its own is not sufficient to include students who stand out in their academic performance during the pre-clinical stage of medical school. However, this does not necessarily mean that non-cognitive selection methods are useless. It is important to note that we only investigated the effect of different selection methods on performance during the early, pre-clinical years of medical school. The curriculum of these years is mainly cognitively oriented and the students' contact with patients is limited. It could be that non-cognitive selection methods have a stronger predictive power when it comes to performance during the final years of medical school, i.e. during clerkships, and in future careers, when non-cognitive capabilities become increasingly important. Students and medical doctors have to be able to work in teams, communicate with patients and colleagues, and show reliability and empathy (Pulvirenti, McMillan, & Lawn, 2014; Stewart, 2001). In addition, previous research has shown that students who are selected on the combination of cognitive and non-cognitive capabilities, achieved higher grades during clerkships (Urlings-Strop, Themmen, Stijnen, & Splinter, 2011). Another reason not to abandon non-cognitive selection methods is the notion of widening access, i.e. the efforts that medical schools have to employ to make sure that candidates from non-traditional backgrounds, such as students from cultural-ethnic minorities or with a lower socio-economic status, have an equal opportunity to be admitted to medical school. However, such candidates in general perform less well on tests that measure cognitive abilities (Gottfredson, 1986; Hunter, 1986; Ployhart, McFarland, & Ziegert, 2003; Stegers-Jager, Steyerberg, Lucieer, & Themmen, 2015). Thus, cognitive measures have an adverse impact on non-traditional applicants, causing a higher probability of being excluded when such measures are solely used in a selection procedure. Therefore, although the findings of the study described in Chapter 2 imply that cognitive selection has the strongest predictive power for academic performance in the pre-clinical stage of medical school, the inclusion of non-cognitive selection methods may a) improve the predictive validity of the instruments for the clinical stage and b) make the procedure a fairer one, by having less adverse impact on the candidates of non-traditional backgrounds.

Part 2: Thinking outside the box

In the second part of this thesis approaches for selection taken from personnel assessment and organizational psychology were investigated. The first approach was the possibility to use one's level of self-regulated learning skills to determine whether an applicant should be admitted or not. A complication in this approach is that self-regulated learning skills are not intrinsically fixed in a person, but develop gradually (van der Stel & Veenman, 2014; Veenman & Spaans, 2005). Therefore, it was needed to examine whether these skills would still develop during medical school. The studies described in *Chapters 3 and 4*, showed that most self-regulated learning skills do not develop during the pre-clinical stage of medical school, except for the skill reflection in one of the three investigated medical schools. Surprisingly, we found that these skills levels can even decrease over the years in medical school. This means that the environment does seem to influence the development of self-regulated learning skills, both positively and negatively. In addition, the study described in Chapter 3 showed that the skills monitoring, reflection, and effort, were found to be, to some extent, related to academic performance during the first year of medical school, while only the skill effort stood out during the third year. The findings that self-regulated learning skills can decrease during medical school and that there only is a limited relation to academic performance, indicate that most self-regulated learning skills are not of essential importance to successfully conclude medical school. Thus, the medical curricula do not seem to encourage students to take control of their learning activities. One reason for this may be that the workload for students in medical education is high, and consequently many medical schools offer highly structured curricula to help students absorb the large amount of information. However, such structured curricula hinder the development of self-regulated learning skills (Premkumar, et al., 2013). When medical schools really wish their students to develop these skills, they should consider how they can change the curriculum in a way that it encourages students to actively take control, while retaining the content of the education. Nevertheless, based on these findings, medical school selection committees may still consider to include self-regulated learning in their selection procedure, given that these skills only minimally change during medical school. However, it is important to note that most applicants have already quite high levels of self-regulated learning skills at entrance, precluding any fair determination of the selection cut-off value based on a measurement of these self-regulated learning skills.

In *Chapter 5*, the expert performance approach was examined: what makes some people perform better than others? In several other complex domains of human endeavour, it has been shown that one is not able to reach the top without prolonged and highly structured training, also named *deliberate practice* (Ericsson, 2009; Ericsson, Krampe, & Tësch-Romer, 1993). The results of this study showed that those residents who were assessed as better performers by their supervisors, clearly showed some patterns of deliberate practice. In particular, these high-performing residents indicated that they more often reread the literature, asked for feedback and admitted that they made mistakes, which is one of the necessary conditions to receive appropriate feedback. Although the clinical setting is not the most optimal environment to fully engage in deliberate practice (van der Wiel, 2011), it appears that those who are willing to improve their performance through using behavioural aspects of deliberate practice, do find possibilities to do so. However, not all residents are expected to fully engage in deliberate practice. Deliberate practice is effortful, often not enjoyable, does not lead to immediate rewards but is solely done with the aim of reaching and maintaining expert performance (Ericsson, et al., 1993). Nonetheless, deliberate practice is an interesting approach since this study has shown that adapting a few principles will already help to increase the residents' performance level. Thus, encouraging residents to behave accordingly some principles of deliberate practice and creating a setting in which development is supported, will be beneficial to the residents' level of performance.

With respect to medical school selection, it is important to consider whether the potential to engage in deliberate practice is visible and measurable at the start of medical school, and would allow identifying applicants who have the ability to become excellent medical specialists. All participants in Chapter 5 were in residency training, which means that they have a specific goal in mind, i.e. becoming a certain medical specialist. The drive to engage in deliberate practice could be in one's nature, but could also be stimulated when a clear goal is apparent. Thus, the awareness of a clear goal would stimulate investing a considerable amount of time and effort in training. This may be less so in the case of medical school applicants, who, at the start of their medical training, albeit highly motivated, quite often do not know what to expect of the medical profession and may not have a clear idea of the directions in which their efforts should develop (Marley & Carman, 1999). In addition, medical school applicants have less opportunities to engage in deliberate practice activities related to the medical domain, since they have much less access to skills labs, patients and supervisors. These factors complicate the possibility of identifying applicants' potential to engage in deliberate practice. Nonetheless, the studies in Chapter 3 and 4 have shown that at the start of medical school, some of the basic requirements needed to engage in deliberate practice, are present. People who engage in deliberate practice have to take control over their learning behaviour to develop in the most optimal way, and most first-year students are able to do so. In addition, the first-year medical students showed high levels of effort and self-efficacy, which are also necessities to engage in deliberate practice, since deliberate practice requires a lot of time and perseverance. The presence of these basic requirement is auspicious for the students' future development. Though, there could be a possibility to gauge the potential of deliberate practice in medical school applicants, namely by evaluating whether these candidates have shown excellent results in other, non-clinical domains, such as in the arts, sports, literature and science. According to Ericsson and colleagues (1993), one is not able reach excellence in performance without engaging in deliberate practice. Thus, applicants who have shown outstanding performance *before* medical school would have had to engage in deliberate practice which would be promising for their future medical careers.

Finally, in *Chapter 6*, the criterion - predictor relationship played an important role. To measure whether a selection method is able to predict job performance, an effective measure of proficiency on the job is required. Unfortunately, performance as a medical doctor is very difficult to assess (Benbassat & Baumal, 2007; Mahon, Henderson, & Kirch, 2013; McGaghie, 1990; Powis, 1994), making it complex to determine whether selection methods do select applicants who become good medical doctors. However, performance of a resident, i.e., a medical doctor in training to become a medical specialist, is continuously monitored during their training. In many aspects residents function as fully trained medical doctors, albeit under supervision. Therefore, the performance of a resident can be used as a proxy criterion for performance as a medical doctor. In Chapter 6, the relation between pre-training variables, available at time of application for surgical residency training, and performance as a surgical resident, was investigated. This study revealed that several variables related to performance, which are gender, secondary school mathematics grade, type of secondary school, mean pre-clinical grade, and, though negatively, the amount of clinical work experience. Although this study is a first explorative study with a small sample size, some variables can be of interest to the selection of surgical residents. Especially the secondary school mathematics grades may be useful to include the most promising future surgeons. The relation between this grade and surgical residency performance is noticeable, as previous research has shown that pre-university grades are predictive for performance in the first years of university, but also showed that the predictive validity of pre-university grades decreases in later years (de Koning, Loyens, Rikers, Smeets, & van der Molen, 2012; Ferguson, James, & Madeley, 2002; Lambe & Bristow, 2011; McManus, et al., 2005). It is therefore enchanting that this secondary school grade is predictive for postgraduate performance.

Nevertheless, this study is not the first that showed that there is a relation between cognitive ability, in this case presented as mathematics grades, and medical performance. Previous research has shown that students selected on their cognitive capabilities perform better during the pre-clinical stage medical school, i.e. drop out less often and earn more ECTS in the first year than students admitted by lottery (Chapter 2; Urlings-Strop, Stijnen, Themmen, & Splinter, 2009), and that first-year performance the best predictor is for performance during the entire pre-clinical stage of medical school (Stegers-Jager, Themmen, Cohen-Schotanus, & Steyerberg, 2015; Stegers-Jager, Steyerberg, Cohen-Schotanus, & Themmen, 2012). In addition, it has also been shown that applicants who earned high grades during secondary school perform best during the pre-clinical stage of medical school (Baars, 2009; Schripsema, van Trigt, Borleffs, & Cohen-Schotanus, 2014), and performance during the pre-clinical stage relates to surgical residency performance (Chapter 6). Taken together, different studies have shown that cognitive ability remains an important predictor throughout medical education.

SUGGESTIONS FOR FUTURE RESEARCH

The aims of this thesis were to provide further insight in the predictive validity of non-cognitive selection methods and to explore insights from outside the domain that may be of interest for medical school selection. Especially the word “explore” hints that the first steps have been taken, but more work needs to be done before the findings of these studies can be applied to medical school selection. The results of the studies described in this thesis have led to suggestions for future research, and will be described in the following section.

In the first study, an experiment was set up to compare the academic performance of two groups of selected students and lottery-admitted students. However, only the academic performance during the pre-clinical stage of medical school was analysed. As the pre-clinical stage is mainly cognitive, it was not surprising that cognitive selection showed the strongest predictive power for performance in this stage. Regarding the non-cognitive selection, its predictive power was absent during the pre-clinical stage, but could be present during the clinical stage of medical school. The students included in this experiment started medical school in 2008 and 2009, which means that the clerkship grades of all students who have completed medical school within the regular six years, will be available at the end of the summer of 2015. It would be insightful to perform a follow-up study in which the performance during the clerkships of these three groups of admitted students will be compared. In addition, performance after medical school could also be monitored to investigate whether there is a relation between selection and scientific output and preferred specialty training.

With respect to the study described in Chapter 5, this was a first explorative study on the role of deliberate practice during residency training. The study did produce some promising results and a follow-up study in which the questionnaire can be validated, would be valuable to answer the question whether these behavioural characteristics fitting in the concept of deliberate practice are visible and measurable at the time of application for medical school. In a future study, the survey items could be re-examined to investigate whether respondents interpreted the items as we intended. A pilot study in which respondents think aloud while completing the questionnaire may be helpful for this. In addition, if necessary, focus groups can be very useful to create new items. Finally, when the questionnaire is adapted, its reliability and content, criterion and construct validity should be examined before the questionnaire can be used as a valid measure of deliberate practice. Once a valid test for deliberate practice is available, the next step will be to investigate how the expert performance approach can be applied to medical school selection. A valid questionnaire could help identify those applicants who are able to show principles of deliberate practice that enables them to become excellent medical doctors.

The final study in this thesis, Chapter 6, showed there is a relation between secondary school mathematics grade and performance during surgical residency training. Mathematics grades are available at time of application for medical school and could also be of use to medical school selection procedures. However, as we do not want to only include future surgeons, future research should examine the relation between secondary school grades and performance in other specialty trainings. These studies can have the same nature as the study described in Chapter 6. Variables known at application for specialty training can be extracted from the residents' application files and could be related to global performance scores and, when available, scores on in-training knowledge tests. If mathematics or other secondary school grades relate to performance in other specialty programmes, these variables could receive another weighting in medical school selection.

Furthermore, there are also other issues regarding medical school selection, which are not directly related to the studies described in this thesis, that remain a challenge and require future research. The first challenge is the homogeneous character of the applicant pool. In the Netherlands, only those who have completed the highest level of secondary school with subjects adequate for medical school, are allowed to participate in selection procedures (Cohen-Schotanus, 1999; ten Cate, 2007), and overall, medical school applicants are more likely to be female, of high socio-economic status and have higher grades than other secondary school graduates who do not apply for medical school (Ferguson, James, Yates, & Lawrence, 2012). The majority of this applicant pool seems suitable to become a medical doctor (Benbassat & Baumal, 2007). Indeed, in the Netherlands, the success rate of candidates who are admitted into medical school by lottery is about 90% (Cohen-Schotanus, 1999). Nowadays, medical school selection focuses on selecting the best and brightest students, while all applicants are very intelligent. However, not all cognitively strong students have the non-cognitive qualities to perform well during their professional career. Thus, instead of only differentiating among the top ranks of outstandingly qualified applicants, selection committees should also focus on identifying applicants with unsuitable non-cognitive qualities, to decrease the chance that these applicants are selected (Powis, 2015). In Flanders, Situational Judgments Tests (SJT) are used to select out these unfavourable applicants (Lievens, Peeters, & Schollaert, 2008). There may not be many unfavourable applicants, but identifying and selecting out the few who are, will be valuable. On an SJT, applicants are presented a situation they may encounter and should judge to what extent responses that are provided are appropriate (Lievens, 2013; St-Sauveur, Girouard, & Goyette, 2014). It is expected that applicants who do not know how to behave appropriately in an interpersonal situation, can be identified by their low scores on an SJT. This hypothesis can be investigated in future research by asking all applicants to complete an SJT during the selection process, without judging them upon their performance in the SJT. Once these applicants have started medical school, their SJT scores can be related to notes on unprofessional behaviour during medical education. When there is a link between low SJT scores and unprofessional behaviour, this would indicate that it is possible to exclude unfavourable applicants during the admission procedure.

In addition, the risk of selecting applicants is that the homogeneity of the group of admitted students becomes even larger than it already was in the applicant pool. It is known that physicians have similar, high levels of cognitive abilities, and even share common personality traits (Borges & Savickas, 2002). However, at the end of medical education, a wide range of medical doctors is required to serve society, i.e. medical doctors that reflect society in terms of ethnicity, race and gender (Andriole, et al., 2010; Cohen

& Steinecke, 2006; Norcini, Zanten, & Boulet, 2008), and medical doctors ranging from primary care to non-primary care providers (Salsberg & Forte, 2002). Medical schools should thus ideally yield entrance to a mix of students with different competences and ambitions (Bandiera, Maniate, Hanson, Woods, & Hodges, 2015). Previous research has shown that cognitive selection is unfavourable for minorities, supporting the addition of non-cognitive methods in the selection procedure (Ployhart, et al., 2003; Stegers-Jager, et al., 2015), but further research is required to examine the effect of different selection methods and different weightings of these methods on student diversity (Cleland, Dowell, McLachlan, Nicholson, & Patterson, 2012; Stegers-Jager, et al., 2015). The challenge is how to deal with the diversity-validity dilemma, i.e. that some of the most valid predictors for job performance show large ethnic and sex subgroup score differences (Plyburn, Ployhart, & Kravitz, 2008), as the main goal remains to include the most promising students into medical school. In addition, retrospective research can be performed to relate results obtained in medical school selection with specialty choice, to examine whether it is possible to predict specialty choice at time of application, and whether applicants with certain specialty preferences are disadvantaged by selection. Of course, many factors, as for example clerkship experience, will probably influence the eventual career choice as well (Wong, Greenhalgh, Westhorp, & Pawson, 2012).

A final challenge is not specific to medical school selection, but to selection in general. The moment applicants are aware of the consequences of their answers, many complete questionnaires in a way they believe provides them the highest chance of being accepted, and not in a way that reflects themselves (Arthur, 2001). Selection committees are afraid that questionnaires, which are used to measure non-cognitive capabilities, do not represent the actual character of the applicant (Ziegler & Buehner, 2009). It is possible, at least to some extent, to control for socially desirability in integrity and personality tests (Ones & Viswesvaran, 1998; Ones, Viswesvaran, & Reiss, 1996). Options used to prevent social desirability include explaining to applicants that it is not in their best interest to distort questions and using forced-choice formats, empirically keyed subtle questions and scales designed to detect social desirability (Ones, et al., 1996). Nevertheless, other selection methods of which the outcome cannot be consciously influenced by the applicant may be of interest to medical school selection. An example of such a selection methods is the use of the format of a computer role-playing game. Previous research has shown that such games can function as a covert test for measuring player characteristics, as for example for personality. There was found to be a link between personality and player behaviour: neurotic players take longer to complete the game, people who score high on openness more frequently start conversations with other characters in the game, extraverted players move around more and agreeable players avoid unfriendly remarks (van Lankveld, 2013). For the purposes of medical school selection, it would be interesting, though no sinecure, to investigate whether such a computer game could be designed that measures intrapersonal skills such as integrity and professionalism. Such a game could be created, based on a strong theoretical framework, around scenarios developed with the help of focus groups. Preferably, these scenarios take place in the medical setting to increase the applicants' feeling that the selection method is relevant and fair (Patterson, Zibarras, Carr, Irish, & Gregory, 2011). Based on these scenarios, a game can be created in which these scenarios occur and require the applicant to respond upon. Finally, a valid questionnaire is needed to compare the applicants' responses to the outcome of the game in order to determine whether and how the correct attributes are displayed in the game. In addition to preventing social desirability, the advantage of such a selection method is that it is possible to take the tests in an online and unsupervised setting. In addition, it would be difficult to coach someone how to play the game in the most favourable way. However, there are also certain drawbacks to this method, including the time and money needed to develop and maintain the game.

CONCLUDING REMARKS

Selecting students for medical education is a high-stake and challenging process, placing an enormous pressure and responsibility on selection committees to admit the most suitable applicants, and to reject the least suitable ones. However, as there is no suitable criterion for a good medical doctor available, an applicant is selected whom is expected that he or she will perform the best during medical school. Nevertheless, it is possible to identify factors that help a medical doctor to provide good healthcare and perform at a higher level, as for example self-regulated learning skills and using deliberate practice. Creating selection methods that are able to measure or predict these factors will help to improve selecting students for medical school.

REFERENCES GENERAL DISCUSSION

Andriole, D. A., Jeffe, D. B., Hageman, H. L., Ephgrave, K., Lypson, M. L., Mavis,B., McDougle, L. & Roberts, N. K. (2010). Variables associated with full-time faculty appointment among contemporary US medical school graduates: Implications for academic medicine workforce diversity. *Academic Medicine, 85*(7), 1250-1257.

Arthur, W., Woehr, D. J., Graziano, W. G. (2001). Personality testing in employment settings - Problems and issues in the application of typical selection practices. *Personnel Review, 30*(5-6), 657-676.

Baars, G. J. A. (2009). *Factors related to student achievement in medical school.* Thesis: Erasmus University Rotterdam, the Netherlands.

Bandiera, G., Maniate, J., Hanson, M. D., Woods, N., & Hodges, B. (2015). Access and selection: Canadian perspectives on who will be good doctors and how to identify them. *Academic Medicine, 90*(7), 946-952.

Benbassat, J., & Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. *Advances in Health Sciences Education, 12*(4), 509-521.

Borges, N. J., & Savickas, M. L. (2002). Personality and medical specialty choice: A literature review and integration. *Journal of Career Assessment, 10*(3), 362-380.

Cleland, J. A., Dowell, J., McLachlan, J., Nicholson, S., & Patterson, F. (2012). *Report - identifying best practice in the selection of medical students.*

Cohen-Schotanus, J. (1999). Student assessment and examination rules. *Medical Teacher, 21*(3), 318-321.

Cohen, J. J., & Steinecke, A. (2006). Building a diverse physician workforce. *JAMA, 296*(9), 1135.

de Koning, B. B., Loyens, S. M. M., Rikers, R. M. J. P., Smeets, G., & van der Molen, H. T. (2012). Generation Psy: Student characteristics and academic achievement in a three-year problem-based learning bachelor program. *Learning and Individual Differences, 22*(3), 313-323.

Ericsson, K. A. (2009). *The Cambridge handbook of expertise and expert performance:* edited by K. Anders Ericsson. New York [etc.]: Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tësch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Ferguson, E., James, D., & Madeley, L. (2002). Factors associated with success in medical school: systematic review of the literature. *British Medical Journal, 324*(7343), 952-957.

Ferguson, E., James, D., Yates, J., & Lawrence, C. (2012). Predicting who applies to study medicine: Implication for diversity in UK medical schools. *Medical Teacher, 34*(5), 382-391.

Gottfredson, L. S. (1986). Societal consequences of the g factor in employment. *Journal of vocational behavior, 29*(3), 379-410.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of vocational behavior, 29*(3), 340-362.

Lambe, P., & Bristow, D. (2011). Predicting medical student performance from attributes at entry: a latent class analysis. *Medical Education, 45*(3), 308-316.

Lievens, F. (2013). Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Medical Education, 47*(2), 182-189.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review, 37*(4), 426-441.

Mahon, K. E., Henderson, M. K., & Kirch, D. G. (2013). Selecting tomorrow's physicians: The key to the future health care workforce. *Academic Medicine, 88*(12), 1806-1811.

Marley, J. & Carman, I. (1999). Selecting medical students: a case report of the need for change. *Medical Education, 33*(6), 455-459.

McGaghie, W. C. (1990). Qualitative variables in medical school admission. *Academic Medicine, 65*(3), 145-149.

McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D., & Richards, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal, 331*(7516), 555-559.

Norcini, J. J., van Zanten, M., & Boulet, J. R. (2008). The contribution of international medical graduates to diversity in the U.S. physician workforce: graduate medical education. *Journal of health care for the poor and underserved, 19*(2), 493-499.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*(2-3), 245-269.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660-679.

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education, 45*(3), 289-297.

Ployhart, R. E., McFarland, L. A., & Ziegert, J. C. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactionsresearch. *Human Performance, 16*(3), 231-259.

Plyburn, K.M., Ployhart, R.E., & Kravitz, D.A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61*(1), 143-151

Powis, D. (2015). Selecting medical students: An unresolved challenge. *Medical Teacher, 37*(3), 252-260.

Powis, D. A. (1994). Selecting medical students. *Medical Education, 28*(5), 443-469.

Premkumar, K., Pahwa, P., Banerjee, A., Baptiste, K., Bhatt, H., & Lim, H. J. (2013). Does medical training promote or deter self-directed learning? A longitudinal mixed-methods study. *Academic Medicine, 88*(11), 1754-1764.

Pulvirenti, M., McMillan, J., & Lawn, S. (2014). Empowerment, patient centred care and self-management. *Health Expectations, 17*(3), 303-310.

Salsberg, E. S., & Forte, G. J. (2002). Trends In the physician workforce, 1980-2000. *Health affairs, 21*(5), 165-173.

Schripsema, N. R., van Trigt, A. M., Borleffs, J. C. C., & Cohen-Schotanus, J. (2014).

Selection and study performance: comparing three admission processes within one medical school. *Medical Education, 48*(12), 1201-1210.

St-Sauveur, C., Girouard, S., & Goyette, V. (2014). Use of situational judgment tests in personnel selection: Are the different methods for scoring the response options equivalent? *International Journal of Selection and Assessment, 22*(3), 225-239.

Stegers-Jager, K. M., Steyerberg, E. W., Lucieer, S. M., & Themmen, A. P. N. (2015). Ethnic and social disparities in performance on medical school selection criteria. *Medical Education, 49*(1), 124-133.

Stegers-Jager, K. M., Themmen, A. P. N., Cohen-Schotanus, J., & Steyerberg, E. W. (2015). Prediciting performance: relative importance of students' backgournd and past performance. *Medical Education, 49*(9), 933-945.

Stegers-Jager, K. M., Steyerberg, E. W., Cohen-Schotanus, J., & Themmen, A. P. N. (2012). Ethnic disparities in undergraduate pre-clinical and clinical performance. *Medical Education, 46*(6), 575-585.

Stewart, M. (2001). Towards a global definition of patient centred care: The patient should be the judge of patient centred care. *British Medical Journal, 322*(7284), 444-445.

ten Cate, O. (2007). Medical education in the Netherlands. *Medical Teacher, 29*(8), 752-757.

Urlings-Strop, L. C., Stijnen, T., Themmen, A. P. N., & Splinter, T. A. (2009). Selection of medical students: a controlled experiment. *Medical Education, 43*(2), 175-183.

Urlings-Strop, L. C., Themmen, A. P. N., Stijnen, T., & Splinter, T. A. (2011). Selected medical students achieve better than lottery-admitted students during clerkships. *Medical Education, 45*(10), 1032-1040.

van der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: a longitudinal study from a developmental perspective. *European journal of Psychology of Education, 29*(1), 117-137.

van der Wiel, M. W. J., van den Bossche, P., Janssen, S. & Jossberger, H. (2011). Exploring deliberate practice in medicine: how do physicianse learn in the workplace? *Advances in Health Sciences Education 16*(1), 81-95.

van Lankveld, G. (2013). *Quantifying individual player differences*. Thesis, Tilburg University, Tilburg.

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*(2), 159-176.

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

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement, 69*(4), 548-565.

SUMMARY

Chapter 8

Worldwide, medical schools have to select their students as the number of applicants highly exceeds the number of places available to educate future medical doctors. A wide range of cognitive and non-cognitive criteria is used to select the most promising applicants, and to reject the least promising applicants. Medical school selection is a high-stake, but also a challenging process, as the applicant pool is very homogeneous and predictions for performance years from time of application need to be made. The purposes of this thesis were to contribute to the existing knowledge on cognitive and non-cognitive selection methods, and to examine whether approaches used in other domains can be applied to medical school selection.

COGNITIVE VERSUS NON-COGNITIVE SELECTION

Following the first objective of this thesis, the study described in *Chapter 2* focused on the effect of cognitive and non-cognitive selection on academic performance. Previous research investigating the predictive validity of these different measures concerned only the relation between a single or a combination of multiple selection methods and academic performance, but has not compared the two methods. In addition, most studies did not include a control group, and were thus not able to investigate whether the selected students perform academically better than students who are randomly assigned. In the study described Chapter 2, we set up an experiment that enabled the comparison of the academic performance of three different admission groups: students admitted based on their cognitive abilities, students admitted on their non-cognitive abilities, and finally, students who were admitted by lottery. As these students were all enrolled in the same curriculum, it was possible to compare the effect of different admission procedures on academic performance. Interestingly, in contrast to our expectations, non-cognitively admitted students did not outperform lottery-admitted students during the pre-clinical stage of medical school. Although we not necessarily expected them to earn more ECTS or higher grades, it was expected that they would drop out less often and passed the OSCE more often than lottery-admitted students. More in line with the expectations was the academic performance of cognitively elected students. These students dropped out less often and earned more ECTS in the first 52 weeks of medical school than the lottery-admitted students, indicating that cognitive selection is the strongest predictor for academic performance.

THINKING OUTSIDE THE BOX

The second objective of this thesis was to explore whether approaches from other domains can be applied to medical student selection. The investigated approached included: the use of self-regulated learning skills of students in selection, the expert performance approach, and the relationship between a predictor (i.e. a selection method) and a criterion (i.e. a measure of performance).

According to the field of personnel selection, self-regulated learning skills, and especially self-efficacy and self-monitoring, positively relate to job performance. These skills help people to learn more efficiently, and to stay more easily up to date on all the developments in their fields. Only, these skills are not fixed but develop gradually. Hence, before applying these measures to medical school selection procedures, the potential development of these skills during medical school should be investigated. The cross-sectional study described in *Chapter 3* examines the development of self-regulated learning in the pre-clinical stages of medical school, i.e. between year one and year three. In addition, since medical schools also want to include students who perform academically well, the relation between self-regulated learning skills and academic performance is investigated. The main finding here was that the self-regulated learning skills, except for reflection, did not develop between the first and third year of medical school. Apart from this, effort, monitoring, and reflection were slightly related to academic performance in the first year, and only one skill, effort, was related to performance in the third year.

Some researchers argued that self-regulated learning skills develop differently depending on the learning environment. The study reported in *Chapter 4* hence focused on the development of self-regulated learning skills in two distinct medical curricula in Brazil: a problem-based and a lecture-based curriculum. Surprisingly, the main finding here was that in the lecture-based curriculum, some self-regulated learning skills, i.e. planning, monitoring, and evaluation, decreased during medical school. In the problem-based curriculum, sixth-semester students were slightly better than second-semester students in reflection, but showed less effort than second-semester students. Though these findings may be influenced by culture, they do indicate that the environment plays a major role in the development of self-regulated learning skills.

Another approach explored in *Chapter 5* is the expert performance perspective. According to this perspective, one will not reach top levels without prolonged, structured training, also labelled as *deliberate practice*. The expert performance approach usually attempts to describe performance under standardized conditions and representative tasks wherein superior performance can be investigated. The complex character of the medical field makes it impossible to compare performance of medical doctors, and consequently little research has been conducted on the role of deliberate practice in the medical field. Residents, however, are assessed during their training, which enabled us to investigate whether behavioural characteristics fit in the concept of deliberate practice related to residency performance. Using a questionnaire to measure their behaviour and supervisor judgement as indicator of performance, this study revealed that those residents who performed in the top two tertiles, more frequently restudied the literature, asked for feedback, and acknowledged that they made mistakes. The latter is necessary to receive appropriate feedback. These findings indicate that deliberate practice is, to some extent, able to explain differences in performance of residents. However, further research is needed to investigate whether these behavioural differences are already present at time of application for medical school, and could be used in medical school selection to identify promising future medical doctors.

Subsequently, in *Chapter 6*, the criterion-predictor relation is highlighted. In this study, the relation between pre-training variables and performance during surgical residency training was examined. In order to investigate whether selection methods are able to predict how well someone will perform as a medical doctor, a criterion of good performance as a medical doctor is required. As residents, contrary to medical doctors, are assessed, their medical performance is, for now, the most optimal criterion available that can be used to measure the effectiveness of a selection method. In this study, biographical data, secondary and medical school performance, academic credentials and clinical work experience, were related to scores on in-training knowledge tests and to global performance scores. While mean pre-clinical grade related to both indictors of success, males only outperformed females on the knowledge test. As regards the global performance scores, aside from the pre-clinical grade, residents with an Atheneum secondary-school diploma, higher secondary-school mathematics grades and less work experience received higher scores. Selection on some of these variables is not allowed and a few other measures are not available at time of application for medical school, but secondary-school mathematics grades are available and its use is permitted. It would therefore be interesting to measure the relation between this variable and performance in other training programmes, to determine whether these mathematics grades should be valued more in medical school selection.

Finally, in *Chapter 7*, a general discussion on the studies described in this thesis is provided. The discussion starts with a short summary of the main findings and considerations for medical schools and medical school selection committees. These considerations include 1) the advice to use of cognitive selection methods, 2) that self-regulated learning skills can be used in medical school selection but it is perhaps more important to help students develop their self-regulated learning skills during medical school, or at least, make sure they do not decrease, 3) to apply the elements of deliberate practice in residency training as this will be beneficial to the residents' performance, and to investigate the presence of these behavioural characteristics at time of application to medical school and 4) to investigate the relation between pre-training variables and residency performance in other training programmes besides surgical residency training, in order to determine whether secondary school mathematics grades should be valued more in medical school selection. Apart from this, the remaining challenges in medical school selection as well as considerations for future research are provided.

SAMENVATTING

Chapter 9

Overal ter wereld moeten geneeskundeopleidingen hun studenten selecteren omdat meer mensen zich aanmelden voor de opleiding dan dat er opleidingsplaatsen beschikbaar zijn. Een heel scala aan methoden wordt gebruikt om de meest veelbelovende kandidaten te selecteren, en de minst veelbelovende kandidaten af te wijzen. Er staat veel op het spel bij de selectie van geneeskundestudenten, maar het is een zeer ingewikkeld proces. De mensen die zich aanmelden voor de geneeskunde opleiding vormen een heel homogene groep en er moet een voorspelling gedaan worden hoe deze mensen over een x aantal jaar zullen functioneren. Het doel van dit proefschrift is om bij te dragen aan de bestaande kennis over het selecteren met cognitieve en non-cognitieve methoden en om te kijken of benaderingen uit andere domeinen toegepast kunnen worden op de selectie van geneeskundestudenten.

COGNITIEF VERSUS NON-COGNITIEF SELECTEREN

In navolging op het eerste doel van dit proefschrift richt de studie die beschreven staat in *hoofdstuk 2* zich op het effect van verschillende selectiemethoden, één die cognitieve vaardigheden meet en één die non-cognitieve vaardigheden meet, op de academische prestaties van geneeskundestudenten. Eerdere studies die de voorspellende waarde van deze verschillende selectiemethoden hebben onderzocht, hebben zich enkel gericht op de relatie tussen een enkel, of een combinatie van meerdere methoden en academische prestaties, maar hebben de voorspellende waarde van cognitieve en non-cognitieve selectiemethoden niet met elkaar vergeleken. Daarnaast hadden de meeste studies ook geen controlegroep geïncludeerd. Zij waren daarom niet in staat om te onderzoeken of de geselecteerde studenten academisch beter presenteerden dan een groep studenten die willekeurig was toegelaten tot de opleiding. In de studie in hoofdstuk 2 hebben we een uniek experiment opgezet dat het mogelijk maakt om de academische prestaties van drie verschillende groepen studenten te vergelijken: studenten die geselecteerd zijn op basis van hun cognitieve vaardigheden, studenten die geselecteerd zijn op basis van hun non-cognitieve vaardigheden en studenten die via loting zijn toegelaten tot de geneeskundeopleiding. Aangezien al deze studenten deelgenomen hebben aan hetzelfde curriculum, was het mogelijk om het effect van deze verschillende toelatingsmethoden op academische prestaties te onderzoeken. Opvallend was dat, in tegenstelling tot onze verwachting, de studenten die geselecteerd waren op basis van hun non-cognitieve vaardigheden, niet beter presteerden tijdens de preklinische fase van de geneeskundeopleiding dan de ingelote studenten. Hoewel we niet hadden gedacht dat ze meer studiepunten zouden verdienen of hogere cijfers zouden halen, hadden we wel verwacht dat ze minder vaak uit zouden vallen en vaker de klinische-vaardighedentoets zouden halen dan de ingelote studenten. De academische prestaties van de studenten die geselecteerd waren op basis van hun cognitieve vaardigheden lagen meer in lijn met de verwachtingen. Deze studenten vielen minder vaak uit en behaalden meer studiepunten in de eerste 52 weken van hun opleiding dan de ingelote studenten. De bevindingen van deze studie laten dus zien dat selectie op basis van cognitieve vaardigheden de sterkste voorspeller is voor academische prestaties, en dat selectie op basis van non-cognitieve vaardigheden op zichzelf niet voldoende is om de best presterende studenten binnen te halen.

DENKEN BUITEN DE GEIJKTE KADERS

Het tweede doel van dit proefschrift was om te onderzoeken of ideeën van andere domeinen toegepast konden worden op de selectie van geneeskundestudenten. Deze ideeën waren: het gebruik van het zelfregulerend leervermogen van studenten, de aanpak van experts om tot hun niveau van prestatie te komen, en de relatie tussen een voorspeller, dat wil zeggen een selectiemethode, en een criterium, een maat voor prestatie.

Onderzoek op het gebied van personeelsselectie laat zien dat er een positieve relatie bestaat tussen het zelfregulerend leervermogen van mensen en hun prestatie op de werkvloer. Voornamelijk het vertrouwen van mensen in hun eigen kunnen en het monitoren van hun leergedrag spelen hierin een grote rol. Deze vaardigheden helpen mensen om efficiënter te leren waardoor ze makkelijker op de hoogte blijven van alle ontwikkelingen in hun vakgebied. Deze vaardigheden staan niet vast bij iemands geboorte, maar ontwikkelen zich tijdens het leven. Voordat er dus besloten kan worden om deze vaardigheden te meten en kandidaten daarop te beoordelen tijdens een selectieprocedure, is het belangrijk om te kijken of deze vaardigheden zich nog ontwikkelen tijdens de geneeskundeopleiding. Met behulp van het cross-sectionele onderzoek dat beschreven staat in *hoofdstuk 3* is daarom de ontwikkeling van het zelfregulerend leervermogen tijdens de preklinische fase van de geneeskundeopleiding onderzocht, dat wil zeggen tussen het eerste en derde jaar. Omdat geneeskunde-opleidingen het ook belangrijk vinden om kandidaten te selecteren die academisch goed presteren, is daarnaast de relatie tussen het zelfregulerend leervermogen van studenten en hun academische presentaties onderzocht. De voornaamste bevindingen waren dat, over het algemeen, het zelfregulerend leervermogen van de studenten niet beter werd tijdens de preklinische fase van de geneeskundeopleiding, alleen het reflectievermogen nam iets toe. Daarnaast was er een kleine samenhang tussen inzet, het monitoren van het leergedrag en het reflectievermogen met de prestaties van eerstejaarsstudenten te vinden, terwijl er in het derde jaar alleen een relatie was tussen inzet en prestatie.

Er zijn verschillende onderzoekers die zeggen dat de ontwikkeling van het zelfregulerend leervermogen afhankelijk is van de omgeving van de persoon. De studie die beschreven staat in *hoofdstuk 4* richt zich daarom op het verschil in de ontwikkeling van het zelfregulerend leervermogen in twee verschillende Braziliaanse medische curricula: een probleemgestuurd curriculum en een curriculum waarin voornamelijk colleges gegeven worden. De belangrijkste bevinding hier was dat in het college-gestuurde curriculum sommige vaardigheden - zoals planning, het monitoren van het leergedrag en het evalueren van het leergedrag - afnamen tijdens de geneeskundeopleiding. In het probleemgestuurde curriculum waren de studenten die in het zesde semester zaten iets beter in staat om te reflecteren dan de studenten die in het tweede semester zaten, maar ze toonden wel minder inzet. Deze bevindingen impliceren dat de omgeving dus inderdaad een grote rol speelt in de ontwikkeling van het zelfregulerend leervermogen.

In *hoofdstuk 5* is onderzocht hoe experts tot hun prestaties zijn gekomen. Volgens de onderzochte benadering bereikt niemand de top zonder hele lange en gestructureerde training. Deze gestructureerde training wordt ook wel *deliberate practice* genoemd. Het onderzoek naar de prestaties van experts vindt meestal plaats in een gestandaardiseerde omgeving waar representatieve taken kunnen worden uitgevoerd om zo het niveau van prestaties te meten. De geneeskunde is alleen een heel complex vakgebied waarin het moeilijk is om de prestaties van artsen met elkaar te vergelijken. Er is dan ook maar weinig onderzoek gedaan naar de invloed van *deliberate practice* op het prestatieniveau van artsen. Artsen in opleiding tot specialist, AIOS, worden echter wel beoordeeld in het kader van hun opleiding. Dit bood de kans om te onderzoeken of gedragsmatige kenmerken die passen binnen *deliberate practice* gerelateerd zijn aan het prestatieniveau van AIOS. Door middel van een vragenlijst waarbij gedrag van AIOS gemeten werd, en de beoordeling van de opleiders over het algemeen functioneren van de AIOS, liet deze studie zien dat de AIOS die in de bovenste twee tertielen zaten wat betreft hun prestaties vaker de basisliteratuur herlazen en vaker om feedback vroegen. Daarnaast waren ze opener over de fouten die ze maakten, wat noodzakelijk is om goede feedback op het functioneren te krijgen. Deze bevindingen laten zien dat *deliberate practice* in ieder geval deels van toepassing is binnen de geneeskunde. Er is echter verder onderzoek nodig om te kijken of deze verschillen in gedrag al zichtbaar zijn op het moment dat mensen zich aanmelden voor de geneeskunde-opleiding en of het dus mogelijk is om te voorspellen wie een veelbelovende toekomst als arts voor zich heeft.

Vervolgens komt in *hoofdstuk 6* de criterium-voorspeller-relatie aan bod. In deze studie zijn variabelen die bekend waren voordat een AIOS zich aanmeldde voor de heelkundeopleiding vergeleken met prestaties tijdens de heelkundeopleiding. Om te kunnen onderzoeken of een selectiemethode in staat is om te voorspellen hoe iemand presteert als arts, moet er een criterium beschikbaar zijn van wat een goede arts is. Een criterium van wat een goede arts is, bestaat alleen niet. De prestaties van AIOS worden, in tegenstelling tot die van andere artsen, wel beoordeeld. Hun prestatie is dus de meest optimale maatstaf voor prestatie die gebruikt kan worden om de effectiviteit van een selectiemethode te onderzoeken. In deze studie zijn biografische gegevens, informatie over prestaties tijdens de middelbare school en de geneeskundeopleiding, bijzondere academische prestaties en klinische-werkervaring gerelateerd aan scores op kennistoetsen en scores voor algemeen functioneren tijdens de AIOS-opleiding. Er blijkt een positieve relatie te bestaan tussen het gemiddelde preklinische cijfer en de score op zowel de kennistoets als het algemeen functioneren. Daarnaast haalden mannen hogere scores op de kennistoetsen dan vrouwen. Verder haalden, naast dat het gemiddeld preklinisch cijfer hieraan gerelateerd was, AIOS met een atheneumdiploma, met een hoger eindexamencijfer voor wiskunde en met minder klinische-werkervaring hogere scores wat betreft algemeen functioneren. Hoewel het niet is toegestaan om te selecteren op geslacht en een aantal van deze variabelen niet beschikbaar zijn wanneer iemand zich voor de geneeskundeopleiding aanmeldt, kan het eindexamencijfer voor wiskunde wel interessant zijn voor de selectie van geneeskundestudenten. Verder onderzoek is nodig om te kijken of dit cijfer ook gerelateerd is aan prestatie binnen andere specialistenopleidingen, en of het eindexamencijfer voor wiskunde een andere weging kan krijgen in de selectie voor geneeskundestudenten.

Hoofdstuk 7 bevat een algemene discussie over de bevindingen van alle studies in dit proefschrift. De discussie begint met een korte samenvatting van de belangrijkste bevindingen en overdenkingen voor geneeskundeopleidingen en selectiecommissies van de geneeskundeopleidingen. Deze overdenkingen zijn 1) om selectiemethoden te gebruiken die de cognitieve vaardigheden van de kandidaten meten, 2) het zelfregulerend leervermogen mogelijk te gebruiken in de selectie van geneeskundestudenten, maar nog belangrijker, om studenten te helpen bij het ontwikkelen van het zelfregulerend leervermogen, of in ieder geval ervoor te zorgen dat deze vaardigheden niet afnemenen 3) de elementen van *deliberate practice* toe te passen in de AIOS-opleiding, aangezien dit zal bijdragen aan het prestatieniveau van de AIOS. Daarnaast is het belangrijk om de aanwezigheid van deze gedragsmatige verschillen op het moment van aanmelding voor de geneeskundeopleiding te onderzoeken en 4) de relatie tussen variabelen die bekend zijn voor aanmelding en prestatie tijdens de AIOS-opleiding te onderzoeken in andere opleidingen dan heelkunde, om te kunnen bepalen of het eindexamencijfer voor wiskunde een andere weging kan krijgen in de selectie voor geneeskundestudenten. Daarnaast worden in de algemene discussie uitdagingen voor de geneeskundeselectie evenals ideeën voor verder onderzoek besproken.

DANKWOORD

Het is zover. Vier uitdagende, inspirerende, soms frustrerende maar bovenal bijzonder leerzame en waardevolle jaren zijn voorbij gevlogen. In de tijd die ik in dienst ben geweest bij het Erasmus MC heb ik niet alleen veel gelezen, geanalyseerd en geschreven, maar vooral veel mensen om mij heen gehad die mij geholpen hebben om dit project tot een goed einde te brengen. Een proefschrift schrijft je overduidelijk niet alleen, maar is het resultaat van kansen die je krijgt, mogelijkheden die je benut, samenwerkingen die je aangaat en support dat je ontvangt. Zonder de steun van mijn supervisors, collega's, vrienden en familie was het me niet gelukt om mijn proefschrift af te ronden en te verdedigen. Vandaar dat ik dit dankwoord wil gebruiken als een oprechte erkenning aan ieder die mij de afgelopen jaren heeft bijgestaan.

Eerst, mijn promotors, Prof.dr.ir. A.P.N. Themmen en Prof.dr. R.M.J.P. Rikers. Axel, dank je wel voor je vele uren begeleiding, je kritische blik en dat je hebt mij laten inzien dat er niet zoveel beren op de weg staan als ik vaak wil denken. Je hebt me geleerd wat de ongeschreven regels van de academie zijn en me gestimuleerd om zoveel mogelijk samenwerkingen aan te gaan en mensen in ons vakgebied te leren kennen, waardoor ik mijzelf verder kon ontwikkelen. Remy, jouw onuitputtelijke kennis heeft mij keer op keer geïnspireerd. Na elk gesprek met jou groeide mijn stapel "te lezen artikelen" en met enige regelmaat kwam er weer een boek bij in mijn boekenkast. Je eeuwige steun en vertrouwen waren bijzonder waardevol, vooral op momenten dat het even niet zo ging als ik graag zou willen. Dank je wel dat je altijd voor me klaar stond en dat ik alles wat op dat moment belangrijk voor mij was, met jou kon bespreken.

Daarnaast wil ik mijn commissieleden graag bedanken voor alle tijd en moeite die zij gestopt hebben in het lezen en evalueren van mijn manuscript: Prof.dr. M. de Hoog, Prof. dr. G.W.C. Paas, Prof.dr. G.R. Norman, Prof.dr. J.N.M. IJzermans, Prof.dr. H.G. Schmidt en Dr. S. Mamede. Naast dat u lid was van mijn commissie, heb ik ook het genoegen gehad om met enkelen van u samen te werken en veelvuldig om advies te mogen vragen. Prof. dr. de Hoog, dank voor de inspirerende samenwerking, de mogelijkheid om onze studie naar deliberate practice uit te mogen voeren bij uw AIOS en stafleden en het kijkje in de keuken dat ik bij kindergeneeskunde heb mogen nemen. Geoff, thank you for being my consultant throughout my entire PhD project, for your patience to answer my many questions, your ideas to improve my studies, your support and your time. It is an hounor to have you in my committee. Prof.dr. IJzermans, u wil ik graag in het bijzonder bedanken, dankzij u zit ik nu hier. Hartelijk dank dat u mij ruim vier jaar geleden in contact hebt gebracht met Axel en mij heeft voorgedragen voor dit promotietraject! Ik vind het erg bijzonder u aan twee studies een belangrijke bijdrage heeft geleverd en dat u plaats wilt nemen in mijn commissie. Sílvia, muito obrigada pela calorosa acolhida, pela sua generosidade e pelo fato de eu sempre poder bater em sua porta para fazer perguntas. Estou realmente agradecida pelo fato de ter participado da minha defesa e gostaria de desejar tudo de melhor para o seu future.

Veel dank ook voor mijn (overige) co-auteurs, die elk een belangrijke bijdrage geleverd hebben aan de studies in dit proefschrift. Ik heb van jullie allemaal iets waardevols mogen leren: inhoudelijk, qua statistiek of door zinnen die mooier geformuleerd werden dan ik zelf ooit voor elkaar had kunnen krijgen. Jullie hebben mij de mogelijkheid gegeven om rond te snuffelen bij verschillende geneeskundeopleidingen en medisch specialismen en jullie kritische en enthousiaste houding heeft mij keer op keer weer op scherp gezet. Bedankt voor alle inspirerende discussies en samenwerkingen!

Mijn collega's bij iMERR mogen natuurlijk ook niet ontbreken. Dank voor alle feedback tijdens en buiten de labmeetings en de interesse in mijn onderzoek. Karen en Laura, fijn dat ik altijd bij jullie binnen mocht lopen met al mijn vragen en onzekerheden. De gesprekken met jullie beiden hebben me enorm geholpen om keuzes te maken, deze te onderbouwen en om focus te houden. Ik wens jullie allebei veel succes toe met alle mooie studies die nog op de planning staan, en duim dat jullie die subsidies binnenhalen! Mijn roomies, Wendy en Lokke, wat heb ik een geluk gehad om jullie als kamergenootjes te krijgen. We hebben veel gelachen, mooie plannetjes gemaakt, een heleboel slingers en ballonnen opgehangen en kopjes koffie gedronken. Jullie waren altijd bereid om mee te denken over vragen en uitdagingen waar ik tegen aan liep en zorgden er wel voor dat elke werkdag een feestje was. Ik ga jullie missen! Succes met de afronding van jullie proefschriften, ik weet zeker dat jullie over een paar jaar allebei een prachtig boekje mogen verdedigen!

Dank ook aan het management en alle (oud-)collega's van de Desideriusschool, voor alle gesprekken, kopjes koffie, borrels, etentjes, vertrouwen en steun. Jullie zijn allemaal zulke bijzondere en lieve mensen, maak er wat moois van samen. Er zijn een paar collega's die ik graag in het bijzonder wil bedanken. Gonny, super fijn dat je mijn proefschrift wilde controleren op spelfouten en vreemd samengestelde zinnen. Willeke, dank voor het uit handen nemen van alle administratieve rompslomp rondom mijn promotie. Het gaf een hoop rust dat ik erop kon vertrouwen dat jij wist wat er wanneer geregeld moest worden. Rianne, bedankt voor de ongelofelijke hoeveelheid vragenlijsten die je voor me nagekeken en ingevoerd hebt. Wat een monnikenwerk was dat, maar je deed het allemaal bijzonder nauwkeurig, erg fijn!

Eu gostaria de agredecer a todos os meus amigos e colegas do Brasil. Obrigada por me fazerem sentir bem-vinda como se eu estivesse no meu próprio lar em Belo Horizonte. Obrigada a todos! Eu tiva uma ótima estadia no Brasil; a minha visita ao Brasil foi uma experiência incrível que eu nunca irea esquerer. Desejo muito revê-los o mais breve possível.

Verder zijn er nog een aantal mensen die enorm veel voor mij gedaan hebben gedurende mijn hele promotietraject en absoluut een plekje verdienen in dit dankwoord. Sofie, jij hebt aan elk van mijn studies wel een bijdrage geleverd. Dank je wel dat je er altijd was, me elke keer het gevoel gaf dat ik welkom was en elke vraag zorgvuldig onder de loep nam. Ik heb enorm veel van je mogen leren en voel me gezegend dat ik jou als "extraatje" bij Remy kreeg. Marian, wat had ik zonder jou gemoeten, nooit vroeg ik je te veel en bij elke vraag kreeg ik een gezellig gesprek cadeau. Na mijn promotie wil ik je heel graag met een wijntje bedanken in plaats van met de gebruikelijke kop koffie!

Ook de Expertgroep Decentrale Selectie, NVMO Promovendinetwerk, ICO en alle mensen van INReSH mogen niet vergeten worden. Jullie feedback op presentaties en brainstormsessies hebben een positieve bijdrage geleverd aan mijn proefschrift. Het was erg fijn om, ondanks onze raakvlakken wat betreft onderzoek, er een groep collega's bij

te hebbenen niet het gevoel te hebben dat we concurrenten van elkaar waren. Nienke en Anouk, jullie in het bijzonder bedankt voor alle kennis die jullie gedeeld hebben, voor alle gezellige lunches, leuke en lieve mailtjes en berichtjes en de vele borrels die de komst naar meetings extra leuk maakten. Nog even en dan zijn jullie ook klaar, ik heb nu al zin om jullie boekjes te lezen!

Mijn proefschrift was ook nooit tot stand gekomen zonder de medewerking van de vele studenten, AIOS, artsen en secretaresses van alle artsen die vragenlijsten voor mij hebben ingevuld, de tijd hebben genomen om met mij te brainstormen, afspraken voor mij hebben ingepland en waarbij ik interviews af heb mogen nemen. Hartelijk dank!

Naast alle collega's waren er de afgelopen vier jaar ook een heleboel vrienden die mij de nodige support, afleiding en gezelligheid bezorgd hebben. Veel van jullie hebben stukken tekst voor mij nagelezen, maar er vooral voor gezorgd dat ik soms echt kon afschakelen van mijn werk! Madelon, er is werkelijk niets waarvoor ik je niet kan opbellen of mee langs kan komen. Naast dat onze gesprekken ontzettend gezellig zijn, heb je me ook vaak zat met mijzelf geconfronteerd waardoor ik mezelf beter heb leren kennen, dank je wel voor alles. Sandra en Ruth, jullie waren tijdens mijn promotietraject niet alleen vriendinnetjes, maar ook een tijdje collega's. Fijn om samen ervaringen te delen en samen te mogen groeien. Ik heb de kelder bezoekjes gemist de laatste periode! Elma, ik heb je leren kennen via ICO, maar de vele vrijdagen die we samengewerkt hebben, de koffietjes, etentjes en borrels hebben je meer vriendin gemaakt dan collega. Het was erg fijn om iemand te hebben die in hetzelfde schuitje zat en waaraan je werkelijk alles kwijt kunt, dank je wel. Chris, jouw input is terug te vinden in de stellingen en in hoofdstuk drie, maar weet dat je zoveel meer voor me gedaan en betekend hebt. Maud, Roel, Rob, Arnold, Kim en Janneke, jullie laten zien dat afstand niets uitmaakt. Hoeveel of weinig we elkaar ook spreken, ik weet dat jullie er altijd voor me zijn en bij elk gesprek voelt het toch alsof we elkaar gisteren nog gesproken hebben. Maartje, ontzettend leuk dat we weer met elkaar in contact gekomen zijn, de borrels en festivalletjes met jou zijn onvergetelijk! Lonneke, onze vriendschap gaat ruim 20 jaar terug, zullen we daar nog een heleboel jaar aan vast plakken?

Mijn paranimfen en vriendinnetjes, Maya en Merel. Maya, naast alle gezelligheid heb je me ook enorm geholpen met de opmaak van dit boekje. Wat was het fijn dat jij wijs kon uit al die drukkerstaal, je hebt me daar zoveel stress mee uit handen genomen. Merel, werkelijk niemand kon zo enthousiast zijn over een publicatie of andere mijlpaal die ik tijdens dit promotietraject bereikt heb. Ik kan met jullie lachen en huilen, fijn met jullie bijkletsen met een theetje op de bank of urenlang dansen in de stad. Alles is goed. Bedankt dat jullie er altijd voor me zijn en dat jullie mij tijdens mijn promotie bij willen staan. Op naar nog vele jaren vriendschap en een bejaardendisco op Ibiza!

Alle andere lieve vrienden, jullie zijn met teveel om allemaal persoonlijk te benoemen, maar ontzettend bedankt voor alle vriendschappen, gezelligheid, betrokkenheid, de kopjes koffie, etentjes, borrels, terrasjes, feestjes, telefoontjes en berichtjes. Ik voel me echt gezegend met zoveel lieve, mooie mensen om me heen!

Mijn familie heeft ook een belangrijke rol gespeeld tijdens het schrijven van dit proefschrift, bedankt voor jullie interesse en betrokkenheid! Victoire, Jan, Pieter, Victoria, Elly, Marianne en Ellen, jullie hebben de afgelopen jaren enorm met me meegeleefd. Jullie enthousiasme over mijn publicaties was vaak groter dan die van mijzelf. Vic en Jan, volgens mij is er niemand die mijn artikelen zo grondig leest als jullie. Piet, dank je wel voor je luisterend oor en dat jij en Victoria mij een heel lief nichtje hebben gegeven dat me bijzonder goed kan afleiden van mijn werk. Elly, dank je wel voor je knuffels! Zus Jan, een extra bedankje voor het controleren van mijn Nederlandse samenvatting, leuk dat er op die manier ook iets van jou in mijn proefschrift staat. Ellen, dank je wel voor je interesse en medeleven. Het was erg fijn dat jij door je eigen promotie ervaring goed kon aanvullen hoe dingen in elkaar staken.

Lieve pap, jij hebt me werkelijk in elk mogelijk opzicht gesteund en aangemoedigd, tijdens dit traject, voordat ik met mijn promotie startte en jou kennende zal je dit hierna ongetwijfeld ook blijven doen. Je bent altijd geïnteresseerd, leeft mee, denkt mee en laat me zien dat je trots bent. Ik kan me geen betere vader voorstellen, dank je wel voor alles, voor altijd. Lieve mam, ook al was jij niet direct aanwezig de afgelopen jaren, het zou niet juist zijn om jou hier niet te benoemen. Jouw positieve houding heeft mij veel kracht gegeven en geïnspireerd om vooral de zonnige kant van dingen te bekijken, iets dat de afgelopen jaren bijzonder goed van pas kwam. Ik hoop dat je op de een of andere manier mee krijgt hoe het met ons gaat en trots bent, vooral op jezelf. Je hebt ons goed achtergelaten.

Rotterdam, oktober 2015


PAPERS AND
PRESENTATIONS

PAPERS

Susanna M. Lucieer, Karen M. Stegers-Jager, Remy M. J .P. Rikers & Axel P. N. Themmen (2015). Non-cognitive selected students do not outperform lottery-admitted students in the pre-clinical stage of medical school. *Advances in Health Sciences Education*, DOI 10.1007/s10459-015-9610-4

Susanna M. Lucieer, Laura Jonker, Chris Visscher, Remy M. J. P. Rikers & Axel P. N. Themmen (2015). Self-Regulated Learning and Academic Performance in Medical Education. *Medical Teacher*, DOI 10.3109/0142159X.2015.1073240

Susanna M. Lucieer, Jos N. van der Geest, Silvana M. Elói-Santos, Rosa M. Delbone de Faria, Laura Jonker, Chris Visscher, Remy M. J. P. Rikers & Axel P. N. Themmen (2015). The development of self-regulated learning in the pre-clinical stage of medical school: a comparison between a lecture-based and a problem-based curriculum. *Advances in Health Sciences Education,*
DOI 10.1007/s10459-015-9613-1

Susanna M. Lucieer, Axel P. N. Themmen, Robert Jan Stolker, Matthijs de Hoog, Yvo M. Smulders, Abel Thijs, Jan N. M. IJzermans & Remy M. J. P. Rikers. Identifying deliberate practice activities in residency training. *Submitted for publication*

Susanna M. Lucieer, Cornelis J. Hopmans, Pieter T. den Hoed, Reinier Timman, Rosemarijn T. Steensma, Remy M. J. P. Rikers, Axel P. N. Themmen & Jan N. M. IJzermans. Identification of factors associated with performance during surgical residency training. *Submitted for publication*

Karen M. Stegers-Jager, Ewout Steyerberg, Susanna M. Lucieer & Axel P. N. Themmen (2014). Ethnicity and Social Background as Predictors of Successful Application for Medical School. *Medical Education*, 49, 124-133

Nienke R. Schripsema, Anke M. van Trigt, Susanna M. Lucieer, Anouk Wouters, Gerda Croiset, Axel P. N. Themmen, Jan C. C. Borleffs, Janke Cohen-Schotanus. Medical school admissions and performance: Is there a participation or a selection effect? *Submitted for publication*

Wendy E. de Leng, Karen M. Stegers-Jager, Susanna M. Lucieer, Maarten Frens, Axel P. N. Themmen. A Scientific Pre-University Program increases Medical Students' Interest in an Academic Career. *In preparation*

ORAL PRESENTATIONS

Susanna M. Lucieer*, Karen M. Stegers-Jager, Remy M. J. P. Rikers & Axel P. N. Themmen. Studiesucces van cognitief versus non-cognitief geselecteerde geneeskundestudenten. NVMO (Nederlandse Vereniging voor Medisch Onderwijs) conference. November 7th, 2014, Egmond aan Zee, the Netherlands

Susanna M. Lucieer*, Remy M. J. P. Rikers & Axel P. N. Themmen. Self-regulated learning, a new selection criterion? Graduate Research Day of the Erasmus University Rotterdam, October 2nd, 2014, Rotterdam, the Netherlands

Janke Cohen-Schotanus*, Nienke R. Schripsema, Axel P. N. Themmen, Susanna M. Lucieer, Anouk Wouters & Gerda Croiset. Lottery as selection procedure: how (un)fair is it? Ottawa Conference, April 28th, 2014, Ottawa, Canada

Susanna M. Lucieer*, Karen M. Stegers-Jager, Remy M. J. P. Rikers & Axel P. N.Themmen. Academic Performance of cognitive versus non-cognitive selected students. Ottawa Conference, April 27th, 2014, Ottawa, Canada

Karen M. Stegers-Jager*, Susanna M. Lucieer & Axel P. N. Themmen. Ethnicity and social background as predictors of successful application for medical school. Ottawa Conference, April 27th, 2014, Ottawa, Canada

Susanna M. Lucieer*, Jos N. van der Geest, Silvana M. Eloi-Santos, Rosa M. Delbone de Faria, Chris Visscher, Laura Jonker, Remy M. J. P. Rikers & Axel P. N. Themmen. The development of Self-Regulated Learning in Medical Education: A comparison between a Lecture-Based and a Problem-Based Curriculum. ICO Fall school, November 8th, 2013, Maastricht, the Netherlands

Susanna M. Lucieer*, Axel P. N. Themmen & Remy M. J. P. Rikers. De ontwikkeling van het zelfregulerend leervermogen van medische studenten in Brazilië – een vergelijking tussen een traditionele en een probleem gestuurde leeromgeving. NVMO PhD-day, April 28th, 2013, Utrecht, the Netherlands

POSTER PRESENTATIONS

Susanna M. Lucieer, Cornelis J. Hopmans, Pieter T. den Hoed, Reinier Timman, Rosemarijn T. Steensma, Remy M. J. P. Rikers, Jan N. M. IJzermans & Axel P. N. Themmen*. Correlation between pre-training variables and performance during surgical residency training: Is surgical competence predictable? AMEE conference, September 9th, 2015, Glasgow , United Kingdom

Susanna M. Lucieer*, Wendy E. de Leng, Remy M. J. P. Rikers & Axel P. N. Themmen. Self-regulated learning, a new selection criterion? INReSH meeting, November 10th-11th 2014, London, United Kingdom

Janine van der Rijt*, Susanna M. Lucieer, Matthijs de Hoog & Wim H. Gijselaers. Encouraging residents' feedback seeking behaviour. Ottawa Conference, April 27th, 2014, Ottawa, Canada

Susanna M. Lucieer*, Axel P. N. Themmen & Remy M. J. P. Rikers. Selecting excellent health care providers – not just promising medical students. Symposium Teacher Expertise, October 11th, 2012, Utrecht, the Netherlands

Susanna M. Lucieer*, Axel P. N. Themmen & Remy M. J. P. Rikers. Selecting excellent health care providers – not just promising medical students. Graduate Research Day, Erasmus University Rotterdam, October 4th, 2012, Rotterdam, the Netherlands

SYMPOSIA

Susanna M. Lucieer*, Anouk Wouters, Geoffrey R. Norman, Kelly Dore, Fiona Patterson, Axel P. N. Themmen & Gerda Croiset. Selection methods in medical school: Where are we now and where are we heading? AMEE conference, August 27th, 2013, Prague, Czech Republic

*presenter

CURRICULUM VITAE

Curriculum Vitae

Susanna Magdalenda Lucieer was born on October 21, 1985, in Rotterdam, the Netherlands. She completed her secondary education in 2004 at the Roncalli Scholengemeenschap in Bergen op Zoom. In the same year, she started her studies in General Health Sciences, with a focus on Mental Health Sciences, at Maastricht University. She obtained her Bachelor's degree in 2008, and continued her study with a Master, also at Maastricht University, in International Business; Organizations, Management, Change and Consultancy which she completed in 2010 with a thesis on the labour mobility of medical specialists in the Netherlands. During her studies, she spent considerable time in committees of her student society and worked as a project manager at the T&E Group, where she, among others, set up a hostess service for the MECC and TEFAF. After her studies, she had several short-time jobs in sales, as a policy maker and as a research assistant. In June 2011, she started her PhD research on medical school selection at the Institute of Medical Education Research Rotterdam, Erasmus Medical Center. During her PhD research, she spent two months in Belo Horizonte, Brazil, to collect data for one of her studies. This PhD research resulted in various national and international presentations, publications and collaborations. As of November 2015, she started working as a consultant at VvAA.

PHD PORTFOLIO

SUMMARY OF PHD TRAINING AND TEACHING

Name PhD student: Susanna M. Lucieer	**PhD period:** June 2011 - August 2015
Erasmus MC Department: Institute for Medical Education Research Rotterdam (iMERR)	**Promotor(s):** Prof. dr. ir. A. P. N. Themmen Prof. dr. R. M. J. P. Rikers

	Year	Workload (ECTS)
General courses		
- Statistics; regression analysis	2011	2
- Biomedical English Writing and Communication	2013	4
- Research Integrity	2013	2
Specific courses (e.g. Research school, Medical Training)		
- Introductory course ICO	2011	7
- Qualitative research	2013	3
- Multilevel analysis	2014	3
- Excel training	2015	0.25
Total ECTS courses		**21.25**
Seminars and workshops		
- Endnote workshop	2011	1
- NVMO workshop “kwalitatief onderzoek in medisch onderwijs”	2011	1
- VPO workshop “digitale vragenlijsten”	2011	1
- Seminar Sport & Onderwijs UMCG	2011	1
- Graduate Research Day EUR	2012	1
- Feedback workshop	2012	1
- Seminar Selection	2012	1
- Seminar Metacognition	2012	1
- Seminar Teacher Expertise	2012	1
- Seminar Non-cognitive selection	2013	1
- Ottawa Preconference workshop	2014	1
Total ECTS seminars and workshops		**11**
Presentations		
- Poster presentation Graduate Research Day EUR *Selecting excellent healthcare providers – not just promising medical students. Rotterdam, the Netherlands*	2012	1
- Poster presentation seminar teacher expertise UU *Selecting excellent healthcare providers – not just promising medical students. Utrecht, the Netherlands*	2012	1
- Oral presentation NVMO promovendidag *De ontwikkeling van het zelfregulerend leervermogen van medische studenten in Brazilië – een vergelijking tussen een traditionele en een probleem gestuurde leeromgeving. Utrecht, the Netherlands*	2013	1

	Year	Workload (ECTS)
- Oral presentation ICO Fall School *The development of Self-Regulated Learning in Medical Education: A comparison between a Lecture-Based and a Problem-Based Curriculum. Maastricht, the Netherlands*	2013	1
- Oral presentation Ottawa Conference *Academic Performance of cognitive versus non-cognitive selected students. Ottawa, Canada*	2014	1
- Oral presentation NVMO conference *Studiesucces van cognitief versus non-cognitief geselecteerde geneeskundestudenten. Egmond aan Zee, the Netherlands*	2014	1
- Oral presentation Graduate Research Day EUR *Self-regulated learning, a new selection criterion? Rotterdam, the Netherlands*	2014	1
- Poster presentation INReSH meeting *Self-regulated learning, a new selection criterion? London, United Kingdom*	2014	1
- E-poster presentation during AMEE Conference *Relation between pre-training variables and performance during surgical residency training. Glasgow, United Kingdom*	2015	1
Total ECTS presentations		**9**
Other		
- Symposium selection in medical schools AMEE conference, 2013 Prague, Czech Republic	2013	1
- International research meetings in Ottawa & London	2014	1
- iMERR research meetings	2014/ 2015	1
Total ECTS other		**3**
Lecturing		
- Course 1.6 *"Normaal of Abnormaal, inleiding klinische psychologie"* (Psychology, Erasmus University Rotterdam) Supervising practicals and excursions, Tutoring	2012	4
- Supervising students honours class – questionnaire development	2012	0.5
Other		
- Research self-regulated learning Brazil		
- Organisation PhD Day Erasmus MC 2013		
Total ECTS		48.75