Readiness for clinical practice

Studies about transitions in medical education, the influence of vertically integrated curricula and the assessment of readiness for practice

Marjo Wijnen-Meijer

Beoordelingscommissie:

Prof. dr. E.W.M.T. ter Braak Prof. dr. M.R. van Dijk Prof. dr. F. Scheele Prof. dr. J.W.F. van Tartwijk Prof. dr. C.P.M. van der Vleuten

ISBN 978-90-393-5866-5 © 2012, Marjo Wijnen-Meijer

All rights reserved. No part of this thesis may be reproduced, stored in retrieval systems, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior permission of the author.

A kind thanks for the use of their photographs goes to: Beeldbank UMC Utrecht (pages 22–23, 36–37, 52–53, 66–67, 84–85, 94–95, 110–111, 120–121, 146–147, 164–165 and 178–179), Vincent Boon (pages 202–203 and 224–225), Emmanuel Keller (pages 232–233), John Picton (pages 256–257), Alex E. Proimos (pages 1–2).

Readiness for clinical practice

Studies about transitions in medical education, the influence of vertically integrated curricula and the assessment of readiness for practice

Voorbereid op de klinische praktijk (met een samenvatting in het Nederlands)

Bereit für die klinische Praxis (mit einer Zusammenfassung in deutscher Sprache)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 21 februari 2013 des middags te 12.45 uur

door

Marjo Wijnen-Meijer geboren op 10 november 1971 te Leersum

Promotoren: Prof. dr. Th.J. ten Cate Prof. dr. J.C.C. Borleffs

Co-promotor: Dr. M.F. van der Schaaf

Table of contents

1

General introduction

A description of the context of the thesis, based on the developments in the field of medical education in relation to educational theories; a specification of the purpose, research questions and the outline of the thesis.

1

Introduction•3; Expertise development•4; Transitions in medical education•7; Curricula at medical schools•9; Measurement of readiness for practice•14; Purpose of the thesis•15; Research questions•17; Methodology•19; Outline of the thesis•20

2

Stages and transitions in medical education around the 23 world: clarifying structures and terminology

An overview of stages and transitions in medical education including the moment that unrestricted practice is allowed, in 40 different countries, based on the results of a questionnaire study and a document analysis.

Abstract•25; Introduction•26; Methods•28; Results•29; Discussion•35

3

The impact of various transitions in the medical education 37 continuum on perceived readiness of trainees to be entrusted with professional tasks

The influence of different types of transition in medical education on the perceived competence development and readiness for practice of trainees: a questionnaire study among trainees and supervisors.

Abstract•39; Introduction•40; Methods•43; Results•45; Discussion•48; Conclusions•51

4 The influence of a vertically integrated curriculum on the 53 transition to postgraduate training

A questionnaire study among medical graduates from two consecutive cohorts of the Utrecht Medical School (one vertically integrated and one non-vertically integrated curriculum) to determine whether a vertically integrated curriculum at medical school affects the transition to postgraduate training.

Abstract•55; Introduction•56; Method•59; Results•60; Conclusion•65

5

Vertical integration in medical school: effect on the transi- 67 tion to postgraduate training

A questionnaire study among medical graduates of one cohort from six medical schools (three with a vertically integrated and three with a non-vertically integrated curriculum) to determine whether a vertically integrated curriculum at medical school affects the transition to postgraduate training.

Abstract•69; Introduction•70; Research questions•72; Methods•74; Results•76; Discussion•80

6

Postgraduate supervisors value graduates from a vertically 85 integrated undergraduate curriculum

A questionnaire study among supervisors of postgraduate training programs in Utrecht (vertically integrated curriculum) and Hamburg (non-vertically integrated curriculum) about their medical graduates' preparedness for work, knowledge and capabilities.

Abstract•87; Introduction•88; Methods•90; Results•91; Discussion•93

7

Essential facets of competence that enable trust in 95 graduates: a Delphi study among physician educators in the Netherlands

A Delphi study of two rounds among experienced physician educators to uncover essential facets of competence that determine decisions to entrust a trainee with critical clinical tasks.

Abstract•97; Introduction •98; Methods •99; Results•104; Discussion•108; Conclusions•109

8

Essential facets of competence that enable trust in medical 111 graduates: a ranking study among physician educators in two countries

A ranking study among physician educators in the Netherlands and Germany to answer the question whether clinical educators in different educational climates agree on the importance of the facets of competence resulting from the Delphi study.

Abstract•113; Introduction•114; Methods•115; Results •117; Discussion•118; Conclusion•119

9

An argument-based approach to the validation of 121 UHTRUST: can we measure how recent graduates can be trusted with unfamiliar tasks?

The evaluation of the validity of an authentic simulation procedure to assess whether medical graduates are ready to be entrusted with unfamiliar clinical tasks, following the argument-based approach for validation.

Abstract•123; Introduction•124; Method•129; Results: argument for validation of UHTRUST•134; Discussion•144

10

Vertically integrated medical education and the readiness 147 for practice of graduates

Comparison of two groups of medical graduates who followed either a vertically integrated curriculum or a non-vertically integrated curriculum in their competences to cope with unfamiliar clinical situations.

Abstract•149; Introduction•150; Method•153; Results•157; Discussion•161

11

General discussion

Summary and discussion of the findings; discussion of limitations and strengths as well as implications for educational practice and suggestions for further research.

Introduction•167; Main findings of the thesis•170; Strengths and limitations•173; Suggestions for the practice of medical education•175; Implications for further research•176; In conclusion•177

12

Summary Summary of the thesis in English, German and Dutch. Summary•180; Zusammenfassung•180; Samenvatting•180

179

165

A Example of a UHTRUST Case

During the UHTRUST assessment candidates had to face five challenging patient cases and seven distracting tasks. Appendix A contains a description of one of the cases, including two disturbances which are related to this patient case.

Description for assessors •205; Description for standardized patient •211; Letter from emergeny unit •215; Results additional examinations •216; Role descriptions disturbances for this case •220

В

Scoring forms UHTRUST

For the UHTRUST assessment five scoring forms were used: three were completed by physicians, one by nurses and one by standardized patients. Appendix B presents (parts of) all scoring forms.

Facets of competences (physicians)•227; Entrustable Professional Activities (physicians)•228; Post Patient Encounter Form (physicians)•229; Facets of competences (nurses)•230; CARE Questionnaire (standardized patients)•231

References	233
Dankwoord/Acknowledgements	251
Curriculum vitæ	255
Index (chapters)	257



General introduction

Introduction

The research project described in this thesis was carried out in the field of medical education. Medical education consists of initial (undergraduate) training at medical school and additional postgraduate training. This thesis is related to the main goal of medical education: to prepare the trainees for working in the clinical practice, by stimulating their development into competent physicians. This research project had both practical and theoretical relevance. It aims to contribute to the knowledge about vertical integration and transitions in medical education in relation to expertise development of trainees.

In addition, this project aims to be useful for educators and curriculum developers in the field of medical education. The first section of this chapter describes the context of the thesis, based on developments in the field of medical education in relation to educational theories. Next, the purpose and research questions are defined. Finally, the structure of this thesis will be described.

Expertise development

Expertise and competency

Experts are distinguished by non-experts, because of their knowledge and skills in a specific field, which makes them able to solve problems in that field. For instance, medical experts are able to diagnose a disease, to prescribe and to perform surgery correctly. Expertise can be seen as the skills, knowledge and characteristics of an expert. Ericsson *et al.* (2006) state that expertise has three characteristics. First, expertise leads to performance that is consistently better than the performance of others. Second, real expertise produces concrete outcomes. Third, expertise can be replicated and measured.

Another term that is often used in the context of education is competency, which originally means "the ability of an individual to do something successfully" (Ten Cate & Scheele, 2007, p543). Next to the ability to do something, in many situations competency also has the meaning of "having the permission or authority to do it" (Sultana, 2009, p19).

Competency is often defined as the integration of knowledge, skills and attitude (Baartman & De Bruijn, 2011). Several authors have defined the features of competencies more specifically. Based on this literature, competencies are: specific, integrative, durable, focused on performance, learnable and mutually dependent (Van Merriënboer *et al.*, 2002). In addition, competencies should reflect external expectations and should lead to performance that is measurable using absolute standards (Albanese *et al.*, 2008). Furthermore, competencies are often seen as context-dependent (Ten Cate *et al.*, 2010).

A competency can be seen as a component or subset of expertise. A competency is context-specific performance within the domain of the individual's expertise (Herling, 2000). Therefore, for the achievement of relevant competencies, expertise development of trainees is needed.

Expertise development in medical education

The goal of medical education is to prepare their trainees for functioning as professionals in the complex clinical context. Several authors have described the process of expertise development, with a focus on consecutive phases.

Dreyfus & Dreyfus (Dreyfus & Dreyfus, 1986; Dreyfus, 2004) described the "five stage model of skill acquisition", which actually refers to professional development. In this model *skill* is not just a routine, but it involves also perceptions and the decisions to use the routines (Eraut, 1994). The model describes the following stages of development: novice, advanced beginner, competence, proficiency and expertise. The novice has to start with learning the rules and applies these rules rigidly. An expert on the other hand can behave flexibly and intuitively, adjusting to the situation (Dreyfus, 2004).

A second model of expertise development focuses on clinical reasoning by physicians and consists of four phases each with a typical knowledge structure. In the first stage medical trainees build causal knowledge about diseases in terms of pathophysiological processes. In second stage this knowledge develops into high-level causal models, in which information about diseases is classified under diagnostic labels. When the trainees have experience with real patient cases they reach the third stage, in which information is organized in narrative structures, so called *illness scripts*: coherent and generalized information about diseases. Experienced physicians will reach the fourth stage of development, in which they use memories of previous patients in the diagnosis of new patients (Schmidt *et al.*, 1990).

These two models emphasize learning from experience. This matches the idea that deliberate practice is important for the development of expertise (Ericsson *et al.*, 1993; Duvivier *et al.*, 2011). To reach expert performance, skills need to be practiced continually at more challenging levels over a longer period of time (Ericsson *et al.*, 1993). In addition it is needed that trainees consciously plan their learning activities and that they receive focused feedback from their supervisors (Duvivier *et al.*, 2011).

Support of expertise development

Medical training programs are effective if they are organized in such a way that they make it possible for the trainees to pass through the phases of expertise development and consequently to develop themselves to medical experts.

Vygotsky shows in his theory about the Zone of Proximal Development (ZPD) that the aim of education should be the development of psychological functions (Vygotsky, 1978; Chaiklin, 2003). Trainees need assistance by peers and more competent persons to undergo this development, for instance through modeling or scaffolding (Collins *et al.*, 1989). Based on this ZPD theory, Dunphy & Williamson described four general stages of learning in relation to the required amount of assistance (Dunphy & Williamson, 2004). In the first stage of external regulation the trainee is not able to perform a particular task independently and needs help from experienced others in carrying

out this task. The second stage is that of internal regulation. In this stage, the trainee is able to carry out the task without assistance, but the performance is not yet fully developed. In the third phase, the task execution has been fully internalized and automatized and assistance is no longer needed. If trainees are faced with major changes or stress it often happens that performance is no longer automatized and the trainee returns to stage one or two. This is designated as the fourth stage of the developmental process

Transitions in medical education

Medical trainees make multiple transitions during their period of education. These include transition from theoretical to clinical training, transition to an increased level of responsibility and seniority, transition from undergraduate to postgraduate training, transition from preregistration to registration and transitions to other hospitals or departments.

The moments and types of transitions differ around the world, but all trainees experience the same two major transitions. The first one is the transition from preclinical to clinical education and the second the transition from medical school to postgraduate training.

A number of studies have reported that many trainees experience these transitions as stressful and demanding. Trainees do not feel well prepared for the next step or do not have enough confidence for their new role with more responsibilities (Moss & McManus, 1992; Whitehouse *et al.*, 2002; Goldacre, 2003; Lempp, 2004; Van Hell *et al.*, 2008; Cave, 2009; Brennan, 2010). They often experience difficulties in the application of knowledge and skills, such as taking histories and performing physical examinations (Prince *et al.*, 2000; Van Hell *et al.*, 2008; Chittenden *et al.*, 2009). Some problems are related to the process of professional socialization, as finding their way in the organization and adjusting to the clinical context (Chittenden *et al.*, 2009; Kilminster *et al.*, 2010). Furthermore, several studies indicate that there are increasing safety risks for patients in periods that many newly qualified doctors start to work (Haller *et al.*, 2009; Jen *et al.*, 2009; Young *et al.*, 2011).

Kilminster *et al.* (2010) define transitions as "critically intensive learning periods", with possible implications for the performance of the trainees during this period. This matches the stage of de-automatization in the model of Dunphy & Williamson (2004). A concept that is important for the understanding of difficulties around transitions is transfer: the ability of trainees to apply knowledge and skills that are learned in one setting in another setting (Gick & Holyoak, 1983, Salomon & Perkins, 1989). Salomon & Perkins (1989) described two main types of transfer: low-road transfer and high-road transfer. Low-road transfer is the transfer of highly practiced skills almost automatically to a new context. High-road transfer concerns an intentional and conscious abstraction of something from one context and application in a new context.

Both types of transfer are needed for medical trainees to become *adaptive experts* in the ever-changing clinical context. Adaptive experts are able to carry out routine tasks efficiently and, next to it, to find solutions to new problems and to adapt their way of working to new circumstances (Bereiter & Scardamalia, 2003; Mylopoulos & Regehr, 2009).

Although it is not possible to prepare the medical trainees for every aspect of their work in the new environment (Kilminster *et al.*, 2010), it is important to know whether the medical school curriculum equip the students well enough for the next phase, and especially for the transfer of what is learned (Bransford & Schwartz, 1999).

Curricula at medical schools

Traditionally, the medical school curriculum was built upon the different medical disciplines or specialties (Papa & Harasym, 1999). The knowledge and skills belonging to these disciplines were taught separately from and after each other. Students had to make the connections between the different subjects on their own (Harden *et al.*, 1984; Papa & Harasym, 1999; Vidic & Weilauf, 2002) and between the theory about diseases and patient cases. This is an example of high-road transfer, which is difficult for inexperienced trainees (Salomon & Perkins, 1989; Fallon *et al.*, 2009; Norman, 2009).

Although many medical school curricula are still discipline based, since the second half of the 20th century curriculum developers have promoted integrated learning (Harden et al., 1984; Vidic & Weilauf, 2002). The idea behind integration is that it is easier to learn and apply knowledge, when the information fragments are connected to each other in a meaningful way and when there is similarity between the learning context and the context in which the knowledge is retrieved (Regehr & Norman, 1996). Integration is also one of the education strategies in the SPICES model (student-centered problem-based integrated community-based electives systematic) for curriculum development. This model describes six curriculum issues and presents each of these issues as a spectrum between two extremes and can be helpful for curriculum developers to make choices regarding the education strategies (Harden *et al.*, 1984). Also newer curriculum development frameworks emphasize the importance of integration, for instance the six-step approach for curriculum development (Kern et al., 1998), the PRISMS framework (product related relevant interprofessional shorter/smaller multi sites; Bligh et al., 2001) and the RIFLE framework (realistic integrated feedback learning evaluation; Bligh & Brice, 2010).

Horizontal integration

The first type of integration is called horizontal integration. This means that different disciplines, taught in the same year or phase of the curriculum, are combined around themes or patient cases (Harden *et al.* 1984; Prince *et al.*, 2000). The change from a discipline-based into a system-based approach led to the first form of a horizontal integrated medical curriculum. This means that all relevant knowledge about for instance the cardiovasculair or musculoskeletal system is taught in the same period (Harden *et al.*, 1984; Leinster, 2009).

This kind of integration is increasingly important because of the growth of medical knowledge and the tendency towards subspecialization (Cooke *et al.*, 2010). An early example of horizontal integration is the new curriculum that was developed at Case Western Reserve University in the 1950s. In the pre-clinical curriculum knowledge from anatomy, physiology, biochemistry, pathophysiology, pharmacology and epidemiology was combined in courses about organs or body systems (Williams, 1980).

In many medical schools horizontal integration is implemented as problem-based learning or case-based learning. In these approaches, students work in small groups and they gather knowledge about a subject in the context of realistic problems or patient cases. Next to this, problem-based learning aims to develop higher cognitive skills as generating learning goals and problem solving (Foley *et al.*, 1997; Norman & Schmidt, 1992; Albanese, 2000; Prince *et al.*, 2000; Prideaux, 2009).

These learning approaches stimulate transfer from basic science to clinical cases, because the concepts are learned in relation with problem contexts (Bransford, 1999; Norman, 2009). These approaches also improve the building of a knowledge framework which is needed for clinical reasoning (Schmidt *et al.*, 1990).

Although the primary goal of horizontally integrated learning approaches was to improve the preparation of students for clinical practice (Albanese, 2000; Norman & Schmidt, 1992), the distinction between the theoretical pre-clinical phase and the practical clinical phase still existed (Foley *et al.*, 1997; Margetson, 1999). As a result, the medical students experienced difficulties in making the transition from theoretical to clinical education (Prince *et al.*, 2000; Chittenden *et al.*, 2009).

Vertical integration

To improve the connection between theoretical and practical training and from undergraduate medical school to postgraduate training, medical curricula have reformed into vertically integrated curricula. Vertical integration means that the disciplines are arranged around themes which are programmed in consecutive years of the curriculum (Harden, 1984; Prideaux, 2009). This kind of integration implies that basic sciences are integrated with and taught in conjunction with clinical cases. To break down the traditional division between the pre-clinical and the clinical phase, during the whole curriculum both basic science and clinical practice is programmed. During the years the amount of theoretical training decreases and the amount of clinical practice increases (Ten Cate, 2007; Leinster, 2009; Prideaux, 2009). A vertically integrated curriculum can be represented as an "inverted triangle curriculum" or a Z-shape curriculum (Harden *et al.*, 1984; Ten Cate, 2007; Leinster, 2009). The more traditional curriculum, in which theory is programmed in the first years and clinical education in the final years, can be depicted as an H-shape (Ten Cate, 2007).

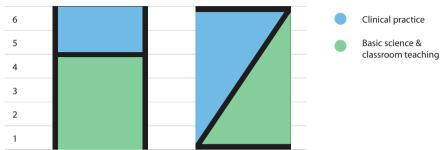


Figure 1 The traditional H-shaped medical curriculum is being replaced by a Z-shaped model

Another important feature of vertical integration is that clinical experience, by means of clerkships or other types of patient contacts, is programmed at an early phase of medical school (Dent *et al.*, 2001; Dahle *et al.*, 2002; Kamalski *et al.*, 2007; Ten Cate, 2007; Leinster, 2009). Several studies have reported the advantages of early clinical experience, such as increased motivation and improvement of clinical skills (Dornan, 2004; Kamalski *et al.*, 2007). Contacts with real patients stimulate the students to practice physical examination skills (Duvivier *et al.*, 2012). Because their experience in clinical contexts, they also better understand the relevance of the theoretical knowledge they have learned (Prince *et al.*, 2000; Dornan *et al.*, 2004; Kamalski *et al.*, 2007). Further, they learn at an early stage what it means to behave professionally and to collaborate in a team (Dornan *et al.*, 2004; Littlewood, 2005).

The common definition of vertical integration (integrating clinical teaching with basic sciences) can be considered too limited. A broader definition is proposed with additional elements. The first is the programming of extended clerkships in the final year of medical school. The second is progressive increase of responsibility for patient care allowed to medical students. At the end of some medical schools, students are given similar responsibility to starting residents (Ten Cate, 2007). This increased responsibility under supervision, motivates students for learning and possibly eases the transition to postgraduate training (Brennan, 2010; Cantillon, 2008).

An illustration of this approach is the General Medical Clinic, an outpatient clinic of the University of Colorado School of Medicine, which was established in 1954. Senior students participated in this clinic with the goal to learn them the principles of comprehensive medical care, which means that the physician is responsible for the patients' total health. During this clerkship, which lasted six months, students were given the maximum responsibility for the patients. Afterwards, students were positive about the degree of responsibility they got for the diagnosis and treatment of patients. In addition, they state that this clerkship was helpful for acquiring more basis facts of clinical medicine, learning to organize clinical work and feeling more like a physician (Hammond *et al.*, 1959).

Central features of vertical integration are the combination of theoretical training with practical training, which are both needed for the development of expertise (Schmidt *et al.*, 1990; Regehr & Norman, 1996; Drevfus, 2004; Aarkrog, 2005; Tynjälä, 2008) and the greater role of workplace learning. Workplace learning gives trainees the opportunity to carry out authentic activities in a real-life context (Billet, 1995; Koens *et al.*, 2005; Chittenden *et al.*, 2009). They can participate in teams by both observing and performing tasks under supervision (Billet, 1995; Dornan et al., 2007; Tynjälä, 2008). Interacting with peers in the workplace fosters learning and sharing of information. The term community of practice (CoP) was introduced by Lave and Wenger (1991) to demonstrate the importance of the participation of individuals in a professional community. Because new members and more experienced members of a community work together, exchange of knowledge is stimulated. Within this context, trainees follow the path from legitimate peripheral participant to core participant of the community (Lave & Wenger, 1991; Li *et al.*, 2009).

Based on the components of vertical integration described in the section above, in this thesis a fully vertically integrated undergraduate medical curriculum is defined by the following five features:

- the delivery of basic science teaching in conjunction with a clinical context;
- 2. the occurrence of real patient contacts from the first year of training;
- 3. clinical clerkships from halfway or earlier during the curriculum;
- 4. a progressive increase of clinical responsibility of students over the course of all clerkships and
- 5. long clinical clerkships with an elective nature and substantial responsibility for patient care in the final year of the curriculum.

Vertical integration can improve the readiness for practice of medical graduates in several ways. As horizontal integration, the connection between theoretical concepts and clinical cases stimulates transfer and the building of relevant knowledge frameworks (Schmidt *et al.*, 1990; Bransford, 1999; Norman, 2009). In addition, the increased role of workplace learning is important. The models of expertise development emphasize the importance of learning by experience (Vygotsky, 1978; Schmidt *et al.*, 1990; Ericsson *et al.*, 1993; Dreyfus, 2004; Dunphy & Williamson, 2004).

Workplace learning makes it possible to work under decreasing supervision by experienced others, to practice clinical skills and to experience multiple patient cases. This combination of learning opportunities increases the possibilities for both low- and high-road transfer (Salomon & Perkins, 1989) and stimulates the trainees to develop themselves to adaptive experts (Bereiter & Scardamalia, 2003; Mylopoulos & Regehr, 2009) and to cope with unfamiliar medical situations in clinical practice.

Longitudinal integrated clerkships

In addition to horizontal and vertical integration, there is increasing attention to integration in the clinical phase by means of so called longitudinal integrated clerkships (LICs). Aim of these LICs is to address the problem of fragmentation of the clinical experience of the students resulting from the traditional system of block rotations, i.e. many short clerkships after each other. In LICs students follow patients, with a wide range of diseases, which stimulates the sense of patient-centeredness. In addition, the students are supervised during the whole period by clinicians from each discipline. In this situation the supervisors have more opportunities to give feedback to the students and to follow their development over time (Ogur *et al.*, 2007; Mazotti *et al.*, 2011; Poncelet et al., 2011). Recent evidence suggests that LICs can contribute positively to the learning process of medical students. Students who followed LICs seem to better understand the course of illness and are more committed to the patients (Hirsh *et al.*, 2012). Furthermore, both students and supervisors appreciate the extensive possibilities for feedback (Mazotti *et al.*, 2011).

Measurement of readiness for practice

Not only the curricula, but also the way of determining the readiness for clinical practice of medical graduates has changed. A few decades ago, graduates were only assessed on knowledge and clinical skills. Later both society and the medical profession did not consider this as enough. To perform the job as a physician well enough, also other qualities are needed, such as communication, collaboration, professionalism and reflection. These qualities are often defined as competencies. In various competency frameworks, for instance CanMEDS (Frank, 2005) and competency framework of the General Medical Council in the UK (General Medical Council, 2009) are the competencies described that medical trainees have to acquire in different stages of their training. Assessment instruments that are used for the observation of trainees in clinical practice, for example the mini-clinical evaluation exercise (mini CEX, Norcini *et al.*, 1995), are often based on competencies.

A recent approach to the development for curricula and the measurement of readiness for clinical practice is based on the concept of entrustability of trainees (Ten Cate, 2005; Ten Cate, 2006; Ten Cate & Scheele, 2007; Sterkenburg *et al.*, 2010; Boyce *et al.*, 2011; Jones *et al.*, 2011). By linking competencies with so called "Entrustable Professional Activities" (EPAs) competency frameworks can be translated to clinical practice, which is helpful for the development of medical school curricula or postgraduate training programmes. An EPA is an observable activity and successful execution of an EPA requires possession of multiple domains of competence at the same time (Ten Cate, 2005; Ten Cate & Scheele, 2007).

EPAs are also useful for the determination of readiness for clinical practice, because the question enclosed is: "Will you trust this trainee with this clinical task?" (Freidson, 1970; Ten Cate, 2005). EPAs can be carried out on consecutive levels, which are defined as the degree of supervision needed. This fits into the context of clinical workplace learning. Much patient care is provided by doctors who are still in training, who are learning while they are at work. Senior doctors have to make continual judgments about trainees' work activities—what can trainees do unsupervised, what activities do they need some help with and what are they not yet able to do?

The focus of the entrustability approach on decreasing supervision and progressive independence can be related to Vygotsky's theory about the Zone of Proximal Development (Vygotsky, 1978; Chaiklin, 2003; Dunphy & Williamson, 2004).

Purpose of the thesis

An important question for curriculum developers is whether the medical school curriculum prepares students well enough for clinical practice and whether vertical integration stimulates transfer from theoretical to practical training and from medical school to postgraduate training. Therefore, the main topic of this thesis concerns the effectiveness of medical curricula. One of the fundamental questions of school effectiveness studies is whether schools make a difference in the achievement of students and whether this difference can be attributed to the school. Various conceptual models concerning effectiveness of educational institutes have been developed to answer this question (Scheerens, 1997). Some of these models consider education from an economic point of view: the relation between input variables like pupil/teacher ratio and student outcome is expressed by an arithmetic function. Other models, so called "educational productivity models" include variables with regard to the quality of instruction.

Marzano (2000) examined three decades of research on school effectiveness and he defined the following factors that correlate with student achievement:

- School-level factors: guaranteed and viable curriculum, challenging goals and effective feedback and community involvement, safe and orderly environment, collegiality and professionalism;
- Teacher-level factors: instructional strategies, classroom management, classroom curriculum design;
- Student-level factors: home environment, learning intelligence and background knowledge, motivation.

This thesis will focus on the effectiveness on school level, in particular the curriculum. The medical school curriculum includes theoretical school-based training, practical workplace-based training and the interaction between these two components (Aarkrog, 2005; Tynjälä, 2008). Therefore, the studies described in this thesis will be about the medical curriculum as a whole.

Many medical schools in Western countries have gradualy developed their curricula to represent a more vertically integrated approach. This development has heightened the need for research into the effects of vertical integration. The main goal of this thesis is to determine whether a vertically integrated curriculum at medical school enhances the transition to postgraduate training. Theories about expertise development and transfer in combination with workplace learning principles, suggest that a positive effect of vertical integration is to be expected (Vygotsky, 1978; Salomon & Perkins, 1989; Schmidt *et al.*, 1990; Lave & Wenger, 1991; Ericsson *et al.*, 1993; Billet, 1995; Bransford, 1999; Dreyfus, 2004; Dunphy & Williamson, 2004; Aarkrog, 2005; Dornan *et al.*, 2007; Tynjälä, 2008; Norman, 2009).

The many curricular changes at medical schools, encouraged several studies on the effectiveness (Schmidt *et al.*, 1996; Kaufman & Mann, 1999; Whitehouse *et al.*, 2002; Goldacre *et al.*, 2003; Lempp *et al.*, 2004; Schmidt *et al.*, 2006; Cave *et al.*, 2009; Distlehorst *et al.*, 2009; Bleakley & Brennan, 2011; Scicluna *et al.*, 2012). The results of these studies indicate that the type of the curriculum can affect the self-confidence and perceived preparedness of medical graduates for work and postgraduate training. However, little information is available about the influence of a fully vertically integrated curriculum.

Related to the main goal, to measure the effectiveness of vertically integrated curricula at medical schools, this thesis addresses three other issues. First, this thesis describes the existing structures of medical education around the world, including the transitions between successive phases. Second, the influence of different types of transitions was studied. Further, we describe an assessment instrument that we developed to measure the readiness for practice of medical graduates, based on the entrustability approach.

Research questions

This thesis describes several studies related to readiness for clinical practice and the transitions in medical education. The following research questions are addressed in this thesis, divided among four topics:

1. Structures of medical education around the world:

- a. What are existing models of medical education, including transitions to a next phase or level? (chapter 2)
- b. At which moment is unrestricted practice allowed? (chapter 2)
- c. Which degrees are granted after medical school? (chapter 2)

2. Influence of different types of transition on perceived competence development and readiness for practice:

- a. How do medical trainees evaluate their readiness to be entrusted with selected clinical tasks at three stages of training? (chapter 3)
- b. How do clinical supervisors evaluate the readiness to be entrusted with selected clinical tasks of medical trainees at three stages of training? (chapter 3)
- c. What is the differential impact of a responsibility transition and a licensure transition on self-perceived competence and supervisor perceptions? (chapter 3)

3. Influence of vertically integrated curricula at medical school on the transition to postgraduate training, by means of comparison with non-vertically integrated curricula:

- a. During which time period was the definite specialty career choice made? (chapters 4 and 5)
- What was the influence of experiences in the final year of medical school on the process of making a career choice? (chapter 4)
- c. How long was the period between graduating medical school and starting residency? (chapters 4 and 5)
- d. How many applications were required to obtain a residency position? (chapters 4 and 5)
- e. How generally satisfied do graduates feel with their careers since leaving medical school? (chapters 4 and 5)
- f. To what extent do graduates feel prepared for work and postgraduate training after finishing medical school? (chapter 5)

- g. What are the differences regarding preparedness for work, medical knowledge and capabilities to manage important specific parts of the work as a physician between starting residents graduated from a medical school with a vertically integrated curriculum and from a medical school with a non-vertically curriculum, according to supervisors in postgraduate training programs? (chapter 6)
- What are the differences between graduates from a medical school with a vertically integrated curriculum and from a medical school with a non-vertically integrated curriculum in their competences to cope with unfamiliar clinical situations? (chapter 10)
- 4. Development of an authentic performance assessment to measure the readiness for clinical practice of medical graduates:
 - a. What do experienced clinical educators consider as essential facets of competence that determine decisions to entrust a trainee with critical clinical tasks? (chapter 7)
 - b. How well do clinical educators in different educational climates agree on these facets? (chapter 8)
 - c. How can validity in the assessment of readiness for practice be attained? (chapter 9)

Methodology

A variety of methods are used to address the research questions. We conducted questionnaire studies among medical graduates and supervisors to gather information about their perceptions regarding the readiness for practice of graduates (chapters 3, 4, 5 and 6). We carried out a qualitative questionnaire study among key informants, supplemented with document analyses to find out about the medical education system in several countries around the world. In addition, we conducted design-based research to develop a procedure for the assessment of readiness for practice. In this process, we used a Delphi method among experts, followed by several implementations of the assessment procedure to collect evidence for validity.

Outline of the thesis

Chapter two describes a questionnaire study and literature review to get an overview of stages and transitions in medical education in 40 different countries, including the moment that unrestricted practice is allowed.

Chapter three focuses on the influence of different types of transition in medical education on the perceived competence development and readiness for clinical practice of trainees. We studied two types of transition, namely the transition to an increased responsibility level and the transition to the status of licensed physician. In some countries these two transitions coincide (for instance in the Netherlands at the end of medical school). In other countries these two transitions are separated by a year or more, like in the United Kingdom, which provides the opportunity to study the difference in impact of these types of transition. To answer the research question, we carried out a questionnaire study in Leeds (UK).

Chapters four and five describe the results of two questionnaire studies among medical graduates. The aim of these studies was to determine whether a vertically integrated curriculum at medical school affect the transition to postgraduate training, based on the perspective of the graduates. We carried out questionnaire studies among graduates who followed either a vertically integrated or a non-vertically integrated curriculum at medical school in the Netherlands. Items in the questionnaire focused on preparedness for work and postgraduate training, the time and number of applications required to be admitted to residency, and the process of making career choices. Chapter four describes a study among graduates of two consecutive cohorts of the Utrecht medical school. Chapter five presents a comparable study among graduates of the same cohort from six different medical schools.

In the sixth chapter we focus on the perspective of supervisors in postgraduate training programmes. A questionnaire study was carried out among supervisors of postgraduate training programs at Utrecht (the Netherlands; vertically integrated curriculum at medical school) and Hamburg (Germany; non-vertically integrated curriculum at medical school). The supervisors were asked about their medical graduates' preparedness for work, knowledge and capabilities to manage some specific parts of the work as a physician. Chapters seven, eight and nine address issues regarding the development of an authentic assessment. The purpose of chapter seven was to uncover general features of trainees that facilitate clinicians' trust of trainees with critical clinical tasks. We conducted a Delphi study of two rounds among 18 experienced physician educators in the Netherlands. The question concerned, what they consider as essential facets of competence that determine decisions to entrust a trainee with critical clinical tasks. In chapter eight, we describe a ranking study among physician educators in the Netherlands and Germany, to answer the question whether clinical educators in different educational climates agree on the facets of competence resulting from the Delphi study. The aim of the study described in chapter nine was to provide a validity argument for the authentic assessment, following the argument-based approach for validation.

In chapter ten we answer the question whether there are differences in readiness for clinical practice between graduates from a vertically integrated curriculum and from a non-vertically integrated curriculum. In particular we wanted to know what the differences are between these two groups in their competences to cope with unfamiliar clinical situations. To answer this question, we designed and implemented the authentic performance assessment, as described in chapter nine.

Chapter eleven concludes this thesis by a summary of the findings and a general discussion. Limitations of the studies are discussed as well as implications for educational practice and suggestions for further research.

This thesis can be read as a whole with subsequent chapters, but is also a collection of related papers. For that matter, repetition and overlap across chapters are inevitable.



Stages and transitions in medical education around the world: clarifying structures and terminology

Chapter based on publication: Wijnen-Meijer, M., Burdick, W., Alofs, L., Burgers, C. & Ten Cate, O. (Accepted for publication). Stages and transitions in medical education around the world: clarifying structures and terminology. *Medical Teacher*.

Abstract

Background

In a world that increasingly serves the international exchange of information on medical training, many students, physicians and educators encounter numerous variations in curricula, degrees, point of licensing and terminology.

Aims

The aim of this study was to shed some light for those trying to compare medical training formats across countries.

Methods

We surveyed a sample of key informants from 40 countries. Survey questions included: structure of medical education, moment that unrestricted practice is allowed, various options after general medical licensing, nomenclature of degrees granted and relevant terminology related to the medical education system. In addition, we searched the literature for description of country-specific information.

Results

Based on the results, we described the six models of current medical training around the world, supplemented with a list of degrees granted after medical school and an explanation of frequently used terminology.

Conclusions

The results of this questionnaire study lead to the conclusion that while there are many differences between countries, there appear to be six dominant models. The models vary in structure and length of medical training, point of full registration and degrees that are granted.

Introduction

To the outside world, no profession seems as well defined worldwide as that of a doctor, and no training as universal as medical training. This is, however, not the case. Medical educators describe an array of pathways and terminology globally leading to potential confusion (Schwarz, 2001; Wojtczak, 2002). In this paper, we attempt to codify the various pathways that lead to completion of doctor training.

There are several reasons why clarification of structures and terminology is useful. One reason is the increasing mobility of medical students and graduates. Many trainees now do part of their medical training abroad (Hallock *et al.*, 2007; Boulet *et al.*, 2006; Harden, 2006; Teichler, 2003). They often discover unexpected differences in the educational structures. For instance, a fourth year medical student in Europe may suddenly appear to be equal to a second year medical student in the USA. Also, for institutions it is important to value students' academic level: how many years of education have they attended? What does the diploma mean regarding their knowledge and experience?

Another reason is the mobility of doctors, which requires similar information about the educational systems (Ineson, 2005). While several countries around the world have agreements regarding mutual recognition of medical diplomas, it is not always clear whether educational programmes and graduation levels are equal (Christensen, 2004; Garcia-Perez *et al.*, 2007; Karle, 2008-1; Schwarz & Wojtczak, 2002).

Third, an issue in debate within Europe is the harmonization of higher education. The purpose of the Bologna Agreement, signed by 46 countries, is to make European higher education comparable and competitive by applying uniform academic degrees and quality assurance standards. One major objective of the Bologna Process is the adoption of a Bachelor and Master system. There is ongoing discussion and uncertainty whether this two-cycle system is suitable for medical schools. "While some countries have adopted this model, others will not do so. Bologna aims to harmonize higher education across Europe but in practice, for medical curriculum models, it leads to divergence rather than convergence." (Patricio *et al.*, 2008; Patricio *et al.*, 2012). Worldwide, the World Federation of Medical Education (WFME) has also launched an initiative to reduce the differences between medical programs. This is called the global standards program and provides standards for quality improvement in basis medical education as well as postgraduate training and continuing professional development (Karle, 2008-2).

Finally, curriculum comparisons frequently bring about a confusion of tongues at international conferences and other international contacts (Wojtczak, 2002). For instance, in presentations about a topic regarding medical education an extensive explanation is required about the structure of medical education in the concerning country, before the actual research can be discussed.

While there are many articles about the medical education system in one country (Amin *et al.*, 2008; Ibrahim, 2007; Kurdak *et al.*, 2008; Nikendei *et al.*, 2009; Pales & Gual, 2008; Prideaux, 2009; Reisi *et al.*, 2009; Reni *et al.*, 2008; Schwarz et al, 2004; Sood, 2008; Suzuki *et al.*, 2008; Ten Cate, 2007) we did not find an article that provides us with a complete overview. Clarification of the structure and terminology of medical education in different countries would therefore be a useful tool for international medical educators and others seeking to compare systems of education. The purpose of this paper is to provide an overview of course structures and terminology occurring in medical education across countries.

Methods

We carried out a qualitative questionnaire study among 44 key informants about the medical education system in their country, supplemented with document analyses. Key informants were selected because they have shown to be experienced medical educators, knowledgeable about the curriculum structure in their country, as derived from publications or conference presentations. Some informants were also involved with the FAIMER faculty development programs. In total 40 countries were involved in this study. These countries represent all continents and are in most cases the countries with the largest amount of medical schools in the continent.

Topics of the questionnaire were:

- Structure of medical education, both undergraduate and postgraduate: different stages (medical school, internship, mandatory clinical service, residency) and length of these stages.
- Point that unrestricted practice is allowed.
- Names of degrees that are granted after medical school.
- Relevant terminology of the medical education system.

After analyzing the answers, results were sent back to the respondents to verify whether we had correctly understood their answers. In some cases we had some additional questions for clarification.

Results

Terminology

There is often confusion about the terminology which is used in different countries to describe components of medical training. Table 1 contains a reference list of some important terms in medical education.

Term	Description
Phases	
Clerkship	Rotation (fixed period of practice) at medical school.
Foundation programme	A two-year, general postgraduate medical training programme in the UK which forms the bridge between medical school and specialist training.
House officer	Period of practice between medical school and full registration in several countries. Also called: medial officer or housemanship.
Internship	A first postgraduate year. In many countries internship is the period of practical training after medical school en before full registration. In many countries the term internship is used to indicate the first year of residency.
Medical school	First phase of medical education. Is a tertiary educational institution, with a curriculum that leads to a medical degree.
Residency	Period of training in a specific medical specialty (postgraduate training).
Rotation/Placement	Fixed period of practice during education.
Social service	Period of mandatory clinical service after medical school, often in rural areas (also called national service).
Persons	
Clerk	Student who is following a clerkship.
(Foundation) House officer	A person who is undertaking the Foundation Programme (UK).
Intern	A trainee during internship.
Medical student / student doctor	Student at medical school.
Physician	A person who is legally qualified to practice medicine.
Resident	A trainee during residency.

 TABLE 1
 Relevant terminology in medical education

Structure of medical education

Figure 1 describes six models of medical education occurring around the world involving the successive phases of medical training. Some models are found in many countries, while others occur in a few or just one country. Entry-level medical education programs are tertiary, or college level, courses undertaken at a medical school. In most countries, students can enter medical school directly after finishing secondary education. In some countries, such as North America and the Philippines, a bachelor's degree in another field is required. In many countries, newly graduated doctors are often required to undertake a period of practice before they are allowed to start with specialty training. This 'in between phase' (internship, foundation programme or housemanship) usually lasts 1–2 years. In several other countries, medical school graduates immediately apply for a residency position whereas in others a period of clinical experience before residency is optional (e.g. in the Netherlands). The length of the postgraduate specialty training (residency) also varies between countries and it depends on the specialty as well.

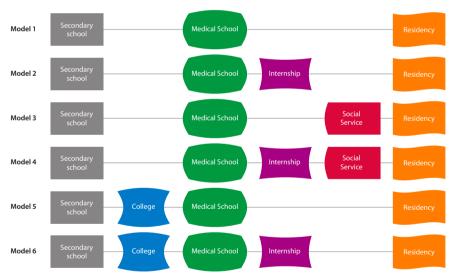


FIGURE 1 GENERAL MODELS OF MEDICAL EDUCATION

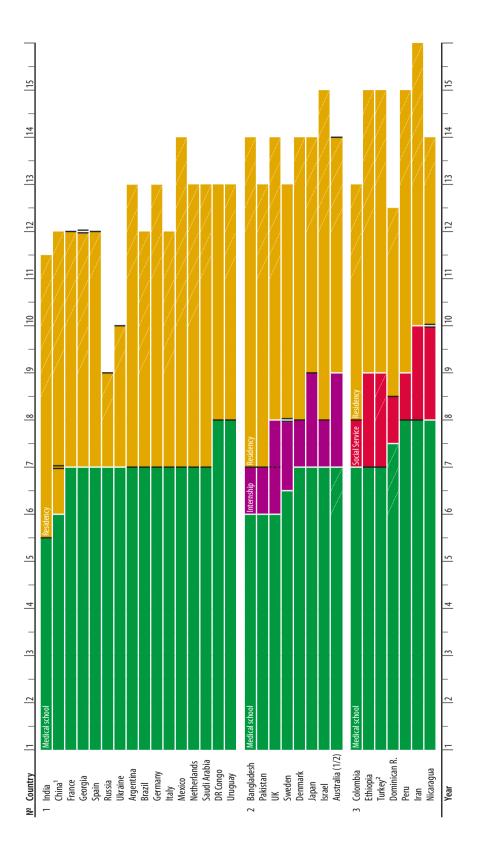
Figure 2 shows the duration of the different stages of medical education per country. In this figure the moment when full registration is granted is also indicated—after this point, full independent practice is allowed. In many countries after the MD degree is bestowed, the moment of full registration as a doctor takes place at a variable interval of time after medical school—after the first year of specialty training, after finishing specialty training, or at a time somewhere between these points.

Names of degrees

As can be seen from Table 2, many different degrees are granted after finishing medical school.

Name of degree	Abbreviations	Country
Bachelor of Medicine, Bachelor of Surgery	MBBS	Australia, Bangladesh, India, Nigeria, Pakistan, Saudi Arabia
Bachelor of Medicine, Bachelor of Surgery	MBBCh, MBBS, MB, ChB	South Africa
Bachelor of Medicine, Bachelor of Surgery	MbChB, MBBS, BMBS, MB	United Kingdom
Bachelor of Medicine, Bachelor of Surgery	MBBCh	Egypt
Bachelor of Medicine, Bachelor of Surgery	MBChB	South Sudan, Sudan
Medical Doctor	MD	Canada, Dominican Republic, Ethiopia, France, Georgia, Germany, Iran, Israel, Japan, Mexico, Netherlands, Phillippines, Peru, Sweden, Turkey, Ukraine, Uruguay, USA
Licenciado en Medicina		Spain
Doctor in Medicine and Surgery (after MS and mandatory social service)		Nicaragua
Doctor in Medicine		DR Congo, Italy, Russia
Medico Cirujano (Surgeon Physician)		Colombia
Dokter	dr	Indonesia
Medico		Argentina, Brazil

 TABLE 2 NAMES OF DEGREE AFTER MEDICAL SCHOOL (BASIC PROGRAM)



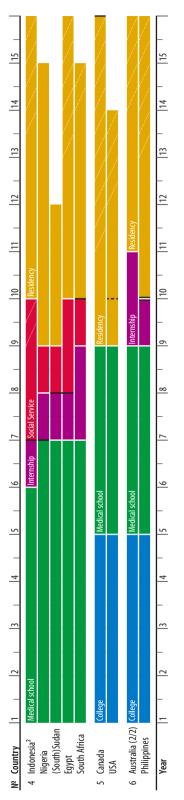


FIGURE 2 MEDICAL EDUCATION SYSTEM PER COUNTRY GROUPED IN SIX GENERAL MODELS (FIGURE 1): AMOUNT OF YEARS AFTER SECONDARY EDUCATION PER PHASE



Transitions

- Irainee is allowed unrestricted practice after finishing this phase and some additional requirements (for example exam) Trainee is allowed unrestricted practice of medicine at this point (other moment than end of phase) End of phase and trainee is allowed unrestricted practice of medicine at this point A transition between two phases
- After 1 year of residency and passing the National Medical Examination residents are allowed unrestricted practice -2
 - In Turkey and Indonesia, mandatory (social) service can take place after residency as well

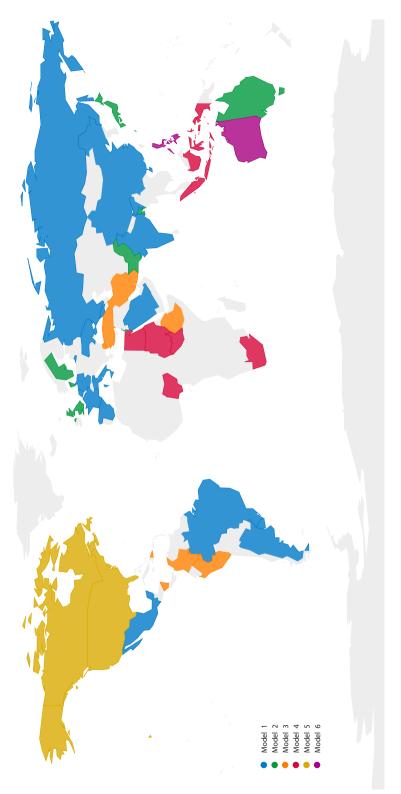


Figure 3 Spread of six general models across the world

Discussion

The purpose of this study was to describe different systems of medical education. The results of this questionnaire study lead to the conclusion that there are many differences between countries. There is variation in structure and length of medical training, moment of full registration and degrees that are granted.

The most important conclusion is that names of stages and degrees are not very informative about the education received and the level of the medical student or graduate. This is even the case for countries that mutually recognize each others' diplomas. E.g., all EU countries must recognize each other's degrees and EU regulations prohibit national authorities to demand any assessment of such international graduates, while for non-EU medical graduates assessments can be made mandatory in EU countries.

Interesting questions are whether it is desirable to make medical education more uniform and whether harmonization of terminology should be recommended to make international mobility of students and doctors easier and to decrease confusion in international contacts. Some initiatives, like the World Federation for Medical Education and the Bologna Agreement, also have the goal of stimulating harmonization and quality improvement in higher and medical education.

This study has some limitations. We received most of our information from representatives of the different countries, who possibly do not have complete or up-to-date information about their medical education system. Another limitation is that not all countries are involved in our study. Nevertheless, we believe that with the response from 40 countries, together with literature information, we have a sufficiently solid foundation to draft the most common models of current medical training around the world, and to list and explain the most frequently used terminology. This can serve as a reference for anyone who is involved in cross-national medical education. More detailed information can be found in dedicated publications as listed in our reference list and on websites of medical schools around the world. Further research might explore the content of the medical curricula.

Finally, this overview just describes the medical education system at this moment, while the situation keeps changing. For that reason it should be interesting to repeat this study for instance after 10 years, preferably with all countries with medical schools included.



The impact of various transitions in the medical education continuum on perceived readiness of trainees to be entrusted with professional tasks

Chapter based on a publication: Wijnen-Meijer, M., Kilminster, S., Van der Schaaf, M. & Ten Cate, O. (2012). The impact of various transitions in the medical education continuum on perceived readiness of trainees to be entrusted with professional tasks. *Medical Teacher, 34* (11), 929–935.

Abstract

Background

Medical trainees go through various transitions during the medical education continuum.

Aims

The aim of this study was to understand how transitions in licensure and increased responsibility may affect trainees' competence development.

Method

We carried out a questionnaire study in Leeds (UK). Trainees and supervisors were asked to determine the competence trainees have at different stages of training to carry out a diversity of medical activities on a 5-point Likert scale. The questionnaires were completed by final year medical students, trainees of foundation year 1 (FY1) and 2 (FY2), and their supervisors.

Results

For all activities listed (1) the trainees' presumed competence increased over time and (2) the mean scores given by trainees were significantly higher than the mean scores given by supervisors across all years the questionnaire covered. As estimated by both trainees and supervisors the impact of a responsibility transition, i.e. from medical school to FY1, is larger than that of a licensure transition, i.e. from FY1 (pre-MD) to FY2 (post-MD).

Conclusions

The transition to formal responsibility with a license to practice medicine seems to have less impact on confidence of trainees and their supervisors to execute critical activities than the transition to informal but significantly higher responsibility.

Introduction

Medical trainees experience many transitions in their learning career. Particularly notable transitions are: start of medical school, transition to clinical training, transition to higher level responsibilities, such as legitimated by a medical diploma or specialist registration, transition from undergraduate to postgraduate training and several transitions related to university and hospital settings. Transitions challenge trainees and are generally recognized as stressful (Cave *et al.*, 2009). A transition to a new level or context of training should carefully balance new challenges with a trainee's competence to cope with these challenges adequately.

Competence development proceeds through stages. Vygotsky's theory about the Zone of Proximal Development (ZPD) indicates that the aim of education should be the development of psychological functions (Vygotsky, 1978; Chaiklin, 2003). To stimulate this development assistance by a more competent other can be helpful, for instance through modeling or scaffolding (Collins *et al.*,1989) or with explicit instruction. Based on this ZPD theory, Dunphy & Williamson described four general stages of learning in relation to the amount of this assistance (Dunphy & Williamson, 2004).

The first is a phase of external regulation. In this phase the learner (e.g. a trainee) is not able to perform a particular task independently and needs assistance from experienced others (e.g. a supervisor) in carrying out this task. The amount of assistance can decrease during this phase as the responsibility the trainee can bear increases. The second stage is that of internal regulation. The trainee is able to carry out the task without assistance, but the performance is not yet fully developed. When the trainee arrives at the third phase, the task execution has been fully internalized and automatized and assistance is no longer needed.

It often happens that performance is no longer automatized and the trainee returns to stage one or two. This de-automatization occurs for instance in cases of major (environmental) changes or stress. Because this situation of de-automatization and recursion occurs so regularly, this is designated as the fourth stage of the normal developmental process. According to Vygotsky's theory, assistance by a supervisor should only be provided to the trainee for those (parts of) tasks a trainee really needs assistance with. Too much help of the supervisor hinders the development of trainee's competences (Vygotsky, 1978; Dunphy & Williamson, 2004).

Supervisor assistance can bridge the distance between the trainee's actual developmental level in a particular task and the level that can be accomplished with guidance. This gap is called his or her ZPD (Vygotsky, 1978), also called "constructive friction" between what a learner can accomplish and what not yet (Vermunt & Verloop, 1999; Ten Cate *et al.*, 2004).

The levels described above can be translated to the medical training context by referring to degrees of clinical supervision needed. One of the characteristics of medical professionals is that they may be trusted to work without supervision (Freidson, 1970). Ten Cate & Scheele (2007) have described five levels of suitable supervision for a trainee, related to the degree that critical activities can be entrusted to this trainee. In Figure 1 we have made a connection between the stages described by Dunphy & Williamson, based on the ZPD theory and the levels of supervision, as described by Ten Cate & Scheele.

N⁰	Level of suitable supervision	Stages of learning based on ZPD theory
1	Not yet able to fulfill a particular task	—
2	May act under full supervision	External regulation of performance
3	May act under moderate supervision	—
4	May act independently	Internal regulation of performance
		Performance is fully developed and automatized
5	May act as supervisor and instructor	_

FIGURE 1 STAGES OF LEARNING COMPARED TO FIVE LEVELS OF SUPERVISION

In clinical settings, supervisors and trainees make daily judgments about trainees' competence to undertake specific tasks and activities. Supervisors must decide to what extent trainees need supervision and whether a trainee is ready to bear increased responsibility (Sterkenburg *et al.*, 2010; Dijksterhuis *et al.*, 2009; Kilminster *et al.*, 2007). Trainees' self-assessment of competence may also be important, because it influences whether they ask for support (Kennedy *et al.*, 2009).

Several studies have demonstrated that trainees often do not feel well prepared for a next level of medical training (Cave *et al.*, 2009; Goldacre *et al.*, 2003) or that they must get used to their new role, with more responsibility (Van Hell *et al.*, 2008). A few studies have found increasing safety risks associated with transitions in medical education (Haller *et al.*, 2009; Jen *et al.*, 2009), and it is not well known how transitions affect trainees' perceived competence for relevant professional tasks.

In addition, the influence of the type of transition, i.e. moving to an increased responsibility level or moving to the status of licensed physician, is not known.

In some countries, the end of undergraduate medical education coincides with licensure for the medical profession (e.g. the Netherlands); in other countries these two transitions are separated by a year or more (e.g. the UK). The separated condition provides the opportunity to study the difference in impact of a responsibility transition versus a transition related to formal licensure.

Aim and research questions

Our aim was to gain understanding about the differential impact on the perceived competence development of responsibility versus licensure transitions. We operationalized this by comparing the transition from an undergraduate medical school setting to the first postgraduate year, in the UK called "foundation year 1" (FY1) and the transition from non-licensed graduate to a licensed graduate with an MD degree, i.e. the transition from the first to the second foundation year (FY2) in the United Kingdom. Finishing medical school results in an MBBS degree but no formal licensure; moving from FY1 to FY2 is based on formal licensure, i.e. full registration at the General Medical Council (GMC).

The research questions are:

- How do trainees evaluate their readiness to be entrusted with selected clinical tasks at three stages of training: final year medical school, FY1 and FY2?
- How do clinical supervisors evaluate the readiness to be entrusted with selected clinical tasks of final year medical students, FY1 trainees and FY2 trainees?
- What is the differential impact of a responsibility transition (medical school to FY1) and a licensure transition (FY1 to FY2) on self-perceived competence and supervisor perceptions?

Methods

We carried out an explorative questionnaire study. We used a questionnaire consisting of 16 so-called "Entrustable Professional Activities (EPAS)" to measure trainees' and supervisors' assessments of trainees' competences. An EPA is a realistic activity and has several characteristics. An EPA is defined as a critical unit of professional work to be entrusted to a trainee once competence at a sufficient level has been reached. Furthermore, an EPA is independently executable and should be observable and measurable in process and outcome. Successful execution of an EPA requires possession of multiple domains of competence at the same time (Ten Cate, 2005; Ten Cate & Scheele, 2007).

The EPAs used in this study describe a diversity of medical activities. They were chosen by consensus discussion of physicians. The questionnaire was used in the past, among in total 663 graduates of six medical schools in the Netherlands. For these data, Cronbach's alpha was 0.69 for the *clinical activities* and 0.82 for the *general activities*. Eleven of the 16 EPAs concern clinical activities; five of them contain more general activities. In this paper we will refer to these EPAs as respectively *clinical activities* and *general activities* (Table 1).

The questionnaire had two versions: one for the trainees and one for the supervisors. Trainees were asked to indicate to what extent they were able to carry out EPAs independently. Supervisors were asked at which level the EPAs should be carried out in the successive years of medical training. The trainees and supervisors assessed the trainees' competence to carry out activities on a 5-point Likert scale, according to the levels of responsibility and required supervision as defined by Sterkenburg *et al.* (2010) and Ten Cate & Scheele (2007).

The rating scale for clinical activities contained five values:

- 1. I am/they are not yet able to do this.
- 2. I am/they are able to do this under direct supervision.
- 3. I am/they are able to do this if supervision is available if I need it.
- 4. I am/they are able to do this independently.
- 5. I am/they are able to supervise others in performing this activity.

The rating scale for general activities contained slightly different values:

- 1. I am/they are not yet able to do this.
- 2. I am/they are able to do this after detailed consultation with my supervisor.
- 3. I am/they are able to do this after brief consultation with my supervisor.
- 4. I am/they are able to do this independently.
- 5. I am/they are able to supervise others in performing this activity.

Procedure

The questionnaire study was carried out among students of the final year of the Medical School of the University of Leeds, first and second year trainees of the Leeds Foundation Training School and their supervisors. All students (n=238) of the final year of medical school (year 2008–2009) were invited to complete the questionnaire by email in electronic format. The questionnaires for the trainees (n=92) of the foundation programme were available in paper at their training days. They also received an electronic format by email. The supervisors (n=107) received paper questionnaires which were posted to them directly.

Data analysis

The questionnaire consisted of two sections; one focused on responsibility regarding clinical activities and one on responsibility in general activities. We conducted a principal component analysis, to establish how many components there were in each part. Additionally, we carried out a reliability analysis to find out whether the items form reliable scales. In order to compare different groups of respondents we used *t*-tests and ANOVAS with post-hoc tests (Norman, 2010).

Ethical approval

Ethical approval was required for this study. Because it involved NHS staff ethical approval was obtained from Leeds Central Research Ethics Committee.

Results

Response

In total 110 trainees completed the questionnaires: 41 final year medical students (response rate 17%), 44 FY1 trainees and 25 FY2 trainees (response rate 75%). Supervisors were asked to fill out a questionnaire for each of the levels at which they supervised trainees. Of them, 45 had returned one or more questionnaires (response rate 42%). In total 86 questionnaires were filled out: 35 about final year medical school, 28 about FY1 and 23 about FY2.

Principal component and reliability analysis

Principal component analysis (Oblimin rotation) on the items concerning *clinical activities* shows a one factor solution that explains a total of 53% of the variance (Eigenvalue 5.9). Principal component analysis on the part *general activities*, also revealed a one factor solution explaining 65% of the variance (Eigenvalue 3.3). Reliability analysis showed high internal consistencies based on a Cronbach's alpha of >0.80 for both scales, for trainees as well as supervisors.

Trainees' competence at different stages of medical training

In Table 1 and Figure 2 overviews of the results are presented. For each stage (medical school, FY1 and FY2) the estimated competences of trainees are presented in two columns: for the self-assessment by trainees and the assessments by the supervisors. In each column the mean score and standard deviation is mentioned per EPA and per scale (clinical and general activities).

Differences in trainee perspective at different stages in their training

Medical school students rated themselves as least able to carry out clinical and general activities and FY2 trainees as most able. The mean scores (Table 1) for medical students were 2.74 and 2.77 for clinical activities and general activities respectively. For FY1 trainees these scores were 3.32 and 3.68 and for FY2 trainees 3.62 and 3.99 respectively. The differences for both scales between medical school and FY1 are statistically significant (based on ANOVA-test, clinical: F(2, 53)=13.54, p=0.00, r=0.46; general: F(2, 75)=27.81, p=0.00, r=0.57). The differences between FY1 and FY2 are not statistically significant.

		2	al year	Final year medical school	chool		Foundati	Foundation Year 1			Foundation Year 2	on Year 2	
		Tra	Trainees	Sup	Supervisors	Trai	Trainees	Supe	Supervisors	Train	Trainees	Super	Supervisors
٩	Entrustable Professional Activities	×	S	×	s	×	S	×	SD	×	s	×	S
	Scale clinical activities	2.74	0.71	1.41	0.36	3.32	0.37	2.28	0.61	3.62	0.34	2.77	0.68
Ĺ.	To 'break bad news'—to explain to a patient there is no longer any curative treatment for them.	3.00	0.99	1.06	0.24	3.47	0.76	2.00	0.81	3.48	0.92	2.36	1.09
2.	To draw up an examination plan for a new patient at the outpatient department.	2.74	1.01	1.89	0.64	3.11	0.96	2.24	0.75	3.05	0.95	2.55	1.00
з.	To explain a prescription of a specific drug when a pharmacist has questions about the choice.	2.79	0.84	1.50	0.66	3.54	09.0	2.74	0.86	3.76	09.0	3.22	0.80
4.	To hold a conversation with a patient and his/her family to explain a mistake.	3.41	0.83	1.33	0.59	3.55	0.75	2.48	1.25	3.88	0.53	2.64	0.95
5.	To remove a small lipoma from the upper leg.	1.38	0.71	1.28	0.46	1.88	0.95	1.33	0.49	2.06	0.90	2.27	1.01
6.	To establish an IV access for a patient with damaged blood vessels, such as can result, for example, from chemotherapy.	2.51	1.27	1.32	0.48	3.94	0.67	2.96	0.96	4.04	0.86	3.40	1.00
7.	To treat a skin laceration of about 4 centimetre at one of the extremities.	2.64	1.11	1.44	0.51	2.57	1.15	2.27	0.88	3.29	1.12	2.65	1.00
∞.	To render emergency assistance to a patient with acute decompensated cardiac failure.	2.92	0.98	1.42	0.58	3.29	0.71	2.65	0.80	3.68	0.85	3.22	0.85
9.	To ask a representative critical questions about the pharmaceutical product.	2.84	1.01	2.09	1.20	2.94	1.28	2.68	1.30	3.05	1.20	3.14	0.96
10.	To give a junior doctor effective feedback about the way he/she does the physical examination.	3.07	1.39	1.06	0.24	3.93	0.45	2.88	1.05	4.13	0.45	3.10	1.04
11.	To summarise succinctly the patient's problems and your questions during ward rounds or team meetings.	3.27	1.03	2.37	0.99	4.00	0.44	3.60	0.76	4.12	0.44	3.88	0.74
	Scale general activities	2.77	0.76	1.70	0.57	3.68	0.52	2.67	0.63	3.99	0.47	3.19	0.64
12.	To give a presentation at the clinical team meeting after a night shift.	2.59	0.96	1.96	0.89	3.71	0.61	3.36	0.90	4.18	0.59	3.68	0.82
13.	To write a referral letter to a colleague.	2.76	0.93	1.42	0.51	4.30	0.51	2.88	0.78	4.24	0.83	3.50	0.74
14.	To discuss a patient in a multidisciplinary meeting with nursing and social work professionals.	2.68	0.97	1.85	0.91	3.49	0.99	2.62	1.02	4.21	0.59	3.30	0.98
15.	To give a presentation to colleagues on a relevant medical topic.	3.67	0.74	2.59	0.98	3.95	0.58	2.96	0.79	4.28	0.46	3.59	0.73
16.	To write an evidence based case-report for colleagues with limited preparation time (three hours).	2.45	1.20	1.57	0.79	2.85	1.00	2.27	0.80	2.91	1.00	2.22	0.94
	(

TABLE 1 OVERALL TABLE OF MEANS AND STANDARD DEVIATIONS (EPAS)

Supervisors' perspectives on trainees' competences at different stages in their training

When comparing the mean scores given by the supervisors for each of the three levels, we found that for both scales (i.e., clinical and general activities) the scores for medical students are lowest and highest for FY2 trainees. The mean scores (Table 1) for medical students were 1.41 and 1.70 for clinical activities and general activities respectively. For FY1 trainees these scores were 2.28 and 2.67 and for FY2 trainees 2.77 and 3.19 respectively. The differences between medical students and FY1 trainees are statistically significant for both clinical and general activities (based on ANOVA-test, clinical: F(2, 28)=18.52, p=0.00, r=0.66; general: F(2, 42)=23.61, p=0.00, r=0.63). The differences between FY1 and FY2 trainees are not statistically significant.

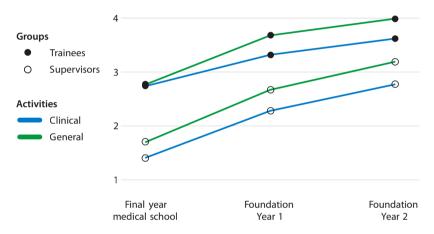


Figure 2 Mean scores of trainees and supervisors of estimated (self) competence (per scale)

Differences between trainee and supervisor perspectives

At all levels (final year medical school, FY1 and FY2) and for both scales (clinical and general activities) there are statistically significant differences between the scores of trainees and the scores of supervisors. In all cases, the scores of trainees are higher than the scores of supervisors (Table 1 and Figure 2), based on 2-tailed *t*-tests, medical school clinical: t(32)=6.28, p=0.00, r=0.74; medical school general: t(43)=4.88, p=0.00, r=0.65; FY2 clinical: t(22)=4.06, p=0.00, r=0.65; FY2 general: t(33)=4.31, p=0.00, r=0.60.

Discussion

The aim of this study was to examine competence development over the transition from medical school to the first foundation year and from the first to second foundation year. As was expected, the mean perceived competence scores for both clinical activities and general activities were lowest for medical school students and highest for FY2 trainees, suggesting that the trainee competence in general is perceived as increasing over time.

We particularly aimed to investigate the impact of the different transitions on the perceived readiness of the trainees to be entrusted with professional tasks. The transition from the final year of medical school to FV1 can be seen as a responsibility transition. After this transition, they are junior doctors instead of students and they are legally allowed to prescribe drugs. The transition from FV1 to FV2 is a licensure transition; after this transition trainees get full registration as a physician with the General Medical Council. In both FV1 and FV2 the trainees work under clinical supervision (Modernising Medical Careers, 2011).

We have seen that from both trainee and supervisor perspectives the differences in competence development between the final year medical school and FY1 are statistically significant, but the differences between FY1 and FY2 are not. The results of this study indicate that the transition to working as a registered doctor, licensed to practice medicine, has apparently less impact than the transition from medical school to foundation programme. It seems that the transition from student to doctor is very important for both supervisors and trainees, despite the fact that the trainees had not attained full registration yet. This may also relate to the fact that after finishing medical school, trainees are part of the community of doctors, which may entail feelings of commitment (Ryan & Deci, 2000). Also, this possibly stimulates supervisors to give more responsibilities to the trainees resulting in an increase of trainees' motivation. This matches the context theory of Koens *et al.* (2005) that suggests that a high level of commitment, together with trainees' responsibility for patient care, stimulates the motivation of the trainees and the willingness to invest effort.

Another notable finding of this study is that for both clinical and general activities and for all levels (final year medical school, FY1 and FY2), the mean scores of the trainees (self-assessment) are significantly higher than the mean scores of the supervisors. In terms of Vygotsky's theory, the trainees and the supervisors disagree about the degree of necessary external regulation, which can be interpreted as the amount of supervision needed (Vygotsky, 1978; Dunphy & Williamson, 2004). This result accords with the findings of Sterkenburg *et al.* (2010).

In their study about decisions of supervisors regarding the tasks they entrust to trainees they found discrepancies between what anesthesiology trainees are expected to do as judged by their supervisors, what they actually do and what they think they are able to do. This may be explained by different frames of reference: trainees may think of common situations or patients, whilst supervisors are more aware of the possible complexities and risks and consider that trainees (particularly the more junior ones) would not be able to deal with these. Trainees who have experienced incidental successes with clinical tasks may draw too positive conclusions about their generalized competence. This could also explain why the differences between the scores of trainees and supervisors are smaller for more experienced trainees, a finding also reported by Sterkenburg *et al.* (2010). This issue deserves further empirical study. Other studies have demonstrated that the abilities of trainees and physicians to self-assess are poor (Davis et al., 2006; Ward *et al.*, 2002; Barnsley *et al.*, 2004; Eva & Regehr, 2011).

The opportunity to carry out activities just beyond one's current competence is necessary for the development of competence. This "constructive friction", in Vygotsky's theory the "zone of proximal development" (Vygotsky, 1978; Chaiklin, 2003; Dunphy & Williamson, 2004), helps to enable trainees to progress in learning (Vermunt & Verloop, 1999; Ten Cate *et al.*, 2004). It may be useful for trainees to be confident about their own competences, as this may motivate them to carry out activities which are challenging for them. Consequently, it is important that both trainees in transition and their supervisors are aware of the fact that these trainees are learning and not necessarily performing perfectly (Kilminster *et al.*, 2010). In the meanwhile, supervisors should guard the patients' safety, as Sterkenburg *et al.* (2010) and Babbott (2010) pointed out. Here is where supervision must serve both educational and patient safety needs.

Our study has some limitations. One is the fact that it is carried out in only one medical school and one foundation training school and that the response rate of medical school students in their final year was low. Due to the voluntary participation it is possible that the sample consisted especially of students and supervisors with a positive attitude towards (self)assessment. Consequently, the sample may not be representative of the population of all medical school students and their supervisors, which implies that the results cannot be generalized to other contexts. Nevertheless, the effect sizes are large and so represent substantive findings. Another possible limitation is the fact that the trainees and students who participated in the study are different groups at different educational stages instead of the same groups at consecutive stages. Next to this, neither we, nor Sterkenburg *et al.* (2010) made a direct comparison between the self-assessment of a trainee and the assessment by the supervisor about particular trainees and their actual competence. Further research must provide such information.

It is important that supervisors can establish how much supervision is needed both to ensure optimal development of trainees and for delivering safe and effective patient care. Trainees should know when to ask for supervision, and therefore must be able to determine their own limits (Eva & Regehr, 2007). Further, it is important that both trainees and supervisors know when trainees are ready to make the transition to the next educational level. Additional investigations about the assessment of actual competence of trainees at different stages of medical training and the implications for the transitions in education could therefore be very useful.

Conclusions

Our aim was to gain further understanding of the competence development over transitions in medical education and to learn about the influence of different types of transitions on this competence development. The results of our study indicate that the perceived trainee competence increases over time, but that this increase is not steady. Both trainees and supervisors seem to consider the impact of a responsibility transition (from medical school to FY1) on the development of trainees' competences higher than that of a licensure transition (from FY1 to FY2).



The influence of a vertically integrated curriculum on the transition to postgraduate training

Chapter based on publication: Wijnen-Meijer, M., Ten Cate, O.Th.J., Rademakers, J.J.D.J.M., Van der Schaaf, M., Borleffs, J.C.C. (2009). The influence of a vertically integrated curriculum on the transition to post-graduate training. *Medical Teacher*, *31* (11), e528–532.

Abstract

Background

Recently, many medical curricula have been changed into vertically integrated programmes. One of the aims of vertical integration is to facilitate the transition from theoretical to clinical education and from medical school to postgraduate training.

Aims

The aim of this study was to determine whether a vertically integrated curriculum affects the transition from medical school to postgraduate training.

Method

We carried out a survey study among graduates of two cohorts of the Utrecht Medical School, who followed either the traditional or the innovative, vertically integrated, curriculum. Topics of the questionnaire were: (1) activities since medical school, (2) required amount of time and number of applications to get admitted to residency, (3) the process of making career choices.

Results

Graduates from the vertically integrated curriculum had made their definite career choice earlier compared to those who followed a traditional programme. Graduates of the new curriculum also needed less time and fewer applications to obtain a residency position.

Conclusions

A vertically integrated curriculum at medical school positively affects the transition to postgraduate training. Additional research, among a larger population, is required to determine which components of the curriculum cause this effect and to specify under which conditions these effect occurs.

Introduction

Reason and relevance

In recent years, many medical curricula have become more vertically integrated. In vertically integrated programmes basic and clinical topics are studied in parallel and in integration with each other (Lie, 1995). Another important characteristic of most of the vertically integrated curricula is early clinical experience, by means of clerkships or other types of patient contact (Dornan & Bundy, 2004; Kamalski *et al.*, 2007; Ten Cate, 2007).

One of the major goals of vertical integration is to stimulate the transition, both from classroom learning to clinical education and from medical school to postgraduate training (Dornan & Bundy, 2004). The transition from medical school to postgraduate programmes is a central issue in The Netherlands. It is often stated that this period needs to be shorter (Bleker & Blijham, 1999). At present, an average of 2–3 years is spent before a medical school graduate starts with postgraduate training (Van der Velden & Hingstman, 2003).

Situational context

In 1999, the Utrecht Medical School replaced its traditional curriculum by an innovative curriculum. The traditional curriculum (H-shaped, Figure 1) is based on a traditional instructional design. In this curriculum, the theory is mainly discipline based and is programmed in the first 4 years, while clinical education is allocated in the 5th and 6th years. The innovative curriculum (Z-shaped, Figure 1) is vertically integrated and has a more contextualized approach of learning (Ten Cate, 2007).

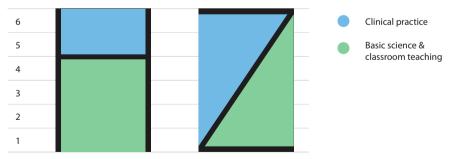


Figure 1 The traditional H-shaped medical curriculum is being replaced by a Z-shaped model

During the first 2 years the basic sciences are studied in combination with clinical cases. In addition, much attention is given to clinical and practical skills, which are required for the early clerkships Internal medicine and Surgery (both 6 weeks) in the 3rd year. The remainder of the 3rd year is spent on thematic blocks and electives. Regular clerkships are programmed in the fourth and fifth years.

An important feature of the innovative curriculum is what has been called a transitional year, which is a rearranged final year of the medical course. In this year, students are asked to choose residency-like internships from a range of possibilities offered. By making adequate choices, they can distinguish themselves from other students and prepare themselves for a specific postgraduate programme. During these internships, students work more independently and are given more responsibility than before, comparable with the responsibility of starting residents.

In The Netherlands two (of eight) universities have introduced this transitional 6th year, both of them in 2004, namely Utrecht University and Leiden University. Other medical faculties have shown interest or intentions to do so in the future. The transitional year can be best compared to the internship in the USA and to what in the UK is called preregistration house officer year or the first year of the foundation programme. A significant difference between these years and the Dutch transitional year is that the latter is an integral part of the 6-year university programme and not put on top of the programme. A crucial aspect of the transitional year is that the medical school has the authority and the obligation to shape this year and control its educational content, whereas in other countries trainees do this extra year after they have finished medical school.

Throughout the Utrecht transitional year students are stimulated to develop general competencies and to monitor their development in a portfolio. A competency is considered as an integration of knowledge, skill and attitude that is necessary to adequately execute a specific professional activity within a given context. The general competencies in the Utrecht transitional year are founded on the CanMEDS framework, which is initially designed in Canada and which is organised around seven roles: medical expert, communicator, collaborator, health advocate, manager, scholar and professional (Frank *et al.*, 1996). The Dutch Central College of Medical Specialists imported this framework in 2004 and adapted it for the Dutch situation. Instead of roles, seven fields of competencies were defined, covering the same CanMEDS domains (Bleker *et al.*, 2004).

Aim and research questions

The aim of this study is to determine whether the innovative, vertically integrated curriculum including the transitional year, affects the

transition from medical school to postgraduate training. Several components of the innovative curriculum can possibly contribute to the development of necessary competencies and therefore to the facilitation of the transition to postgraduate training.

As a result of early clinical experience and larger responsibility, students learn in an early stage of their medical training programme to act as a professional (Dornan & Bundy, 2004). Besides this, as students must choose internships during the final year of undergraduate medical training, they need to think about their specialty career choice and the way to create a distinct profile for themselves (Hannafin, 1984; Kinzie *et al.*, 1988). Because of the use of portfolios, students may be more aware of their own strengths and weaknesses and are stimulated to develop their competencies (Driessen *et al.*, 2003).

The hypothesis in this study is that an innovative curriculum that includes a transitional year facilitates the transition to postgraduate training. Medical graduates from the innovative curriculum feel better prepared for a specific postgraduate programme of their choice and feel more confident about their competencies and career choice. Because of that, they need less time to decide to start postgraduate training and the period before becoming a resident is shortened. We expected that the graduates who followed the innovative curriculum had made the definite specialty career choice at an earlier stage and that they needed less time and less applications to obtain a residency position in comparison with the graduates from the traditional curriculum. Because of that, we also expected to find that they are more satisfied with the course of their careers.

To investigate this hypothesis a case study was carried out among graduates of two cohorts of the Utrecht Medical School, who followed either the traditional or the innovative, vertically integrated programme. Questionnaires were used to gather data.

The following questions were addressed:

- At which moment was the definite specialty career choice made?
- What was the influence of the transitional year on the process of making a career choice?
- What was the length of the period between graduating medical school and starting residency?
- What was the number of applications to obtain a residency position?
- What was the general satisfaction about the career since medical school?

Method

Population

The study population consisted of all medical school graduates from Utrecht University who entered medical school in 1998 (traditional curriculum) and 1999 (innovative curriculum) and graduated before March 2007. A questionnaire was sent in July 2007 to a total of 416 graduates: 205 of them started medical school in 1998 and 211 in 1999. Addresses were acquired from the student administration of Utrecht University. After one month a reminder was sent.

Instrument

The graduates were asked about the internships they had chosen during the final year of medical school, their preferences with regard to the specialty career choices at several moments during medical school and after graduating, activities they had carried out since medical school, whether they were involved in a postgraduate training programme, required amount of time and applications to get admitted to a postgraduate training programme. Finally, they were asked to indicate on a 10-point Likert scale (from 1 = 'very dissatisfied' to 10 = 'very satisfied') to what extent they were satisfied with the course of their careers at the moment of the research. In addition, questions on respondent's characteristics were included: gender, age, ethnic background and domestic circumstances, for instance, whether they had children.

Analysis

For the analysis, the results of the two cohorts were compared with regard to the process of making specialty career choices, the length of the period before the graduates got admittance to postgraduate training, the number of applications to obtain a residency position, and general satisfaction about the course of their careers since medical school. Statistical differences for most of the topics were analysed with Chi-square-tests. The difference between the two groups regarding satisfaction about their careers was analysed with a *t*-test.

Results

Response and characteristics

In total, 81/205 graduates who started in 1998 and 72/211 graduates of cohort 1999 returned the questionnaire (response rate 40% and 34%, respectively). Graduates, who had started with the traditional curriculum and switched over to the innovative curriculum, were excluded, as well as foreign students who were already licensed in their home country but had to follow an additional programme. Consequently, groups of 72 (cohort 1998) and 71 (cohort 1999) subjects were left for statistical analysis.

The difference in gender was almost the same for the two groups: 64% (cohort 1998) and 66% (cohort 1999) were women, respectively. Graduates of cohort 1998 were logically older than the graduates who started in 1999 at the moment of research. When the age is corrected for the cohorts, it turns out that the average age of cohort 1998 was also a little higher when they entered medical school (average 19.2 years versus 18.9). Other characteristics of the two cohorts (ethnic background and domestic circumstances) did not differ from each other.

The process of making choices

The graduates were asked (in retrospect) about the certainty of their career choice on several occasions, namely, before the start of the final (6th) year of medical school and at the time of graduating. These data are presented in Table 1.

-	Coho	rt (%)
-	Cono	1((70)
	1998 (<i>n</i> =72)	1999 (<i>n</i> =71)
Before the start of the final year (Year 6) of medical schoo		
(Very) certain	57	65
Not entirely certain	31	31
(Very) uncertain	12	5
At the the time of graduation		
(Very) certain	74	84
Not entirely certain	23	12
(Very) uncertain	3	4

 TABLE 1 DEGREE OF CERTAINTY ABOUT CAREER CHOICE (IN RETROSPECT)

The table above shows the degree of certainty about career choice before the start of the final year and at the time of graduation. More than half of all respondents were certain about their future career before the start of their final year. Of the respondents who followed the traditional curriculum 57% were certain about their preference for a specialty. Among graduates who followed the innovative curriculum, this percentage was 65%. The percentage of students who were uncertain of their career choice was 12% and 5%, respectively.

As also shown in Table 1, at the time of graduation the differences were slightly more pronounced. Of the 1999 cohort 84% reported that they were certain about their career choice at the time of graduation with only 16% being uncertain or not entirely certain. Among the 1998 graduates 74% were certain and 26% were uncertain or not entirely certain. The differences were not statistically significant.

			Cohort (%)		
	(*) <i>p</i> <0.	05	1998 (<i>n</i> =72)		1999 (<i>n</i> =71)
Before entering medical s	chool		3		6
During medical school	1st or 2nd year 3rd, 4th or 5th year 6th year	1 35 31	67*	3 50 30	83*
After graduation from me	edical school		30*		11*

 TABLE 2 PERIOD IN WHICH STUDENTS MAKE THEIR DEFINITE CAREER CHOICE

Table 2 illustrates in which phase the respondents had made the definite career choice. Based on the structure of the innovative curriculum, the time spent at medical school was divided in three phases, namely years 1-2, years 3-5 and year 6. In the first 2 years the emphasis was laid on the study of basic sciences. In the third, fourth and fifth years, students completed their clerkships. The sixth year was the transitional year.

In both groups a small number of graduates had made their career choice already before entering medical school. A few graduates made their decision during the 1st or 2nd year of medical training. In the group of 1998 graduates, 35% had made their decision during the years 3–5 and 31% during the final year. In the 1999 group these percentages were 50% and 30%, respectively.

Comparing graduates who made their decision during the whole period, more graduates of the innovative curriculum made their decision before graduation than graduates of the traditional programme (83% versus 67%). Thirty per cent of graduates who followed the traditional programme made their decision about career choice after graduating. In the 1999 cohort this was only 11%. This difference was statistically significant (χ^2 =0.008, *p*<0.05).

	Coho	rt (%)
	1998 (<i>n</i> =48)	1999 (<i>n</i> =42)
Less than 6 months after graduating	27	48
Between 6 and 12 months after graduating	31	17
More than 12 months after graduating	42	36

 TABLE 3 (Required) TIME BETWEEN GRADUATING AND ADMITTANCE TO RESIDENCY

		Coho	rt (%)
	(*) <i>p</i> <0.05	1998 (<i>n</i> =48)	1999 (<i>n</i> =42)
0 or 1 applications		71*	93*
2 or 3 applications		19	5
4 or more applications		10	2

 TABLE 4 NUMBER OF APPLICATIONS TO OBTAIN A RESIDENCY POSITION

Admittance to residency

The respondents who already had obtained a residency position at the moment of research (cohort 1998: n=48, cohort 1999: n=42) were asked about the length of the period between graduating and admittance to residency. Table 3 shows that more graduates from cohort 1999 (48%) obtained a residency position within 6 months after graduating in comparison with those from cohort 1998 (27%). Though this difference was not statistically significant (χ^2 =0.09), it suits the expectation that graduates from the innovative curriculum need less time to obtain a residency position.

As shown in Table 4, respondents who followed the traditional curriculum had to apply for a residency position more times than those who followed the innovative curriculum. Of the graduates from cohort 1999, 93% were invited for a residency position or had to apply only once (cohort 1998: 71%). This difference was statistically significant (χ^2 =0.03, *p*<0.05).

General satisfaction

The graduates were asked to indicate on a 10-point scale to what extent they were satisfied with their careers since medical school (from 1 ='very dissatisfied' to 10 = 'very satisfied'). In general, respondents of both groups were satisfied, but the graduates from cohort 1999 were a little more positive (average 8.15 versus 7.82). This difference is not statistically significant (*t*=0.76 based on 2-tailed *t*-test).

Discussion

The purpose of the study was to determine whether a different curriculum at medical school influences the transition to postgraduate training. To investigate this, a questionnaire study was carried out among medical school graduates from Utrecht University. One group of graduates had followed the traditional curriculum and the other group the innovative, vertically integrated curriculum, including a so-called transitional year.

There were several aspects in which graduates of these cohorts differed from each other. Firstly, it appeared that more graduates from the innovative curriculum had made the definite career choice before graduating, despite the fact that they were somewhat younger. Of particular interest is the fact that more graduates from the innovative curriculum had made this decision during the 3rd, 4th or 5th year of medical school (50%), in comparison with the traditional curriculum (35%).

Possibly this difference had to do with the fact that, due to the design of the innovative curriculum with elective internships in the transitional year, students had to think earlier about their future plans. Furthermore, it is notable that the percentage of respondents from the traditional curriculum who were (very) uncertain about their career choice had decreased from 12% to 3% during the 6th year, whereas for cohort 1999, it had remained almost unchanged (successively 5% and 4%).

Secondly, graduates of the new curriculum needed less time to get admitted to residency: Forty-eight percent had been admitted to postgraduate training within 6 months after graduating versus 27% of cohort 1998. This result deviates from the results of a recent Groningen study, which indicates that graduates of a conventional curriculum needed less time to find a residency position than those of a problembased learning programme (Cohen-Schotanus *et al.*, 2008).

The graduates of the innovative curriculum also needed fewer application attempts to obtain a residency position. Ninety-three per cent of graduates who followed the innovative programme were invited for a residency position or had to apply only once. In the 1998 cohort this percentage was 71%, which is significantly lower.

There are a number of possible explanations for this difference. The fact that more graduates from the innovative curriculum had decided about their career before graduating is probably important. Besides this, they had more opportunities to show their competencies, because they were given more responsibilities during their final clerkships.

Finally these graduates seemed somewhat more satisfied with the course of their careers until the moment of the research, but this difference was not significant.

Conclusion

In conclusion, the results of this study support our hypothesis that the vertically integrated curriculum at Utrecht Medical School, including a transitional year, affects the transition to postgraduate training in a positive way. Although not all of the differences between the two groups are statistically significant, they suit our expectations. As our findings are based on data from one medical school in The Netherlands, and the response rate was mediocre, the results cannot automatically be generalized to other populations. Furthermore, the groups compared in this study, had started and finished medical school in different years, with possible different labour market circumstances. Therefore, additional research is needed, preferably with more graduates from different universities who entered medical school at the same time. In particular, further research should attempt to identify which components of the innovative curriculum—like early clinical experience, larger responsibility during internships, internship electives or portfolio—causes the effects found in this case study and elsewhere (Cave *et al.*, 2007). Also further research is needed into the specification of conditions under which these curriculum effects occur.



Vertical integration in medical school: effect on the transition to postgraduate training

Chapter based on publication: Wijnen-Meijer, M., Ten Cate, O.T.J., Van der Schaaf, M., Borleffs, J.C.C. (2010). Vertical integration in medical school: effect on the transition to postgraduate training. *Medical Educa-tion*, 44, 272–279.

Abstract

Objectives

Recently, many medical schools' curricula have been revised so that they represent vertically integrated (VI) curricula. Important changes include: the provision of earlier clinical experience, longer clerkships, and the fostering of increasing levels of responsibility. One of the aims of vertical integration is to facilitate the transition to post-graduate training. The purpose of the present study is to determine whether a VI curriculum at medical school affects the transition to postgraduate training in a positive way.

Methods

We carried out a questionnaire study among graduates of six medical schools in the Netherlands, who had followed either a VI or a non-VI curriculum. Items in the questionnaire focused on preparedness for work and postgraduate training, the time and number of applications required to be admitted to residency, and the process of making career choices.

Results

In comparison with those who have followed non-VI programmes, graduates of VI curricula appear to make definitive career choices earlier, need less time and fewer applications to obtain residency positions and feel more prepared for work and postgraduate training.

Conclusions

The curriculum at medical school affects the transition to postgraduate training. Additional research is required to determine which components of the curriculum cause this effect and to specify under which conditions this effect occurs.

Introduction

During medical training two major transitions take place. The first is the transition from theoretical to clinical education at medical school. The second involves the transition from medical school to postgraduate medical specialist training, when trainees are licensed as MDs and are given much more responsibility. Several studies have demonstrated that many trainees experience difficulties with these transitions: trainees face a sudden change in role and must shoulder more responsibilities (Van Hell *et al.*, 2008); many trainees do not feel well prepared for the next phase of medical training, (Goldacre *et al.*, 2003; Lempp *et al.*, 2004; Cave *et al.*, 2009) and trainees often experience difficulties in the application of knowledge and skills (Prince *et al.*, 2000; Van Hell *et al.*, 2008).

The Dutch medical education continuum has several moments of transition. Students can enter one of the eight medical schools in the Netherlands after finishing a 6-year programme at secondary school. Medical school training also takes 6 years. Full registration is granted after graduation from medical school. This means that trainees are allowed to work as doctors and to enter specialist training programmes. Most medical school graduates apply for postgraduate training. They choose between 33 medical specialties (in which training lasts 4–6 years), general practice training (3 years) and public health training (2–4 years).

In the Netherlands, the transition from medical school to postgraduate training, especially in terms of the length of the intervening period, is an issue of considerable debate. Nowadays, Dutch medical graduates wait an average of 2–3 years before entering postgraduate training (Van der Velden & Hingstman, 2003). Many graduates work as MDs for several years to gain some experience or to find out which specialty they want to choose before they apply for a residency position.

There is wide opinion that the duration of the total training (undergraduate, postgraduate and intervening period combined) needs to be shorter in order to allow specialists to begin to practise earlier and thus to practise for a longer period of their lives. One way to reduce the length of the intervening period is to better gear medical school towards postgraduate training (Bleker & Blijham, 1999).

Recently several Dutch medical curricula have been revised so that they now represent more vertically integrated (VI) programmes. In a VI curriculum, biomedical sciences and clinical cases are integrated and studied in parallel or in connection with one another (Schmidt *et al.*, 1996; Custers & Ten Cate, 2002; Vidic & Weilauf, 2002). Another feature of the VI curriculum is that it provides for early clinical experience by means of clerkships or other types of patient contact (Dornan & Bundy, 2004; Ten Cate, 2007; Kamalski *et al.*, 2007).

Figure 1 represents a schematic depiction of the characteristics of the revised curriculum. In the non-VI (H-shaped) curriculum, theory is mainly discipline-based and is programmed for delivery in the first 4 years, whereas clinical education is delivered in Years 5 and 6. The VI (Z-shaped) curriculum has a more contextualised approach to learning. As Figure 1 shows, in the VI curriculum clinical practice is programmed throughout the whole period of medical school. The proportion of clinical practice-based education increases over the years (Ten Cate, 2007).

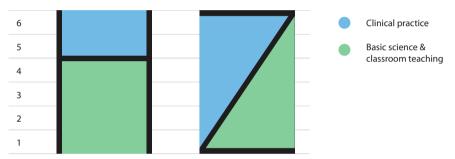


Figure 1 The traditional H-shaped medical curriculum is being replaced by a Z-shaped model

Furthermore, extended clerkships are programmed into the final year of many VI curricula. During these longer clerkships, students are given similar responsibilities to starting residents. To emphasise this widening range of responsibility, several Dutch medical schools use the term *semiphysicians* to describe students in this phase of medical training (Ten Cate, 2007). Thus vertical integration does not only involve differences in the number and distribution of hours of clinical training across the curriculum, but includes a philosophy that supports progressive increases in the responsibility and independence allowed to medical students.

One of the major goals of vertical integration is to ease the transitions from classroom learning to clinical education and from medical school to postgraduate training (Dornan & Bundy, 2004; Ten Cate, 2007). In postgraduate training, doctors in the Netherlands undertake clinical work under supervision while engaging in education. In the remainder of this paper we will refer to this changed role as *work and postgraduate training*.

Research questions

This study aimed to determine whether a VI curriculum at medical school affects the transition to postgraduate training. The results of several recent studies indicate that curriculum changes can contribute towards facilitating this transition; for instance, early clinical experience can increase motivation for learning and support students to act as professionals at an early stage in their medical training (Dornan & Bundy, 2004; Kamalski, *et al.*, 2007).

Goldacre *et al.* (2003) found large differences between medical schools in how graduates perceive their preparedness for a postgraduate career. This implies that the nature of the curriculum can affect the perceived preparedness for work. Other studies in the UK showed that a reformed medical curriculum, with greater levels of 'real-life' clinical practice, makes graduates feel better prepared and therefore smoothes the transition to postgraduate training (Whitehouse *et al.*, 2002; Hayes *et al.*, 2004; Lempp *et al.*, 2004; Watmough *et al.*, 2006; Cave *et al.*, 2007; Cave *et al.*, 2009). A case study carried out among different intern rotations demonstrated that increased responsibility leads to greater motivation for learning (Cantillon & MacDermott, 2008).

In a case study among two consecutive cohorts of graduates of Utrecht Medical School, who had followed a non-VI and a VI curriculum, respectively, we found that graduates of the VI curriculum made their definitive career choices earlier. We also observed that these graduates needed less time and fewer applications to obtain residency positions (Wijnen-Meijer *et al.*, 2009, *chapter 4*). As these findings are based on data from only one medical school and other causes of the findings could not be excluded, additional information is needed to determine the effect of vertical integration. One way to obtain this information is to compare data pertaining to graduates who entered different medical schools at the same time and followed programmes that differed in their degree of vertical integration.

The main hypothesis of the current study was that a VI curriculum facilitates the transition from medical school to postgraduate training. We expected to find that graduates of VI curricula make definitive decisions on specialty career choice at an earlier stage and that they feel better prepared for postgraduate training after finishing medical school. We also expected that, because of their experience as semi-physicians and the greater responsibility this entails, these graduates would need less time and fewer applications to obtain a residency position. Consequently, the period between medical school and residency may be shortened for graduates who have followed VI curricula in comparison with those who have followed traditional curricula at medical school. Presumably, these graduates may also be more satisfied about their careers since they graduated from medical school.

To investigate these hypotheses, we carried out a questionnaire study among graduates of six medical schools in the Netherlands who started medical school in 2000. At that time, three medical schools followed non-VI curricula and the other three had implemented VI programmes.

The following questions were addressed:

- How long was the period between graduating from medical school and starting residency?
- How many applications were required to obtain a residency position?
- > During which time periods was a definitive career choice made?
- To what extent do graduates feel prepared for work and postgraduate training after finishing medical school?
- How generally satisfied do graduates feel with their careers since leaving medical school?

Methods

Participating medical schools

Six medical schools in the Netherlands participated in this study. We gathered information about the curricula by means of interviews and document analysis. Three medical schools (Nijmegen, Maastricht and VU University Amsterdam) ran non-VI curricula in the period 2000–2008. The other three schools (Utrecht, Leiden and Groningen) ran VI curricula. Vertically integrated programmes were defined by the following characteristics: basic and clinical topics were studied in parallel and in integration with one another; students received early clinical experience (\geq 20 weeks of clerkships before the start of Year 5) and undertook longer clerkships in Year 6 (\geq 12 weeks).

Instrument

To determine whether these curriculum characteristics affect the transition from medical school to postgraduate training, we designed and administered a questionnaire for medical graduates. This questionnaire was successfully piloted in a study among graduates in 2007 (Wijnen-Meijer *et al.*, 2009, *chapter 4*). Afterwards, we asked five participants in the pilot study about their experiences with the questionnaire and analysed the distributions of the answers. As a result, we made three adjustments: (1) we adapted the answer options for the question on when participants made a definitive career choice, (2) we asked how experiences in Year 6 of medical school had affected specialty choice and (3) we asked how respondents felt their training in medical school had prepared them for work and postgraduate training.

The topics examined in the final questionnaire included: rotations taken during the final year of medical school; preferences with regard to specialty choices at several stages during medical school and after graduating; the amount of time and number of applications required to gain admittance to residency, and perceived preparedness for work and postgraduate training. We also asked graduates to rate their overall level of satisfaction with the course of their careers on a scale of 1 = 'very dissatisfied' to 10 = 'very satisfied'. Finally, we asked for graduates' views on their training in medical school and the process of making career choices in retrospect.

Participants and procedure

All alumni of the six participating medical schools who entered medical school in 2000 and graduated before April 2008 (n=1217) were asked to participate. This represented graduates of the six medical schools as follows: Utrecht (VI), n=214; Leiden (VI), n=185; Groningen (VI), n=214; Nijmegen (non-VI), n=220; VU University Amsterdam (non-VI), n=148, and Maastricht (non-VI), n=236.

In June 2008 all graduates were sent the questionnaire, accompanied by a letter that explained the purposes of the study. We guaranteed that all data would be treated confidentially and asked the graduates for their cooperation. Four weeks after the initial mailing, a reminder was sent.

Analysis

The results of the two groups of alumni were compared for the purposes of analysis. Statistical differences for most of the topics were analysed with chi-squared tests. The difference between the two groups regarding career satisfaction was analysed with a *t*-test. Finally, we checked relationships between the variables with Pearson's correlations.

Ethical considerations

In accordance with national practice in the Netherlands, ethical approval is not required for educational studies. However, we can confirm that participants cannot be identified from the material presented and that no plausible harm to participating individuals could have been caused by this study. The participants freely decided whether they wanted to respond to the questionnaire. There were no adverse consequences if they decided not to take part.

Results

Response rate

Of the 1217 questionnaires we sent to the graduates of cohort 2000 from the six participating medical schools, 571 were returned, giving a response rate of 47%. In total, 280/613 graduates who had followed VI programmes and 291/604 graduates of medical schools with non-VI curricula returned the questionnaire (giving response rates of 46% and 48%, respectively). Graduates who had followed a customised programme for any reason were excluded. Consequently, responses from 279 (VI curricula) and 285 (non-VI curricula) graduates were left for statistical analysis.

Demographics

Gender distribution was almost identical in both groups: 77% (VI programmes) and 78% (non-VI programmes) of respondents were female. This percentage reflects the gender distribution in cohort 2000. Graduates of VI curricula were slightly younger at the time of the current research than those who had followed non-VI programmes (average 27.3 versus 27.7 years; df=7, $\chi^2=0.000$, p<0.05). Ethnic background and domestic circumstances did not significantly differ between the two groups. The distribution of respondents in terms of ethnic background and age is similar to that across all of cohort 2000.

Admittance to residency

Respondents who already had obtained residency positions at the time of research (n=159 from VI and n=182 from non-VI curricula) were asked about the length of time between graduation from medical school and admittance to a postgraduate training programme. As Table 1 shows, more graduates of VI curricula (50%) obtained residency positions within 6 months of graduating in comparison with graduates of non-VI curricula (33%). This difference is statistically significant (df=2, $\chi^2=0.006$, p<0.05).

		Grou	p (%)
	(*) <i>p</i> <0.05	Non-VI curriculum (<i>n</i> =182)	VI curriculum (<i>n</i> =159)
Less than 6 months after graduating		33*	50*
Between 6 and 12 months after graduating		28	23
More than 12 months after graduating		39*	27*

TABLE 1 TIME BETWEEN GRADUATING AND START OF RESIDENCY

			Grou	p (%)	
	(*) <i>p</i> <0.05		Non-VI curriculum (<i>n</i> =182)		VI curriculum (<i>n</i> =158)
Invited for residency posit	ion		5*		11*
Number of applications	One Two Three or more	75 17 4	95*	71 11 7	89*

 TABLE 2 NUMBER OF APPLICATIONS REQUIRED TO OBTAIN A RESIDENCY POSITION

Table 2 shows that more respondents from VI programmes had been personally invited to fulfil a specific residency position in comparison with those who had followed non-VI curricula (11% versus 5%). This difference is statistically significant (df=1, χ^2 =0.048, p<0.05).

The process of making career choices

Table 3 shows in which phase of medical training the respondents had made their definitive career choice. In both groups a small number of graduates had made definitive choices before the start of Year 4 of medical school.

Of the graduates who had followed VI curricula, 63% had made their decision during Year 4, 5 or 6 of medical school and 25% had done so after graduating. In the group of graduates from non-VI programmes, these percentages were 49% and 35%, respectively. This difference is statistically significant (df=2, χ ²=0.003, p<0.05).

Of the graduates who followed a VI curriculum, 34% had made their career choices during Year 6, in comparison with 24% of graduates of non-VI schools. This difference is also statistically significant (df=7, χ^2 =0.012, p<0.05).

			Grou		
	(*) p<(0.05	Non-VI curriculum (<i>n</i> =280)		VI curriculum (<i>n</i> =272)
Before Year 4 of medical s	chool		16		12
During Years 4, 5 and 6	Year 4 Year 5 Year 6	4 21 24*	49*	9 21 34*	63*
After graduation from me	dical school		35*		25*

TABLE 3 PERIOD IN WHICH STUDENTS MADE A DEFINITIVE CAREER CHOICE

We asked graduates to retrospectively identify the extent to which they had been certain about their specialty choice at various time-points, including before the start of the final year (Year 6) of medical school and at the time of graduating. Data are presented in Table 4. More than half of the respondents had been certain about their career choice before the start of the final year of medical school. In both groups, 11% had been uncertain of their career choice at that moment.

	Grou	p (%)
	Non-VI curriculum (<i>n</i> =266)	VI curriculum (<i>n</i> =264)
Before the start of the final year (Year 6) of medical school	ol	
(Very) certain	52	55
Not entirely certain	38	35
(Very) uncertain	11	11
At the the time of graduation		
(Very) certain	73	79
Not entirely certain	21	18
(Very) uncertain	5	3

 TABLE 4 DEGREE OF CERTAINTY ABOUT CAREER CHOICE BEFORE THE START OF THE FINAL YEAR

 (YEAR 6) AND AT THE TIME OF GRADUATION (IN RETROSPECT)

Table 4 also illustrates that the differences between the two groups were still small at the time of graduation. Overall, 79% and 73% of graduates from VI and non-VI programmes, respectively, were certain about their choice of specialty at graduation. These differences are not statistically significant.

The graduates were asked to what extent their specialty choice had been influenced by their experiences during Year 6 of medical school. Responses to this question are presented in Table 5. Graduates who had followed VI curricula were more likely to observe that their experiences during this final year had exerted strong influence on their career choice, in comparison with those who had followed non-VI curricula (73% versus 59%). This difference is statistically significant ($df=2, \chi^2=0.002, p<0.05$).

		Grou	р (%)
	(*) <i>p</i> <0.05	Non-VI curriculum (<i>n</i> =278)	VI curriculum (<i>n</i> =276)
(Very) Weak influence		17	12
Moderate influence		24*	15*
(Very) Strong influence		59*	73*

TABLE 5 INFLUENCE OF EXPERIENCES DURING YEAR 6 OF MEDICAL SCHOOL ON CAREER CHOICE

		Group (%)		
	(*) <i>p</i> <0.05	Non-VI curriculum (<i>n</i> =284)	VI curriculum (<i>n</i> =275)	
(Very) Bad		3	4	
Moderate		34*	24*	
(Very) Good		63*	72*	

 TABLE 6 PERCEIVED PREPAREDNESS FOR WORK AND POSTGRADUATE TRAINING AFTER MEDICAL

 SCHOOL

Perceived preparedness for work and postgraduate training

Respondents were asked to what extent their training at medical school had prepared them for work as a doctor and for postgraduate training. As **Table 6** shows, graduates who had followed VI programmes were more positive about their preparedness; 72% of these answered that medical school prepared them well or very well, whereas only 63% of graduates from non-VI curricula gave a similar response. This difference is statistically significant ($df=2, \chi^2=0.043, p<0.05$).

General satisfaction

We asked respondents to indicate on a 10-point scale to what extent they were satisfied with their careers since medical school (1 ='very dissatisfied', 10 ='very satisfied'). In general, graduates of both groups were satisfied, but respondents who had followed VI curricula gave slightly more positive answers (average 8.12 versus 8.0). This difference is not statistically significant (t=0.144 based on 2-tailed t-test).

Correlations

We computed Pearson's correlations for the moment of definitive career choice, length of period between graduation from medical school and admittance to residency, number of application attempts, perceived preparedness for work and general satisfaction. For all correlations r<0.30.

Discussion

The main hypothesis of this study was that a VI curriculum facilitates the transition from medical school to postgraduate training. To test that hypothesis, we carried out a questionnaire study among graduates of six medical schools in the Netherlands. One group of graduates had followed a non-VI programme and the other group a VI curriculum. We found several differences between these two groups in support of our hypothesis.

Firstly, it appeared that more graduates from VI curricula had made a definitive career choice before graduating from medical school (75% versus 65%), despite the fact that they were slightly younger. This difference may be attributed to experiences during Year 6 of medical school. The final year of a VI curriculum included longer clerkships (≥12 weeks) designed to foster increasing levels of responsibility.

More graduates of VI curricula mentioned that their experiences during this year had influenced their career choice. Of these, 73% indicated that this influence was strong, versus 59% of the group who graduated from non-VI curricula. Furthermore, because students were required to choose these extended clerkships from a range of options, they were forced to consider their future plans at an earlier phase.

Secondly, more graduates from VI curricula were actively invited to fulfil specific residency positions in comparison with those who had followed non-VI curricula (11% versus 5%). It is possible that this difference reflects the increased number of opportunities students on VI curricula have to show their competencies because they work more independently, are given more responsibility during their final clerkships, and act more as colleagues of the staff.

Thirdly, respondents from medical schools with VI programmes felt more prepared for work and postgraduate training. A large majority (72%) of graduates from VI curricula felt well or very well prepared after medical school. The corresponding figure for the group of non-VI graduates was 63%, which is significantly lower.

Finally, graduates from VI curricula needed less time to obtain residency positions. Of this group, 50% had been admitted to postgraduate training within 6 months of graduating, versus 33% of graduates from non-VI programmes. This difference may reflect the fact that more graduates from VI programmes had made their career choices before graduating and were thus less likely to need more time or more experiences in order to decide which postgraduate training they wanted to apply for. In addition, the facts that they felt more prepared for work and postgraduate training after medical school and that more of them were invited to specific residency positions are probably important.

The comparability of medical curricula in the Netherlands represents a strength of this study. All schools abide by the same blueprint of objectives and students are admitted to medical school largely according to the outcomes of a national lottery system (Ten Cate, 2007). This makes it possible to compare curriculum structures, independently of content and student population, as was recently shown by Schmidt *et al.*, (2009) in their study of the effect of active learning curricula on student progress.

A possible limitation of our study is that we asked graduates about their experiences in retrospect, which may have resulted in the loss of some information, but this possibility is equally true for all respondents. Another limitation is the study's modest response rate. Geographical differences in available positions for residency could be a confounding factor, but as both the VI and the non-VI schools were spread over the country, this effect is unlikely.

Finally, our operationalisation of a VI curriculum was limited to the integrated programming of basic and clinical sciences, early clerkships and longer clinical rotations in Year 6. We could not separately capture the differences in the extent of responsibility and independence of Year 6 students, but assumed that *semi-physicians* in VI curricula would have more of both than students in non-VI curricula. Although some curricula were in the process of transforming at the time of this study, we were able to clearly distinguish curricula, given the variables we used and the cohorts investigated.

In conclusion, the results support our hypothesis that a VI curriculum at medical school affects the transition to postgraduate training in a positive way. The differences we found between the two groups of graduates match our expectations. These findings also concord with the results of a study we carried out in 2007, among two cohorts of graduates from Utrecht University (Wijnen-Meijer *et al.* 2009, *chapter 4*).

The findings raise questions that this study is unable to answer. In particular, we do not know yet which components of the VI curriculum—such as early clinical experience or greater responsibility cause the effects we found. This issue deserves further empirical study, preferably internationally. In addition, supplementary research is needed to specify the conditions under which our findings occur and to establish the extent to which they depend on, for instance, the underlying educational philosophy or educational environment. Additional investigations, in which trainees are questioned during medical training about their experiences and the process of making career choices, could also be useful. Finally, further research into the actual performance of graduates from VI and non-VI curricula, possibly in combination with investigation into the length of time between medical school and residency, could substantiate whether this curriculum structure may affect more than just transition parameters.



Postgraduate supervisors value graduates from a vertically integrated undergraduate curriculum

Chapter based on publication: Wijnen-Meijer, M., Ten Cate, O., Van der Schaaf, M. & Harendza, S. (Accepted for publication). Graduates from vertically integrated curricula. *The Clinical Teacher*.

Abstract

Background

Vertical integration (VI) has been recommended as an undergraduate medical curriculum structure to foster the transition to postgraduate training. Our definition of VI includes (1) the provision of early clinical experience, (2) the integration of biomedical sciences with clinical cases, (3) long clerkships during the final year and (4) increasing levels of clinical responsibility for students. The aim of the current study is to support the hypothesis that medical graduates from VI programs meet expectations of postgraduate supervisors better than those from non-VI curricula.

Methods

A questionnaire study was carried out among supervisors of postgraduate training programs at Utrecht (the Netherlands, VI; n=128) and Hamburg (Germany, non-VI; n=114). The supervisors were asked about their medical graduates' preparedness for work, knowledge and capabilities to manage some specific parts of the work as a physician. They evaluated their performances on a 5-point Likert scale.

Results

The two groups of supervisors did not differ in their judgment of their graduates' preparedness for work and amount of knowledge. However, Utrecht supervisors evaluated their graduates higher with respect to capability to work independently, solving medical problems, managing unfamiliar medical situations, prioritizing their tasks, collaborating with other people, estimating when they need to consult their supervisors and reflecting on their activities.

Discussion

Graduates from vertically integrated medical curricula appeared to be more capable in several facets of a physician's job. Research into the actual performance of graduates from VI and non-VI curricula is needed to further support a firm recommendation for VI curricula.

Introduction

Integration, both horizontally and vertically, has been regarded as a feature of modern medical education, reflected in the oft-cited SPICES model (Harden *et al.*, 1984). Medical schools vary in the extent they have applied vertical integration, which reflects the idea of early attention to traditionally late curriculum elements. Clinical teaching and experience in the first curriculum years, next to attention for the basic sciences is typically subsumed under vertical integration. We believe that also bringing clinical responsibility to earlier curriculum stages, and increasing this responsibility of clerks to a level that traditionally would be asked in postgraduate training, as some schools do, should be considered a feature of vertical integration. We therefore propose an operational definition of vertical integration, including four elements: (1) provision of early clinical experience, (2) integration of biomedical sciences and clinical cases, (3) long clerkships during the final years of training and (4) fostering of increasing levels of clinical responsibility within undergraduate training (Schmidt *et al.*, 1996; Vidic & Weilauf, 2002; Ten Cate, 2007). One aim of vertical integration is to facilitate a smooth transition to postgraduate training.

There is some evidence that vertical integration has the desired transitional effect. In an earlier study we found that a vertically integrated (VI) curriculum at medical school affected the transition to postgraduate training in a positive way. In comparison with graduates from non-VI programs, graduates of VI curricula appear to make their definitive career choices earlier and need less time and fewer applications to obtain a residency position (Wijnen-Meijer *et al.*, 2010, *chapter 5*). It is also known that graduates from VI curricula feel better prepared for work and postgraduate training and that they, in their own opinion, have acquired more of the general competences (Schmidt *et al.*, 2006; Cave *et al.*, 2009; Wijnen-Meijer *et al.*, 2010, *chapter 5*).

This may be interpreted as a sign that applicants from VI curricula are valued by program directors of residency programs, but it can only serve as an indirect measure. A next step to support this hypothesis is to directly investigate the opinions of clinicians about the proficiency of graduates from VI and non-VI curricula. Expert performance, as known from previous research, depends on two major factors: knowledge and experiences in practice (Ericsson *et al.*, 2006). Keijsers *et al.* (2009) reported about one school where students from a VI curriculum had less knowledge of basic sciences compared to students in a previous conventional, non-VI curriculum at the same institution. However little is known about the impact of the increased clinical experience students

usually receive in VI curricula. By hearsay among employers of Utrecht graduates, we know that supervisors of postgraduate training programs highly appreciate the proactive clinical functioning and general competences of current VI graduates. To substantiate this impression a more systematic data collection was deemed necessary.

Research question

Our research question was: do graduates of a vertically integrated curriculum differ from those of a non-vertically integrated curriculum regarding preparedness for work, medical knowledge and capabilities to manage important specific parts of the work as a physician, according to supervisors in postgraduate training programs?

The study was conducted among medical supervisors in Utrecht (the Netherlands) and Hamburg (Germany). The medical school in Utrecht has a VI curriculum, while the program in Hamburg was not vertically integrated during the execution of this study, based on the four criteria listed above (Ten Cate, 2007; Nikendei *et al.*, 2009).

Methods

We designed and administered a questionnaire to supervisors of postgraduate training programs in the Utrecht and Hamburg regions. To evaluate supervisors' judgments of the performances of starting residents, we used a questionnaire with 13 items on different facets of the physician's work (Table 1). Supervisors were asked to what extent medical graduates were prepared for clinical work, had acquired enough knowledge and were capable to manage some important parts of the work as a physician. They evaluated this performance on a Likert scale items with values from 1 = 'very bad' to 5 = 'very good'.

Participants and procedure

In August 2010, all supervisors of postgraduate training programs in the regions of Utrecht (n=128) and Hamburg (n=114) were sent the questionnaire electronically, accompanied by information about the study and an invitation to participate. Two weeks after the initial mailing a reminder was sent.

Analysis

The judgments of the two groups were compared with 2-sided independent *t*-tests. We used parametric tests for the analyses of Likert scale data. Although there is debate about the question whether these can be viewed as interval scales, we follow Norman who indicates that parametric statistics are robust for these types of data (Norman, 2010).

Ethical considerations

In accordance with national practice in the Netherlands and Germany, ethical approval is not required for educational studies. However, we can confirm that participants cannot be identified from the material presented and that no plausible harm to participating individuals could be caused by this study. The participants could freely decide whether they wanted to respond to the questionnaire. There were no adverse consequences if they decided not to take part.

Results

In total, 34/128 supervisors of the region of Utrecht and 40/114 supervisors of the region of Hamburg completed the questionnaire (response rates 27% and 35%).

		Re	gion	
(*) <i>p</i> <0.05, (**) <i>p</i> <0	.01 U	trecht	Ham	burg
№ To what extend	М	SD	М	SD
Practice	3.47	0.76	3.67	0.72
1 are graduates prepared for the work at your department?	3.47	0.76	3.67	0.72
Knowledge	3.19	0.64	3.28	0.54
2 have graduates enough biomedical knowledge?	3.12	0.70	2.97	0.80
3 have graduates enough clinical knowledge?	3.36	0.65	3.31	0.75
4 have graduates enough pathophysiological knowledge of diseases?	3.09	0.88	3.11	0.88
Capabilities	3.65	0.49	3.17	0.46
5 are graduates capable to work independently? **	3.64	0.78	3.00	0.90
6 are graduates capable to solve medical problems? *	3.44	0.71	3.05	0.86
7 are graduates capable to manage unfamiliar medical situations? **	3.35	0.54	2.64	0.71
8 are graduates capable to prioritize their tasks? **	3.41	0.61	2.53	0.77
9 are graduates capable to collaborate with other people? **	4.15	0.62	3.64	0.80
10 are graduates capable to estimate when they have to consult their supervisors? *	÷ 3.97	0.59	3.50	0.95
11 are graduates capable to reflect on their activities? *	3.72	0.68	3.25	0.87
12 are graduates capable to behave professionally with regard to patients?	3.76	0.61	3.45	1.06
13 are graduates capable to manage stressful situations?	3.38	0.65	3.25	0.81

 TABLE 1 OVERALL TABLE OF MEANS, STANDARD DEVIATIONS AND DIFFERENCES (T-TESTS)

Demographics

The distribution in gender was nearly the same for the two groups; 82% (Utrecht) and 85% (Hamburg) were male respectively. These percentages reflect the gender distribution in the total group of supervisors in these regions. The responders of both groups represented a wide range of 18–19 medical disciplines in the regions of Utrecht and Hamburg, respectively.

Preparedness for work

The respondents were asked to what extent medical graduates from Utrecht or Hamburg universities were prepared for the clinical work at their department. The mean scores (Table 1) were 3.47 and 3.67, respectively. This difference was not statistically significant.

Knowledge

We asked the supervisors whether the starting residents had enough biomedical, clinical and pathophysiological knowledge. The mean scores (Table 1) were 3.12, 3.36 and 3.09 for Utrecht graduates and 2.97, 3.31 and 3.11 for Hamburg graduates. The differences between regions were not statistically significant. The three items form a reliable scale (Cronbach's alpha is 0.84), but also the mean scale scores (3.19 for Utrecht and 3.28 for Hamburg) showed no statistically significant difference.

Capability

Supervisors were asked to what extent they judged these medical graduates to be capable of managing some important aspects of a physician's job. For seven of these nine questions the differences between the mean scores (Table 1) from Utrecht and Hamburg were statistically significant, based on 2-tailed independent *t*-tests: 'to work independently': t(69)=3.15, p=0.00, r=0.35; 'to solve medical problems': t(71)=2.10, p=0.04, r=0.24; 'to manage unfamiliar medical situations': t(71)=4.77, p=0.00, r=0.59; 'to prioritize tasks': t(68)=5.29, p=0.00, r=0.34; 'to estimate when to consult supervisors': t(69)=2.46, p=0.02, r=0.29; 'to reflect on activities': t(66)=2.44, p=0.02, r=0.29.

In all cases, the mean scores for Utrecht graduates were higher. The differences between the mean scores (Table 1) of the items 'to behave professionally' and 'to manage stressful situations' were not statistically significant. The difference in mean scores between the two groups for the scale *capability* (Cronbach's alpha is 0.92) was statistically significant, based on a 2-tailed *t*-test: t(47)=3.30, p=0.00, r=0.45.

Discussion

The aim of this study was to explore differences in perceived performance between starting residents from a vertically integrated medical curriculum and those from a non-vertically integrated curriculum, according to supervisors of postgraduate training programs. Based on the results of a questionnaire study among supervisors in the region of Utrecht (VI) and Hamburg (non-VI) we found no differences regarding their judged preparedness for clinical work and the estimated acquired amount of knowledge.

However, the VI graduates were judged to be more capable to work independently, to solve medical problems, to manage unfamiliar medical situations, to prioritize their tasks, to collaborate with other people, to estimate when they have to consult their supervisors and to reflect on their activities than the non-VI students. These results correspond with the results of earlier studies among medical graduates from different curricula (Schmidt *et al.*, 2006; Cave *et al.*, 2009) and can be explained by the fact that these graduates had more clinical experience, together with more responsibility at the end of medical school.

Our study has some limitations. One limitation is that we only evaluated graduates from two schools. A second one is the fact that the two groups of responding supervisors are from different countries with possibly cultural differences affecting their judgments. Also the fact that the supervisors did not score both groups, but only the medical graduates of their own medical school, is a restriction. Another limitation is the modest response rate for both groups. This is a clear drawback. However, because our focus was not on representative sampling, but on comparison of groups, and because we found roughly equivalent response rates, we believe these are acceptable.

Furthermore, the supervisors in this study evaluated the medical graduates as a group instead of as individual trainees. Though the results of this study should be interpreted with care, this study is an important addition to the earlier studies among medical graduates from different curricula.

Further research should include actual performance of graduates from VI and non-VI curricula. This is necessary to justify a firm recommendation for vertically integrated undergraduate curricula.



Essential facets of competence that enable trust in graduates: a Delphi study among physician educators in the Netherlands

Chapter based on publication: Wijnen-Meijer, M., Van der Schaaf, M., Nillesen, K., Harendza, S. & Ten Cate, O. (Accepted for publication). Essential facets of competence that enable trust in graduates: A Delphi study among physician educators in the Netherlands. *Journal of Graduate Medical Education*.

Abstract

Background

There is a need for valid methods to assess the readiness for clinical practice of recently graduated physicians. To develop these methods, it is relevant to know which general features of trainees facilitate supervisors' trust in trainees to perform critical clinical tasks. The aim of this study was to discover such essential *facets of competence* (FOCs), based on the opinion of experienced physician educators.

Methods

We conducted a Delphi study of two rounds among 18 experienced physician educators in the Netherlands. Mean, standard deviation, level of agreement and skewness were calculated for the importance of FOCs to make entrustment decisions. The study yielded a list of 25 FOCs.

Results

In the first round, means were between 6.50 and 7.00 on a 7-point Likert scale (SD 0.42–2.18); in the second round, means ranged from 5.45 to 6.90 (SD 0.3–2.02). The level of agreement was high for 91.7% of the FOCs in the first round and 100% of the FOCs in the second round.

Conclusions

We found consensus about important FOCs for entrustment decisions among experts.

Introduction

Obtaining a correct impression of medical trainees' readiness for clinical practice is important for medical educators. This process requires a valid method for the assessment of clinical competence, which has occupied the minds of medical educators for decades. Despite multiple reviews and authoritative proposals (Van der Vleuten, 1996; Wass *et al.*, 2001; Epstein & Hundert, 2002; Turnbull & Van Barneveld, 2002; Epstein, 2007; Holmboe & Hawkins, 2008), in practice clinical educators, evaluators and examiners still have difficulty assessing medical trainees (Govaerts *et al.*, 2007).

Instruments to assess separate domains of competence in the workplace are scarce (Lurie & Mooney, 2010) and instruments that focus on specific tasks usually rely on simulated conditions to ensure standardization. For instance, the mini clinical evaluation exercise (Norcini *et al.*, 1995) and other direct-observation instruments (Kogan *et al.*, 2009) have turned out to be valuable feedback instruments but appear to have limited or not well-documented psychometric properties (Kogan *et al.*, 2009; Norcini & Burch, 2007).

One problem may be a lack of validity of those assessment approaches because they do not directly consider the important question "whether a medical trainee is ready for independent practice", but rather focus on the assessment of particular skills. Traditional checklists established for these purposes may not capture the essence of such entrustment decisions as they insufficiently align with this essential construct (Crossley *et al.*, 2011). In clinical settings supervisors daily decide whether they trust a medical trainee with a specific task and to what extent supervision is needed (Ten Cate, 2006; Kilminster *et al.*, 2007; Dijksterhuis *et al.*, 2009; Sterkenburg *et al.*, 2010). Specific tasks pertain to a specific procedure or skill that is to be carried out; general features pertain to task-independent characteristics. These general features can be called *facets of competence* (FOCS).

Aim and research question

The aim of this study was to uncover general features of trainees that facilitate clinicians' trust of trainees with critical clinical tasks. The research question was: what do experienced clinical educators consider as essential facets of competence that determine decisions to entrust a trainee with critical clinical tasks? Our goal was not to answer this question for specific tasks, but for critical clinical tasks in general. The answer to this question is relevant when developing an instrument to evaluate medical graduates' readiness for practice.

Methods

Design

A Delphi technique was used to investigate consensus and amount of agreement among experts. The Delphi technique is a widely accepted method in identifying desired features of professionals by eliciting expert opinions in successive rounds (Linstone & Turoff, 1977; Dunn *et al.*, 1985; Clayton, 1997). An advantage of the Delphi process is that face-to-face meetings are not required so that there is no risk of peer pressure, and experts from different regions can easily participate in the study.

The Delphi technique in our study comprised the following procedure: experts were interviewed with electronic questionnaires; answers were collected, aggregated and refined throughout two rounds. Though panel members did not know the individual answers of other participants, after each round general feedback was provided to each panel member by summarizing all judgments of the previous round.

At the start of our study, participants received instructions. Next, they were asked to judge preliminary descriptions of general facets of competence of medical trainees and subsequently to judge successive revisions in two further rounds.

Participants

Experienced physician-educators, acquainted with competence levels of starting residents in the Netherlands, were selected as experts. We approached all 24 experienced clinicians in The Netherlands who met the following criteria: (1) holds an academic chair in medical education, (2) works in clinical practice and (3) supervises residents.

Instrument development

We developed a draft questionnaire with a preliminary list of FOCs drawn from a literature review of selected competency frameworks and relevant publications. First, three competency frameworks were analysed, i.e., the CanMEDS framework (Frank, 2005), the Dutch 'Blueprint of objectives for medical schools' (Dutch Federation of University Medical Centers, 2009) and the competency framework of the General Medical Council in the UK (General Medical Council, 2009). A qualitative analysis of the FOCs in the different frameworks resulted in an initial list with relevant FOCs. The first draft with relevant FOCs was completed through a literature review regarding relevant FOCs for clinicians. This search yielded three relevant articles (Kearney, 2005; Ginsburg *et al.*, 2010; Sterkenburg *et al.*, 2010), two of which had a holistic approach to evaluation of medical graduates' competences (Ginsburg *et al.*, 2010; Sterkenburg *et al.*, 2010). Ginsburg *et al.* (2010) describe those characteristics of trainees that influence assessments by supervisors. Sterkenburg *et al.* (2010) focus on factors that guide decisions of supervisors to trust residents with critical clinical tasks. Kearney (2005) describes the results of a Delphi study regarding the features of professionalism.

We compared these six sources and found many content similarities (see Table 1). There were also differences, mainly in the way the descriptions were ordered and labeled, and in their level of detail. Based on the sources, we constructed a list of 24 FOCs that met the following criteria: content correspondence with the original source, observable and assessable level of detail, and applicability to medical graduates.

Each of the resulting FOCs was scored on two significant aspects for valid competence descriptions: (1) Clarity and completeness of formulation, (2) Importance for the entrustment of critical clinical tasks to starting residents. For each FOC a 7-point Likert scale (Dawes, 2007) was developed that pertained to the following statements:

- The description of this FOC is complete and clear.
- This FOC is important for the entrustment of critical clinical tasks to beginning residents.

Only the two endpoints of the scale were labelled: 1 (strongly disagree) and 7 (strongly agree). Respondents were asked to substantiate their answers with comments and to propose improvements for the FOC descriptions. At the end of the questionnaire, respondents were asked to add any new FOCs that they felt were missing.

Data gathering

Our Delphi study had two rounds. For the first round, the list of 24 FOCs was electronically provided to the panelists. Based on scores regarding completeness and clearness of the descriptions and the experts' suggestions for improvement of the descriptions, all FOC descriptions were slightly modified. Further, two FOCs were divided into two separate FOCs and two other FOCs were combined. In a second round, new list consisting of 25 FOCs, together with a summary of the judgments of the experts in the first round, was sent to the panelists.

The panelists were asked the same questions as in the first round. For three FOCs, multiple choice questions were added to solicit opinions

concerning contradictory or unclear aspects that were made during the first round. Round 2 started two weeks after the closing of round 1. In both rounds, the respondents had two weeks to complete the questionnaire.

Data analysis

After each round, a key issue was the decision whether particular FOCs should be accepted, revised or deleted. This decision was based on expert ratings on the completeness and clearness of formulation and importance for entrustment decisions. We used experts' mean score to calculate their endorsement of the proposed FOCs.

Next we established the degree of consensus among the experts' judgments using standard deviations and levels of agreement. For the calculation of the levels of agreement we used a method described by De Loe (De Loe, 2005; Dekker-Groen *et al.*, 2011; Van der Schaaf & Stokking, 2011) for a 4-point scale, adapted to a 7-point scale. De Loe designates consensus as 'medium' if 70% of the scores are given in two (of four) contiguous scale levels. Because of our 7-point scale, we adapted this criterion to three (of seven) contiguous scale levels, which is slightly more stringent (Table 2). However, consensus can only be assumed if judgments tend to be unidirectional (Sackman, 1975). Therefore, the skewness of the distributions of the ratings was computed to check on symmetry and to check whether the experts' judgments tended to be in one direction.

Our purpose was to revise the FOCs based on experts' judgments and comments until the list satisfied the majority of experts. Therefore we made a list of all written comments from the experts that we used to modify the descriptions of the FOCs. To test the differences between ratings in the two rounds on the completeness and clearness of the formulation, we used 2-tailed depended *t*-tests (Norman, 2010).

Ethical considerations

In accordance with national practice in the Netherlands, ethical approval is not required for education studies. We made sure that participants could not be identified from any materials presented to other participants, that participation would cause no plausible harm and would be fully voluntarily, and that refusal or withdrawal would not incur any adverse consequences.

N⁰	Facet of competence	а	b	с	d	e	f
1	<i>Teamwork and collegiality.</i> The physician cooperates effectively and respectful in a (multidisciplinary) team, taking the views, knowledge and expertise of others into account.	~	~	~	~		~
2	Knowing and maintaining own personal bounds and possibilities. The physician knows the boundaries of his own ability and asks for help (timely) when needed. He reflects on himself and the situation.	~	~	~	~	~	
3	<i>Role differentiation.</i> The physician is aware of the multiple personal and professional roles he fulfils and behaves accordingly these roles.	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
4	Active professional development. The physician aims for quality and professional development by means of a critical attitude towards himself and his environment, study, self-assessment, reflection, asking for feedback and setting and achieving learning goals. He reacts to criticism constructively and is aware of his own responsibility regarding his own abilities.	~	~	~	~		~
5	<i>Coping with mistakes.</i> The physician is aware of the fact that anyone can make and makes mistakes once in a while. He is approachable when someone points out his mistakes and reacts adequately when he thinks that a colleague makes a mistake.	~	~	~	~		~
6	<i>Structure, work planning and priorities.</i> The physician sees the overall picture, has organisational skills, a flexible attitude and sets priorities in his work.	~	~	~	~	~	
7	<i>Responsibility</i> . The physician takes responsibility and shows account- ability for his work. He accepts liability for his work.	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
8	Verbal communication with colleagues and supervisors. The physician gives structured, pithy and unambiguous verbal reports on his findings on a patient and his diagnostic and therapeutic policy. He asks relevant and purposeful questions.	~	~	~		~	~
9	Written (and digital) account/report to colleagues and supervisors. The physician keeps files in a systematic and accurate way and delivers written reports to colleagues and supervisors that are adequate in substance, pithy and that state his thinking in a clear and explicit way.	~	~	~			~
10	<i>Empathy and openness.</i> The physician shows empathy, openness and susceptibility/accessibility in his contact with patients.	\checkmark	\checkmark	\checkmark			\checkmark
11	Adapted informing of patients. The physician informs patients and their relatives in a personally adapted way. He takes account of (among other things) the age, mental development, education, language, psychological wellbeing of the patient and his relatives.	~	~	~	~		
12	Respecting privacy and autonomy of the patient. The physician shows respect for the privacy and autonomy of patients and their relatives. He uses his expertise to lead patients in the right direction and takes their choices into account. He follows rules and aggreements regarding doctor-patient-interactions.	~	~	~			\checkmark
13	Scientific and empirical grounded method of working. The physician uses evidence-based procedures whenever possible and relies on scientific knowledge. He searches actively and purposefully for evidence and consults high quality resources. He uses his scientific knowledge criti- cally and carefully in his work.	~	~	~	~		
14	<i>Ethical awareness.</i> The physician is acquainted with ethical aspects of his work. He distinguishes different points of view in the moral debate and makes deliberate choices when his work confronts him with ethical issues.	~	~	√			~

N⁰	Facet of competence	а	b	с	d	e	f
15	Active listening to patients. The physician listens actively to patients and reacts (verbally and nonverbally) on the things he hears in a way that encourages the sharing of information (by the patients) and confirms his involvement with the patient. He shows attention to non-verbal signals coming from the patients.	~	~	\checkmark			
16	<i>Advising patients.</i> The physician advises patient and his relatives about possible diagnostic and therapeutic strategies and possible long- and short-term outcomes.	~	\checkmark	~			
17	Handling emotions of patients and their relatives. The physician handles emotional patients and relatives in a proper way. If necessary, he gives attention to fear, sadness, confusion, and anger and asks for help if needed.	~	~	~			
18	Attention to individual patient background. The physician is aware of a possible influence of the patient background and characteristics on the origin and development of health problems and on the doctor-patient-relationship. He takes this into consideration when developing his diagnostic and therapeutic policy.	~	√	~			
19	Attention to psychosocial aspects of health problems. The physician gives attention to the illness experience and expectations of the patients and relatives. He is aware of possible psychosocial determinants and effects of health problems.	~	~	~			
20	Attention to relatives and caregivers. Whenever needed the physician involves relatives and caregivers in examination and decision-making. During the whole healthcare-contact they will be sufficiently informed.	~	~	√			
21	Active health promotion. The physician looks beyond treatment of cur- rent health problems whenever this is relevant and, if this is applicable for his function, actively promotes health by means of prevention and education.	~	~	~			
22	<i>Continuity in the care process.</i> He physician ascertains continuity in the care process by adequate follow-up treatments, aftercare and purposeful referrals.	~	~	~			
23	<i>Safety and risk management</i> . The physician is alert and critical. He recognizes risks and responds to them timely. He aims at safety by the use of protocols where possible or the deliberate deviation of these protocols for the benefits of the patient. He reports irresponsible behaviour.	~	~	~			
24	<i>Financial and social awareness.</i> The physician is aware of the Financial aspects of his job and is careful in the use of (limited) resources. He is aware of the costs and benefits healthcare holds for society.	~	~	√			
25	<i>Coping with uncertainty.</i> The physician accepts that uncertainty belongs to his profession. Notwithstanding this uncertainty he works decisive.			\checkmark			\checkmark

TABLE 1 FINAL LIST OF FACETS OF COMPETENCE (FOCS), INCLUDING REFERENCES

References

- a CanMEDS, 2005
- b The new doctor. GMC, 2009
- c Blueprint of objectives for medical schools. Dutch Fed. of University Medical Centers, 2009
- d Ginsburg et al., 2010
- e Sterkenburg et al., 2010
- f Kearney, 2005.

Results

Respondents

Eighteen physician educators from 8 different hospitals and 11 disciplines participated in this study. In the first round, 14 participated in the Delphi study (58%), 13 of which were male. The average age was 59 years (range 52–66 years). In the second round, 15 experts responded (63%). Of them, 13 were male. The average age was 59 years (49–66).

The distribution in gender and age in the sample reflects the distribution in the population of physician educators in the Netherlands. The responders in both rounds represented a wide range of surgical and non-surgical disciplines (anaesthesiology, cardiology, general practice, geriatrics, gynaecology, internal medicine, neurology, oncology, paediatrics, rheumatology and surgery). Eleven experts participated in both rounds, 3 only in the first round and 4 only in the second round.

	agreemen	for the level of t for a 4-point ding to De Loe	Adapted calculation fo agreement for a 7-p		
Agreement	1 category	2 contiguous categories	1 category or 2 contiguous categories	3 contiguous categories	
High	70%	80%	70%	80%	
Medium	60%	70%	60%	70%	
Low	50%	60%	50%	60%	
None		f ratings in ous categories	<60% of ratings in 3 contiguous categories		

 TABLE 2 CALCULATION LEVELS OF AGREEMENT ACCORDING TO THE METHOD OF DE LOE (1995)

Results

Table 3 shows the expert ratings on 'comprehensiveness and clarity' of the descriptions of the FOCs in both rounds of the Delphi study. In the first round the means ranged from 4.77 to 6.46 and the standard deviations ranged from 0.78 to 2.07. The results of the second round were comparable: the means were between 4.82 and 6.27 and the standard deviations between 1.03 and 2.23.

For the first and the second round, the level of agreement among the experts was medium or high for 83.3% and 76.0% of the FOCs, respectively. To test the differences in ratings on comprehensiveness and clarity of the formulations between round 1 and 2, we used dependent *t*-tests. This difference was statistically significant for only one FOC ('active listening to patients'), based on a 2-tailed *t*-test (t(10)=1.5, p=0.03, r=0.64). The mean rating for this FOC in the second round was higher than in the first.

We also asked the panel members to indicate on a 7-point Likert scale the importance of the FOC for entrustment decisions. In the first round these means ranged from 6.50 to 7.00, and in the second round from 5.45 to 6.90. The level of agreement was high for 91.7% of the FOCs in the first round and for 100% of the FOCs in the second round (see Table 4). In both rounds and for both the judgments of 'comprehensiveness and clarity' and the judgments of 'importance', all ratings tended toward the 'strongly agree'-side of the scale, so in all cases the skewness was negative.

After round 2, minor changes were recommended regarding the description of 11 FOCs to complete the final list of 25 FOCs (Table 1).

	(*) difference between round 1 and round 2, $p < 0.05$		8	Kound 1			2	Kound 2	
ĽĽ	Facet of competence (FOC)	Mean	SD	LoA (table 2)	Skewness	Mean	SD	LoA (table 2)	Skewness
цщ	Teamwork and collegiality	Ι	Ι	I	I	5.27	1.90	Medium	-1.33
Ā	Knowing and maintaining own personal bounds and possibilities	5.00	1.73	Medium	-0.85	5.36	2.11	High	-1.37
æ	Role differentiation	5.50	1.92	Medium	-1.49	5.55	1.75	Medium	-0.91
A	Active professional development	5.90	1.60	High	-1.85	6.00	1.41	High	-0.88
J	Coping with uncertainty	5.40	2.17	High	-1.46	5.00	1.76	Low	-0.30
J	Coping with mistakes					5.00	1.70	Low	-0.34
St	Structure, work planning and	5.50	1.90	High	-1.52	5.70	1.16	High	-0.34
Р	Priorities	6.00	1.16	High	-2.17				
Ŗ	Responsibility	6.60	0.70	High	-1.11	5.90	1.85	High	-2.44
ž	Verbal communication with colleagues and supervisors	4.55	2.12	Low	-0.27	5.64	1.43	Medium	-0.71
\leq	Written (and digital) account—report to colleagues and supervisors	5.36	2.20	High	-1.05	60.9	1.14	High	-1.21
Ē	Empathy and openness	5.27	1.55	Low	-0.36	5.82	1.33	Medium	-0.54
At	Adapted informing of patients	60.9	0.94	High	-1.08	5.55	1.81	None	-1.64
æ	Respecting privacy and autonomy of the patient	5.64	1.75	Medium	-0.69	4.82	1.60	Low	-0.53
S	Scientific and empirical grounded method of working	6.10	1.20	High	-1.20	5.6	1.51	Medium	-0.61
Ш	Ethical awareness	5.90	1.45	High	-1.16	5.2	1.81	Medium	-0.65
¥	Active listening to patients*	5.00	1.73	None	-0.28	6.27	1.49	High	-2.79
A	Advising patients	4.64	1.75	Low	-0.55	4.82	2.23	Medium	-0.98
Ť	Handling emotions of patients and their relatives	5.45	1.29	Medium	-0.73	4.91	2.02	Low	-0.47
Ą	Attention to individual patient background	5.55	1.92	High	-1.70	5.64	1.03	High	-1.80
Ą	Attention to psychosocial aspects of health problems	5.73	1.55	High	-1.59	5.00	1.61	Low	-0.53
Ä	Attention to relatives and caregivers	6.18	0.98	High	-1.20	5.36	1.36	High	-1.68
Å	Active health promotion	5.80	1.87	High	-2.17	6.00	1.56	High	-2.18
J	Continuity in the care process	6.20	0.79	High	-0.41	5.80	1.69	High	-1.52
ŝ	Safety and risk management	5.00	1.94	Medium	-0.79	5.20	1.93	Medium	-0.93
Ξ	Financial and social awareness	Ι	I			5.40	1.96	Medium	-1.37
A	Appropriate use of sources	5.20	1.93	Medium	-1.16		Ι		Ι

(CC)			1 - 4 (4-1-1 - 4)	1			1 - 4 /4-1-5/	Ţ
Facet of competence (FOC)	Mean	SD	LoA (table 2)	Skewness	Mean	SD	LoA (table 2)	Skewness
Teamwork and collegiality		Ι			5.45	1.51	High	-1.19
Knowing and maintaining own personal bounds and possibilities	5.73	1.20	High	-1.16	6.82	0.40	High	-1.90
Role differentiation	5.82	2.18	High	-1.83	6.00	1.79	High	-1.79
Active professional development	6.80	0.42	High	-1.78	6.70	0.48	High	-1.04
Coping with uncertainty	6.70	0.48	High	-1.04	6.60	0.52	High	-0.48
Coping with mistakes					6.80	0.42	High	-1.78
Structure, work planning and	6.50	0.53	High	0.00	6.00	0.94	High	-0.99
Priorities	6.10	0.99	High	-1.09				
Responsibility	6.60	0.52	High	-0.48	5.90	1.85	High	-2.44
Verbal communication with colleagues and supervisors	6.73	0.47	High	-1.19	6.73	0.47	High	-1.19
Written (and digital) account—report to colleagues and supervisors	6.55	0.52	High	-0.21	6.73	0.47	High	-1.19
Empathy and openness	6.91	0.30	High	-3.32	6.90	0.30	High	-3.32
Adapted informing of patients	6.73	0.47	High	-1.19	6.64	0.50	High	-0.66
Respecting privacy and autonomy of the patient	6.73	0.47	Medium	-1.19	5.91	1.51	High	-1.93
Scientific and empirical grounded method of working	6.10	1.29	High	-1.79	6.70	0.48	High	-1.04
Ethical awareness	6.20	1.03	High	-1.24	6.30	1.25	High	-2.41
Active listening to patients	6.64	0.67	High	-1.80	6.90	0.30	High	-3.32
Advising patients	5.82	0.98	High	-0.35	5.91	2.02	High	-1.80
Handling emotions of patients and their relatives	5.55	1.51	High	-1.39	6.18	1.25	High	-1.91
Attention to individual patient background	5.91	1.76	High	-2.53	5.64	1.50	High	-1.62
Attention to psychosocial aspects of health problems	6.00	1.55	High	-1.97	6.00	1.48	High	-2.25
Attention to relatives and caregivers	6.27	0.65	High	-0.29	5.82	0.75	High	-1.40
Active health promotion	5.30	1.89	Medium	-1.41	5.80	1.23	High	-0.43
Continuity in the care process	6.30	0.82	High	-0.69	6.00	1.56	High	-2.18
Safety and risk management	6.30	1.25	High	-2.41	5.90	1.52	High	-2.15
Financial and social awareness		Ι			6.00	1.83	High	-2.74
Appropriate use of sources	5.40	1.84	High	-1.68			I	

Discussion

This study exemplifies which facets of the general competence of trainees seem to inform the decisions by clinical supervisors to trust them with critical tasks. It is important to map these facets, in the light of current approaches to assessment in competency-based medical education (Carraccio & Burke, 2010; Frank *et al.*, 2010; Ten Cate *et al.*, 2010; Jones *et al.*, 2011).

We made a preliminary list of FOCs, based on competency frameworks and the literature reflecting an international perspective (Frank, 2005; Kearney, 2005; Dutch Federation of University Medical Centers, 2009; General Medical Council, 2009; Ginsburg *et al.*, 2010; Sterkenburg *et al.*, 2010). After a Delphi study with 2 rounds, we constructed a list with 25 FOCs. The most important finding was that the experts agreed about the formulation of these FOCs. They strongly agreed that these FOCs are important or very important for the entrustment of critical clinical tasks to starting residents.

Earlier studies about entrustment decisions show that these decision processes are complex and relate to several factors (Dijksterhuis *et al.*, 2009; Sterkenburg *et al.*, 2010). These factors include the characteristics and achieved level of competence of the trainees. In the current study we clarified which FOCs of the trainees appear relevant for entrustment decisions.

Our study is limited by the small number of participants. However, because of the expertise of the participants and because of their high level of consensus, we believe that the findings are highly relevant, but the study cannot be viewed as conclusive. Further evaluation of the identified FOCs in other countries and different educational climates will be important to enhance generalizability of these findings and inform understanding of entrustability (Babbot, 2010; Boyce *et al.*, 2011).

The validity of our findings could be further supported by evaluating actual entrustment decisions and their relationships with trainees' FOCs. Finally, our population was selective and not representative of clinical supervisors in general. Participants' academic background made their views informative, but could have led to the listing of certain FOCs, although we do not have specific hypotheses in mind to support this concern. The results of this study identify general features of medical graduates that enable supervisors to entrust them with critical clinical tasks. The results can be useful for the development of a valid method for assessing medical graduates' readiness for clinical practice. The findings may also be useful for frame of reference training (Holmboe & Hawkins, 2008) for clinicians who must regularly make entrustment decisions.

Conclusions

We found consensus support by experts for a list of 25 important facets of competence (FOCs) for the entrustment of critical tasks to medical trainees.



Essential facets of competence that enable trust in medical graduates: a ranking study among physician educators in two countries

Chapter based on publication: Wijnen-Meijer, M., Van der Schaaf, M., Nillesen, K., Harendza, S. & Ten Cate, O. (Submitted for review and publication). Essential facets of competence that enable trust in medical graduates: A ranking study among physician educators in two countries.

Abstract

Background

One way to operationalize the assessment of trainees in a competencybased context is to determine whether they can be entrusted with critical activities. To determine which *facets of competence* (FOCs) are most informative for such decisions a Delphi study had been carried out among Dutch educators (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7).

Methods

In the current study, the resulting list of facets of competence was evaluated among experienced Dutch and German clinical educators to determine which facets appear most relevant and to evaluate the agreement among experts in different countries as a support for their external validity. Eight Dutch and eight German experts scored each FOC on a 5-point scale for relevance.

Results and conclusion

A rank-order comparison showed that there was almost full agreement about the top 10 FOCs, among those were 'Scientific and empirical grounded method of working', 'Knowing and maintaining own personal bounds and possibilities', 'Active professional development', 'Teamwork and collegiality', 'Active listening to patients', and 'Verbal communication with colleagues and supervisors'. We conclude that these facets of competence may be used in a frame of reference training for educators who need to make entrustment decisions about trainees.

Introduction

Following a rapid increase in competency-based medical education (Carraccio *et al.*, 2002; Frank *et al.*, 2010), the methods and concerns around the assessment of competence have been met with increasing interest (Norcini *et al.*, 2008; Lurie, 2012). Assessment tools in a workplace that cannot be standardized and increased interest in the 'softer' skills pose challenges to the assessment procedures (Lurie *et al.*, 2009; Lurie *et al.*, 2011). One approach that has been suggested to operationalize the attainment of competencies is to determine whether or when a trainee can be trusted to execute a professional activity without supervision (Ten Cate, 2005).

Trust in trainees requires observations that do not only draw on standardized skills and knowledge but take other facets of competence into account (Kennedy *et al.*, 2008; Sterkenburg *et al.*, 2010). Using a Delphi approach, Wijnen-Meijer *et al.* have investigated the factors that educators in The Netherlands find important to consider when making entrustment decisions about medical trainees (accepted for publication, *chapter* 7). This yielded a list of factors, or facets of competence (FOCs) as they were called.

The aim of our current study was to determine the external validity, i.e. the generalizability, of the factors that these Dutch educators found essential. Our approach was to ask educators in two countries to rankorder these FOCs and to determine the level of agreement among the countries about the highest scoring FOCs.

The study was carried out among Dutch and German medical educators. The Netherlands and Germany differ in medical education culture, particularly in the sense that in the Netherlands education reform has dominated medical curricula throughout the country since the mid 1970s, while such processes in Germany have started only recently (Ten Cate, 2007; Nikendei *et al.*, 2009). It is fair to say that Dutch medical schools have modern medical curricula (not necessarily better) and most German medical schools have predominantly traditional medical curricula, while a number of other countries in Europe have positions in between. Agreement among medical educators in The Netherlands and Germany about important FOCs would strengthen the relevance of those FOCs.

Methods

After completion of the Dutch descriptions of the FOCs they were translated to German as input for the ranking study.

Participants

To find comparable groups in both countries we approached all 24 Dutch experts who had contributed to the Delphi study. They had been selected on the grounds that they were experienced clinicians, involved in postgraduate training and hold a chair in education (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7). We also approached 36 German physicians of different institutes and disciplines. All of them were experienced clinicians, each with a master's degree in medical education.

Questionnaire

The experts were invited by email to complete a questionnaire in electronic format. To gain insight into the experts' preferences for FOCs, they were asked to assign a score from 1 = 'least important' to 5 = 'most important' to each of the 25 FOCs that had resulted from the Delphi study. They were requested to give each possible score (1-5) five times (so, 5 FOCs had to get score 1; 5 FOCs had to get score 2, etc.).

Data analysis

We calculated means, medians and standard deviations for both the Dutch and German group of experts and a level of agreement, according to an adapted De Loe procedure (De Loe, 1995). De Loe used this method for a 4-point scale, while we used the same method for a 5-point scale (see Table 1).

		the level of agreement le, according to De Loe
Agreement	1 category	2 contiguous categories
High	70%	80%
Medium	60%	70%
Low	50%	60%
None	<60% of ratings ir	a 2 contiguous categories

 TABLE 1 CALCULATION OF LEVELS OF AGREEMENT ACCORDING TO DE LOE (1995)

		Rai	Rank order			Dutch edu	Dutch educators (n=8)			German edu	German educators (<i>n</i> =8)	8)
Facet of competence (FOC)	Comb.	NL	넁	Ave. mean	Median	Mean	S	LoA	Median	Mean	S	LoA
Scientifically and empirically grounded method of working	-	2		4.51	5.0	4.63	0.52	High	5.0	4.38	1.41	High
Knowing and maintaining own personal bounds and possibilities	2	÷	Ś	4.32	5.0	4.63	0.52	High	4.0	4.00	0.93	Low
Active professional development	ŝ	-	5	4.19	5.0	4.75	0.46	High	4.5	3.63	1.69	Low
Teamwork and collegiality	4	7	2	4.01	3.5	3.63	1.06	Low	4.5	4.38	0.74	High
Active listening to patients	5	4	4	3.94	4.0	4.00	0.93	Low	4.0	3.88	1.13	Low
Verbal communication with colleagues and supervisors	9	9	7	3.63	3.5	3.75	1.17	None	3.5	3.50	1.50	None
Empathy and openness	7	10	9	3.44	3.0	3.25	1.04	Low	3.5	3.63	1.30	None
Responsibility	8	5	10	3.38	4.0	3.75	1.39	Medium	3.0	3.00	1.93	None
Coping with mistakes	6	6	6	3.26	3.5	3.38	1.41	None	3.5	3.13	1.64	None
Safety and risk management	10	8	11	3.25	4.0	3.50	1.41	Low	3.0	3.00	1.41	Low
Written (and digital) account / report to colleagues and supervisors	11	11	15	2.94	3.5	3.13	1.13	Medium	3.0	2.75	1.67	None
Attention to individual patient background	12	15	16	2.82	2.5	2.88	1.36	None	3.0	2.75	1.04	Low
Respecting privacy and autonomy of the patient	13	13	17	2.82	2.5	3.00	1.15	None	2.5	2.63	1.30	None
Advising patients	14	14	18	2.75	2.5	3.00	1.20	Low	2.5	2.50	0.54	High
Handling emotions of patients and their relatives	15	12	21	2.69	3.0	3.00	1.20	Low	2.5	2.38	0.74	High
Structure, work planning and priorities	16	20	8	2.65	2.0	2.00	0.93	Medium	4.0	3.29	1.70	None
Ethical awareness	17	17	12	2.63	2.5	2.38	1.30	None	3.0	2.88	1.46	None
Continuity in the care process	18	16	19	2.63	3.0	2.75	1.28	Low	2.0	2.50	1.41	Low
Adapted informing of patients	19	19	13	2.51	2.0	2.13	1.13	Low	3.0	2.88	0.84	Medium
Attention to psychosocial aspects of health problems	20	18	22	2.38	2.0	2.38	1.19	Medium	2.0	2.38	1.19	Low
Active health promotion	21	24	14	2.32	1.5	1.75	0.89	Medium	3.0	2.88	1.25	Low
Financial and social awareness	22	23	20	2.16	1.0	1.88	1.64	High	2.0	2.43	1.51	None
Role differentiation	23	22	23	2.13	1.5	1.88	1.36	High	2.5	2.38	1.30	None
Coping with uncertainty	24	21	24	2.07	1.5	2.00	1.20	Low	1.0	2.13	1.55	Medium
Attention to relatives and caregivers	25	25	25	1.76	2.0	1.63	0.52	High	1.5	1.88	1.36	High

Results

Respondents

In total 8/24 Dutch and 8/36 German experts participated in the ranking study (response rates 33% and 22%, respectively). In the Dutch group, 7 were male and the average age was 61 years (56–66 years). In the German group, the average age was 43 years (33–53 years) with 4 male and 4 female responders. The responders in both rounds represented a wide range of surgical and non-surgical disciplines (Dutch: cardiology, general practice, gynaecology, internal medicine, neurology, oncology and surgery; German: emergency medicine, gynaecology, internal medicine, psychiatry and surgery).

Results

The results of the ranking study are presented in Table 2. The medians and means indicate the ranking of the FOCs. For the Dutch group of responders, medians were between 1 and 5 and the means between 1.63 and 4.75 (SD 0.46–1.64). For the German group, medians varied between 1 and 5 and the means between 1.88 and 4.38 (SD 0.54–1.93). For both groups, the level of agreement varied from none to high.

Discussion

The aim of this study was to determine which FOCs of medical trainees are considered most important to formulate entrustment decisions by experienced supervisors in residency training and to evaluate the international agreement.

There appeared to be a large variation in attributed importance among the FOCs; means ranged between 1.75 and 4.5. However, we found strong agreement between physician educators from the Netherlands and physician educators from Germany in ranking the competency facets relevant for entrustment decisions. The top 10 for the total group of responders, based on medians and means, appeared nearly the same as the top 10 for each of the two countries separately. The only FOC that substantially differed was 'structure, work planning and priorities'. This FOC was ranked 8 in the German group, but 16 among the Dutch. For only 4 of the 25 FOCs the difference in ranking between the two groups was statistically significant (p<0.05).

The fact that two groups physician educators from different medical education cultures highly agreed on the importance of certain FOCs for entrustment decisions, strengthens the relevance of these FOCs. The top-10 align with Kennedy *et al.*'s findings. In their grounded theory study, these authors have also found "truthfulness" (the absence of deception) and "conscientiousness" (the thoroughness in data gathering and dependability) to be an important quality that supervisors value in trainees to determined their readiness for independent clinical work (Kennedy *et al.*, 2008) We had not explicitly included these facets in our list, but it may be assumed that our respondents would score this item highly if they had been included. Implicitly they are reflected in our items that stress scientifically grounded working, openness, responsibility and coping with mistakes, all part of the top-10.

Our findings may be used for a frame of reference training (Holmboe *et al.*, 2011) for clinicians who must regularly take entrustment decisions.

Conclusion

Our aim was to reveal what facets of competence are considered most important for entrustment decisions by supervisors of residents. We found high consensus between experts from the Netherlands and Germany, despite large differences in their curricula. Our findings are relevant for the development of assessment instruments to evaluate whether medical graduates are ready for clinical practice.



An argument-based approach to the validation of UHTRUST: can we measure how recent graduates can be trusted with unfamiliar tasks?

Chapter based on publication Wijnen-Meijer, M., Van der Schaaf, M., Booij, E., Harendza, S., Boscardin, C., Van Wijngaarden, J. & Ten Cate, Th.J. (Conditionally accepted for publication). An argument-based approach to the validation of UHTRUST: Can we measure how recent graduates can be trusted with unfamiliar tasks? *Advanced in Health Sciences Education*.

Abstract

Introduction

There is a need for valid methods to assess the readiness for clinical practice of medical graduates. This study evaluates the validity of UHTRUST, an authentic simulation procedure to assess whether medical trainees are ready to be entrusted with unfamiliar clinical tasks at the highest level of Miller's pyramid. This assessment, in which candidates were judged by clinicians, nurses and standardized patients, addresses the question: Can this trainee be trusted with unfamiliar clinical tasks? The aim of this paper is to provide a validity argument for this assessment procedure. We collected data from various sources during preparation and administration of a UHTRUST-assessment.

Method

In total, 60 candidates (30 from the Netherlands and 30 from Germany) participated. To provide a validity argument for the UHTRUST-assessment, we followed Kane's argument-based approach for validation. All available data were used to design a coherent and plausible argument. Considerable data was collected during the development of the assessment procedure. In addition, a generalizability study was conducted to evaluate the reliability of the scores given by assessors and to determine the proportion of variance accounted by candidates and assessors.

Results and discussion

It was found that most of Kane's validity assumptions were defendable with accurate and often parallel lines of backing. UHTRUST can be used to compare the readiness for clinical practice of medical graduates. Further exploration of the procedures for entrustment decisions is recommended.

Introduction

Background

The goal of academic education is to ensure that students acquire knowledge within the domain of their future profession. Besides knowledge, complex skills such as communication skills, organization skills and problem solving skills can be considered essential for professionals to function in a professional environment (Birenbaum & Dochy, 1996; Fraser & Greenhalgh, 2001). In medical education there is growing support for the idea that trainees should be able to apply knowledge to solve clinical problems (Kreiter & Bergus, 2008; Ten Cate *et al.*, 2010; Wittert & Nelson, 2009).

While standardized assessment of factual knowledge and practical skills is common, medical educators have increasingly realized that such tests have their limitations. For instance, problems are reported regarding the translation of the outcomes of such tests to the real world of health care (Arnold, 2002; Epstein, 2007; Ginsburg *et al.*, 2010; Ginsburg, 2011; Howley, 2004; Newble, 2004; Wass *et al.*, 2001). This has led to a call for a higher validity of assessment procedures in clinical education (Crossley *et al.*, 2011; Govaerts *et al.*, 2007; Tavares & Eva, 2012; Wetzel, 2012).

Miller's pyramid is a well know metaphor that delineates levels of assessment in medical training, (Miller, 1990). Miller's third level ('shows how') and level 4 ('does') reflect work behaviour in a standardized setting and in the actual workplace, respectively. The Objective Structured Clinical Examination (OSCE), in which trainees address different structured case-fragments sequentially (Harden & Gleeson, 1979), is an example of an assessment at level 3. The strength of the OSCE is its potential to provide evidence of reliability, because of the high degree of standardization. Assessment methods at level 4 aim to capture what a trainee actually does in clinical practice; Examples are the mini-clinical evaluation exercise (mini-CEX, Norcini *et al.*, 1995) and Direct Observation of Procedural Skills (DOPS, Barton et al., 2012). These kinds of assessment provide useful information about the readiness for clinical practice of trainees, but the clinical context precludes high levels of reliability, as the clinical context cannot be standardized.

To address this problem, we have attempted to develop a simulation that reflects the authentic clinical situation better than any other test we know of. Critical in our approach is that we acknowledged the ever-changing, unpredictable clinical context. Coping with real life clinical problems means coping with unfamiliar situations. To assess whether medical trainees are ready to be entrusted with unfamiliar clinical tasks, at the highest level of Miller's pyramid, we designed the 'Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test' (UHTRUST). The aim of this paper is to evaluate its validity by building a validity argument for this assessment procedure.

The UHTRUST assessment procedure

An authentic simulation for medical graduates at MD level was developed to evaluate their readiness for clinical practice. The key question we wanted to address was whether they would be ready to be entrusted with critical clinical activities, so called "Entrustable Professional Activities" (EPAs), (Ten Cate *et al.*, 2010; Ten Cate, 2005; Ten Cate, 2006). This matches the idea that one important characteristic of medical professionals is that they must be trusted to work without supervision (Freidson, 1970).

During the assessment, candidates were situated in the role of a beginning resident on a very busy day with the initial instruction: "This is your first day as a resident, at a ward which is yet unknown to you. Unfortunately, your supervisor is called away. It is not possible to cancel the patient appointments, so you will be responsible for them, but you can call your supervisor for help whenever you feel the need to."

The assessment procedure consisted of three phases. In the first phase (1 hour), the candidates saw five consecutive standardized patients that "had been admitted to the hospital" with non-common medical problems. During the second phase (3 hours) the candidates had time to gather information from the internet, to request lab results and x-rays, to determine differential diagnoses and to draw up a management plan to enable a presentation of each patient to the supervisor at the end of the day. During this phase the candidates also faced distracting tasks, such as reported changes in a patient's condition, questions from nurses or junior students and an urgent organizational problem that needed to be solved. The candidates had the opportunity to call their supervisors by phone if needed and halfway the day there was a brief face-to-face meeting with the supervisor to discuss the candidate's progress. In the third phase (30 minutes), the candidates reported their differential diagnoses and management options. Each candidate was independently assessed by two or three clinicians, one nurse and six standardized patients (SPs—five of them acted as patients and the sixth played the role of the husband of one of the patients) on different facets of their competence, resulted from a Delphi study among clinical educators (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7; Wijnen-Meijer et al., submitted, chapter 8).

One of the clinicians acted as the candidate's personal supervisor during the assessment. The second clinician was present all day and listened in to telephone conversations and face-to-face conversations between the supervisor and the candidate. The third clinician only observed the candidate during the reporting phase. The nurses observed the candidates during the second phase and they deliberately disturbed the candidates in a systematic way with several distracting, but real life tasks. An example of a patient case can be found in *appendix A*.

After all observations the clinicians were asked to individually indicate how much supervision they estimated this trainee would require on nine EPAs, unrelated to the observed scenarios. Figure 1 provides a schematic overview of UHTRUST.

Phase	Activities	Duration	Assessors
	Briefing	30 minutes	
Phase 1	Short meeting with supervisor // consultation by five patients	1 hour	Standardized patients
	Walking to next location	10 minutes	
Phase 2	Collection of diagnostic information about five patients // seven distracting tasks // halfway meeting with supervisor // drawing up management plans // if needed: calls with supervisor	3 hours	Clinicians 1 and 2, Nurse
	Walking to next location	10 minutes	
Phase 3	Report and discuss examination plans and treatment plans	30 minutes	Clinicians 1, 2 and 3
	Debriefing	30 minutes	

FIGURE 1 SCHEMATIC OVERVIEW OF UHTRUST

The patient cases and distracting tasks were selected in such way that they covered the breadth of the medical profession to a large extent, while at the same time the assessment fitted within one day. Similar to an OSCE, the tasks were standardized as much as possible. Differences were the addition of distracting tasks, the fact that the different cases ran simultaneously which make it more similar with clinical practice and the possibility to interact with a supervisor who also assesses, also adding to authenticity.

Quality requirements regarding assessments

High stakes assessments need to be reliable and valid. Reliability refers to consistency of scores across repeated measurements and is considered an important condition for validity. Hence reliability is a component of the validity argument (Wass & Archer, 2011; Holmboe & Hawkins, 2008). This definition of reliability has remained generally stable during the past decades (Brennan, 2006). In contrast, the concept of validity has changed over the years. Brennan (2006) describes how the concept was first defined in terms of the predictive power for future performance (Cureton, 1951). In the nineteen seventies and eighties, Cronbach (1971) and Messick (1989) introduced the "unitary" notion of validity. According to their view, validity focuses on the appropriateness of inferences from test scores, and not just on the validity of the assessment instrument itself. According to the Standards for Educational and Psychological Measurement (Brennan, 2006, p.2) the preferred definition of test validity is "the degree of which evidence and theory support the interpretations of test scores entailed by purposed used of tests".

Kane (1992) introduced the *argument-based approach* to validity, which is consonant with Messick's ideas of validity, but focuses on the methodology for validation. The goal is to provide a structured and coherent analysis of all evidence in support of the interpretation of test scores. The argument that underpins validity leads from the test administration to the interpretation of scores. Kane (1992, 2006) labels four components in this inferential chain: scoring, generalization, extrapolation and interpretation. The scoring component of the argument requires evidence regarding the appropriateness of the assessment conditions, data collection and scoring procedure. The generalization component requires evidence that the observed score, coming from the task sample used, is generalizable to a broader domain i.e. the test domain. In the extrapolation component, the interpretation is extended to the practice domain. This requires evidence of the link between the data collected in the test and the behaviours of interest in the real world. The fourth component of the argument concerns interpretation. Here, a final conclusion is drawn: what implications logically result from the observed skill level of the candidate?

The argument-based approach can be applied in two main stages (Kane, 2006). The first stage is during the development of the assessment procedures; the proposed interpretation of assessment scores and use are specified by means of an interpretative argument. The second is when the assessment is finished. In the second stage, when the measurement product is finished, the proposed argument is critically evaluated.

Besides reliability and validity as psychometric aspects of assessments, several authors also mentioned the importance of utility aspects, which are costs, feasibility and acceptability (Van der Vleuten, 1996; Messick, 1995; Linn, 1991). We will also pay attention to these descriptors in this study.

Threats to the validity of authentic assessments

Authentic assessments bring along specific threats to validity. A first threat concerns the reliability of the scoring process. Assessors bring unique characteristics and experiences to the rating process, and it appears to be difficult for assessors to exclude biases stemming from their personal background and to prevent selective observations (Bakker, 2008; Govaerts *et al.*, 2007; Sterkenburg *et al.*, 2010). Candidates can respond to complex tasks in authentic assessments in very different ways, which makes it difficult for assessors to score such performances consistently (Gipps, 1994; Moss, 1994).

A second threat concerns the specificity of assessment tasks. A small number of tasks in an authentic assessment makes it difficult to establish a sample of tasks that is representative to the universe of generalization (Kane, 2006). A high score on one task does not necessarily mean a high score on a different task (Bakker, 2008; Lane & Stone, 2006). Because only a small number of tasks can be included in one assessment, the scores awarded to one task have a relatively large effect on the total score (Kane, 2004).

The third threat to the validity of authentic assessments, particularly to the extrapolation, is construct-irrelevant variance (Messick 1989; Messick, 1995). Assessors inherently bring sources of constructirrelevant variance to the rating process, such as attention to irrelevant features, or tendencies to overly severe or lenient rating (Gingerich *et al.*, 2011; Lane & Stone, 2006; Tavares & Eva, 2012). Also tasks that are too easy or too difficult can lead to construct-irrelevant variance (Albanese, 2011; Messick, 1995).

The fourth threat to validity is construct underrepresentation. This occurs when the content of the test is not reflective of relevant knowledge or skills. Therefore, the results of the test do not make it possible to make statements about the candidate's abilities (Messick, 1989; Messick, 1995). This threat is related to both the extrapolation and interpretation part of the validity argument.

The aim of study reported in this paper is to provide a validity argument for the UHTRUST-assessment procedure, following an argument-based approach for validation. We have chosen this broader approach to integrative interpretations based on different types of relevant evidence because of its appropriateness for complex assessment methods like UHTRUST. We supplemented the argument-based approach for validation with our findings regarding utility aspects.

Method

Participants

The assessment procedure was developed by four educationalists, together with two recently graduated doctors and six experienced physicians. During the assessment days at two locations in July and August of 2011, 60 candidates participated (30 from the Netherlands and 30 from Germany). All candidates were nearly graduated from medical school at the moment of the assessments. They participated voluntarily and had applied in reaction to announcements. In Utrecht (the Netherlands) the candidates were assessed by in total 14 physicians, 4 nurses and 6 SPs. In Hamburg (Germany) 11 physician assessors, 4 nurses and 18 SPs were involved.

All candidates were assessed by at least two assessors; 20 Dutch and 6 German candidates were also judged by a third clinician. The difference between the number of SPs is caused by the fact that in Utrecht all 30 times the 6 patient roles were consistently played by the same SPs, while in Hamburg every role rotated between 3 SPs. The physicians and nurses were invited to participate, based on their clinical experience and experience with supervising trainees. Furthermore, at each location about 30 persons assisted with the organization of the assessment.

Scoring instruments

The physicians completed three kinds of scoring forms for each candidate. One scoring form included seven so called *facets of competence* (FOCs) that can be considered key in making entrustment decisions by supervisors about residents (see Table 2). These facets were developed during a Delphi study among physician supervisors (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7; Wijnen-Meijer *et al.*, submitted, *chapter* 8). For each FOC, the candidates were scored on a 3-point Likert scale of 1 = 'weak' to 3 = 'good' for each of five different patient cases. Next to this, the assessors gave an overall score for each FOC on a 5-point Likert scale, from 1 = 'very weak' to 5 = 'very good' (*appendix B*, *page* 227).

The second questionnaire consisted of nine so called *Entrustable Professional Activities* (EPAs; see Table 3), tasks that are suitable to entrust to a trainee once sufficient capability is attained for unsupervised practice (Ten Cate, 2005). The physicians were asked to indicate on a 5-point scale how much supervision they estimated the candidate needs for these EPAs. (1 = 'he/she is not able to do this', 2 = 'he/she is able to do this under direct supervision', 3 = 'he/she is able to do this if supervision is available', 4 = 'he/she is able to do this independently', 5 = 'he/she is able to supervise others in performing this activity'). The EPAs did not reflect the activities or pathologies involved in the actually observed activities (*appendix B, page 228*).

The third form was a so-called *Post Patient Encounter Form* (PPEF), based on the Post-Encounter Form designed and validated by Durning *et al.* (2012). The candidates summarised on this form for each patient case the most important problems, differential diagnoses and a proposal for treatment. The assessing physicians scored these aspects on a 5-point Likert scale from 1 = 'below expectation' to 5 = 'above expectation' (*appendix B, page 229*).

The fourth scoring form was completed by the nurses. This scoring form contained six FOCs, which were similar to six of the seven FOCs that were scored by the clinicians. For each FOC, the candidate was scored by a nurse on a 3-point Likert scale of 1 (weak) to 3 (good) for their performance regarding five different disturbances. Additionally, the nurse gave an overall score for each FOC on a 5-point Likert scale, from 1 = 'very weak' to 5 = 'very good' (*appendix B, page 230*).

The SPs completed the so called CARE-questionnaire, a validated instrument consisting of 10 questions to measure consultation skills and empathy (Mercer *et al.*, 2004). The SPs scored the candidates on these items with a 5-point scale from 1 = 'poor' to 5 = 'excellent' (*appendix B, page 231*).

Evaluation forms

Three months before the assessment days were scheduled pilot assessments were organized in Utrecht and in Hamburg. At the end of these pilots, all candidates, clinicians, nurses, SPs and staff members evaluated the organization and content of the assessment. This information was used to make adjustments to the assessment and to gather information for the argument for validity.

Procedure to the development of the argument for validity

A theoretical framework for the validity argument was developed based on various theoretical and empirical studies related to the argumentbased approach to validity (Bakker, 2008; Chapelle *et al.*, 2010; Hawkins *et al.*, 2009; Hawkins & Holmboe, 2008; Kane, 2004; Kane, 2006). The framework illustrates the four major inferences that are associated with an argument-based approach to validity and their underlying assumptions (see Table 1).

The writing of the argument for validity was an iterative process. All pieces of validity evidence had to be collected and arranged in a way

that did justice to Kane's argument-based approach to validity. During the development of the assessment, considerations and decisions were discussed and written down. All available data were used to design a coherent and plausible argument. To do so, the data were linked to one or more of the four major inferences in the argument for validity.

Analysis

The results from the scoring forms were analysed by means of Pearson correlation coefficients between all FOCs and EPAs. In addition, a generalizability analysis was conducted on the FOCs and EPAs to determine the proportion of variance accounted by the candidates and by the assessors. We also calculated means and standard deviations of the scores on the evaluation forms.

Ethical approval

Ethical approval for the Dutch part of the study was obtained from the NVMO Ethical Review Board. For the German part, ethical approval was obtained from the State of Hamburg Physicians' Ethics Board.

Infere	ences (I) and assumptions (A)	Warrants licensing the assumptions
l1 A1.1	Scoring: from the observed perfor The assessment conditions are appropriate	 The assessment conditions were standardized, so candidates were provided with equal opportunities to show their abilities and test scores could be compared to one another. A detailed planning was defined to ensure a smooth running of the assessment day. This planning was evaluated after the pilots. It was unlikely that the candidates had access to the assessment tasks prior to the assessment days.
A1.2	The scores are recorded ac- curately	 All candidates were judged by multiple trained assessors. The assessors judged the candidates' behavior on seven FOCs and nine EPAs. Assessors were urged to follow a systematic and transparent scoring procedure. This reduces the risk of invalid and unreliable judgments. Instructed staff members were assigned to the different groups of assessors. These staff members checked whether the scoring forms were filled out correctly and answered assessors' questions. In every room staff members were present to prevent the candidates to exchange information regarding the cases. Security measures were taken to prevent loss of data and to protect the assessments' integrity.
A1.3	The scoring criteria are appropri- ate and acceptable	 A panel of informed experts agreed upon the content and language of seven FOCs. These were used to develop an analytic scoring rubric. Global rating scales were used to score the FOCs. They were expected to be more feasible, equally reliable, and more valid than detailed (dichotomous) behavio- ral checklists. Experts developed nine EPAs that were used to indi- cate to what extent the candidates can be entrusted with critical clinical tasks.
A1.4	Reliable and valid scoring of the performance by the assessors	 The sample and selection procedure of the assessors was acceptable. During a <i>frame of reference</i> training assessors attempted to reach shared understanding of the content and performance standards. Assessors were also informed on how to avoid typical rater errors. The internal consistency (reliability), calculated by means of phi-coefficient, of the raters for the FOCs varies from acceptable to good (see Table 2). The reliability of the raters for the EPAs varies from moderate to acceptable (see Table 3).
12	Generalization: from the observed	score to the expected universe score

- 12 Generalization: from the observed score to the expected universe score
- A2.1 The scores are stable and random error due to different occasions, raters and tasks is controlled
- The standardization measures described in inference 1 controlled the random error caused by administration occasion, rater and tasks.

Infere	nces (I) and assumptions (A)	Warrants licensing the assumptions
		 Multiple assessors per candidate were used to reduce the influence of personal biases of the individual assessors. Candidates were confronted with multiple cases. This reduced the variance caused by tasks specificity, and provided the candidates with the opportunity to demonstrate their competences on multiple occasions. G-studies are conducted to determine the percentage of variance that can be explained by the candidate and by the assessor. For the FOCs the percentage of variance accounted by the candidate is relatively high (see Table 2); for the EPAs this percentage is lower (see Table 3).
A.2.2	The sample of observations is representative of the universe of generalization	 It was only possible to sample a relatively small number of assessment tasks. In order to compensate this deficiency, serious effort was made to draw a representative sample from the universe of generalization. Experts were consulted during the task development and evaluation. They made a blueprint for the content of the assessment in order to make sure that the task sample could not be completed without the use of the defined FOCs.
13	Extrapolation: from the universe s	core to the expected level of skill in the target domain
A3.1	The universe score is related to the level of skill of the graduate in the target domain	 The authentic character of the assessment makes the argument for extrapolation plausible. When a comprehensive construct is measured, the practical limits of assessment must be accepted. UHTRUST provided negative evidence on the candidates' true ability to cope with unfamiliar clinical situations.
A3.2	There are no systematic errors that are likely to undermine the extrapolation	 The standardized assessment conditions and the use of standardized patients brought about an artificial aspect to the assessment. Sources of irrelevant variance caused by systematic differences between SPs (and real patients) and time pressure were identified and controlled.
14	Interpretation: from the level of sk	ill in the target domain to the test interpretation
A4.1	All assumptions are defensible with accurate and plausible evidence	 Most assumptions were defendable with accurate (and often parallel lines of) backing.
A4.2	The data acquired by the as- sessment can be used for the intended purposes	 Because most validity assumptions were defendable with accurate and often parallel lines of backing, this assessment can be used for the intended purpose.

 Table 1
 Framework for an Argument for Validity—inferences, assumptions and warrants licensing the assumptions

Results: argument for validation of UHTRUST

A summary of the inferences is presented in Table 1.

Inference 1: Scoring

From the observed performance to the observed score

Assumption 1.1. The assessment conditions are appropriate The UHTRUST assessment procedure was administered on two different locations and with two different groups of candidates, physicians, nurses and standardized patients (SPs). To enhance similarity between the two administration occasions, the observations of the assessment were made under semi-standardized assessment conditions (Cohen & Wollack, 2006; Kane, 2006).

Near the assessment days standardized instructions (e.g., time limits), conditions of administration and guidelines for scoring were established and checked by the testdevelopers. This implied how a valid response or judgment had to be constructed, what ancillary materials were allowed for the candidates and how much help supervisors and nurses were expected to provide. During the assessment it was ensured that all candidates were assessed with the same clinical content and tasks and under the same conditions.

UHTRUST was meant to be a realistic assessment with open-ended tasks. In this kind of assessment it is difficult to discern all potential threats to standardization, "including those associated with SP portrayal, unanticipated student reactions to the scripted SP responses and case irregularities" (Holmboe & Hawkins, 2008, page 105). To enhance standardization, the evaluations and recordings made during the pilot assessments were used to optimize the cases and strengthen the effectiveness of instructions for all participants.

To ensure smoothly running assessment procedures a detailed planning was constructed. This planning was partly based on empirical and theoretical studies containing useful advice derived from successfully run authentic assessments (Boursicot & Roberts, 2005; Cohen & Wollack, 2006; Holmboe & Hawkins, 2008). Besides detailed time schedules for candidates, physicians, nurses, SPs and staff members, it included descriptions of necessary practical and logistical arrangements, to begin with the pilot study on two sites. Its evaluations among both Dutch and German participants (n=84) revealed that the pilot days were felt well organized. Combined mean score on this item was 4.11 (SD=0.44) on a 5-point scale.

We made sure candidates had no access to the assessment tasks prior to the assessment administration, as the scores of these candidates would then not accurately reflect their ability levels. The cases of UHTRUST were developed especially for this assessment in the months prior to the assessment days. Premature exposure to the assessment tasks was therefore unlikely. One case about hemolytic uremic syndrome was replaced as an unusually large outbreak of this disease at one site (Hamburg) took place between the pilot study and the main study (Harendza, 2011).

Assumption 1.2. The scores are recorded accurately

The use of a systematic and transparent scoring procedure reduces the risk of invalid and unreliable judgments (Bakker, 2008; Kane, 2006). Prior to the assessment, all assessors received a frame of reference training (Holmboe & Hawkins, 2008), including explanations about the impact of scoring errors and biases. In addition, all assessors were orally trained in using the scoring forms and received written instructions about the scoring procedures. At both assessment days, instructed staff members were assigned to the different groups of assessors. They checked whether the scoring forms were filled out correctly and readable, and were available to answer any assessors' questions.

To prevent loss or mixing up of data various security measures were taken. At the beginning of the day, all candidates, assessors, SPs and staff members checked-in. At the end of the day all participants checkedout and staff members made certain that all intended documents had been received.

UHTRUST was administered for research purposes only. Since there was no impact on academic progress or graduation or any risk of harmful consequences for those with high or low scores, and because participation was voluntary, it was assumed that the candidates' motivation to cheat was low. To be sure, in every room staff members were present to prevent candidates to exchange information regarding cases.

Assumption 1.3. The scoring criteria are appropriate and acceptable

Scoring criteria (i.e., content standards) indicate what candidates should know or be able to do. They guide the assessors' judgment about the quality of a candidates' performance (Gipps, 1994). The development of the *facets of competence* (FOCs) was split into two phases. First, in a Delphi study and ranking study ten FOCs were identified as most important to entrust critical clinical tasks to a trainee (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7; Wijnen-Meijer *et al.*, submitted, *chapter* 8). This study was conducted in the Netherlands and afterwards

validated in Germany. Second, in the main study seven FOCs were scored by the physicians and nurses and three FOCs were judged by the SPs. The FOCs were not further specified in sub-criteria, resulting in rather global assessment criteria, which are commonly used for the measurement of discrete constructs (e.g. communication, empathy) as well as in assessing more broad constructs (e.g. the ability to take responsibility) (Hawkins *et al.*, 2009). Detailed (dichotomous) checklists (e.g. makes eye contact, introduces themselves) for such broad constructs often fail to validly capture essential features, due to a difficulty to quantify elements of expert behavior (Holmboe & Hawkins, 2008).

At the end of the day, the assessing physicians were asked to indicate for each observed candidate to what extent they would entrust this person with new critical clinical activities. In order to do so, a variety of nine EPAs was identified (see Table 3). The developers made sure that each of the identified FOCs would be necessary in at least one EPA, to ensure that all EPAs together covered all FOCs. Pearson correlations were calculated to verify these relationships (Table 4). All FOCs correlated significantly with all EPAs (p<0.01).

Assumption 1.4. Reliable and valid scoring of the performance by the assessors For performance assessments, the quality of the assessment as a whole is related to the ability of the assessors to use the scoring criteria to reach a technically and professionally defensible conclusion (Dwyer, 1995). All physicians and nurses involved in UHTRUST were selected because of their active clinical and supervising experience, to ensure they are capable to make profound judgments about a candidate's ability. All assessors participated voluntarily. This is important, because high motivation contributes to the quality of the rating outcomes (Govaerts *et al.*, 2007).

To prepare the assessors for judging and scoring the performance of the candidates, a frame of reference training was delivered (Holmboe & Hawkins, 2008). The training was standardized and was given by the same two instructors. The training sessions at both locations were similar. During the training, the assessors learned to apply the detailed scoring procedure in systematic and consistent way. The assessors had to elaborate and share conceptualizations of what constitutes competent behavior and were asked to formulate standards for acceptable and unacceptable behaviors for each of the FOCs. This was important given the fact that the FOCs refer to abstract qualities which assessors need to infer from the performance of the candidate. The assessors reached consensus about performance standards and rating scale anchors for each of the scoring criteria. According to Knight (2002), there will always be a certain degree of ambiguity about the meaning of criteria and interpretation of the standards. However, it is important to put trust in the judgment of an expert and not to quell creativity (Gipps, 1994; Ten Cate, 2006). The scoring criteria, the detailed scoring procedure and the assessor selection procedure and training were all designed and implemented to maximize objective and reliable scoring. However, they do not yet guarantee a high quality assessment processes (Nijveldt, 2007). That is why the effects of these measures where statistically examined.

To estimate the reliability of the clinician raters on the seven FOCs, the EPAs and the PPEFs a G-study was conducted, for both two physicians (who observed the candidate all day) and three physicians (including the assessor who only observed the reporting phase). The reliability (phi-coefficient) of three raters for the FOCs varies from acceptable (0.73) to good (0.88) (see Table 2). The reliability of three raters for the EPAs varied from moderate (0.36) to acceptable (0.72) (see Table 3). A G-study was also conducted for the PPEFs that summarised for each patient case the most important problems, differential diagnoses and a management proposal. The reliability (phi-coefficient) for both two and three physicians varied from acceptable (0.63 and 0.64 respectively) to good (0.90 and 0.89 respectively) except for 'Problem' of case 3 (0.39 and 0.42 respectively) and 'DD' of case 5 (0.59 and 0.48 respectively) (see Table 5). For all scores there is little difference in reliability when two or three assessors are included.

Inference 2: Generalization

From the observed score to the expected universe score

Assumption 2.1. The scores are stable and random error due to different occasions, raters and tasks is controlled

Any facet that is allowed to vary in the universe of generalization (e.g. tasks, assessors) and that is sampled by the measurement procedure contributes to random error of an assessment score (Kane, 2006; Lane & Stone, 2006). The standardization measures, as described in the section about the scoring inference, were implemented to reduce for random error caused by three variables: the administration occasion, the assessors and the tasks. Furthermore, measures were taken to increase stability of scoring. First, the stability of scoring was enhanced by the use of multiple assessors per candidate. This reduces the influence of personal biases of the individual assessors (Kane, 2006).

Second, the use of multiple cases compensates the psychometric limitations inherent to a single case, assuming that all cases have adequate quality. The candidates encountered five patient cases and seven distracting tasks, designed to sample skills more broadly over the course of the assessment day. The use of multiple cases also enhanced the accuracy of the test scores, because the assessors were given the opportunity to judge the candidates based on their performance on multiple occasions, and the candidates were provided with several opportunities to demonstrate their competence.

In a G-study, the overall variance of the test is calculated for multiple sources. If three assessors are included, we find percentages of variance explained by the candidate to be relatively high for the FOCs (varying from 47% to 71%, see Table 2). For the EPAs they were lower (16% to 46%, see Table 3) and for the PPEFs (Table 5) there is more divergence, particularly for the items 'problems' (19% to 66%) and 'DD' (24% to 74%). For 'treatment' they were more consistent and relatively high (41% to 67%). In addition we found that percentages hardly differ if either two or three assessors are included.

Assumption 2.2. The sample of observations is representative of the universe of generalization

According to Bakker (2008), the selection of representative samples of assessment tasks is an important issue in performance assessments. To make sure that the UHTRUST tasks would cover the selected FOCs, ten medical experts were consulted during the task development phase, to make a blueprint for the content of the assessment. They were consulted how cases should play out and agreed that all tasks included in UHTRUST together portrayed a sufficiently broad content.

The raters' judgments about the candidates' ability to take responsibility for unfamiliar situations were also thought to be influenced by the way candidates would handle the disruptions during the second phase of the assessment. These additional tasks were also thoroughly planned and discussed by the experts.

Inference 3: Extrapolation

From the universe score to the expected level of skill in the target domain

Assumption 3.1. The universe score is related to the level of skill of the graduate in the target domain

Serious efforts were made to achieve a high level of physical and psychological fidelity. First, the choice to work with SPS instead of written patient scenarios, contributed to the level of realism. In performing the simulation, the SP does not only present the gestalt and history of the patient being simulated, "but the body language, physical findings and emotional and personality characteristics as well" (Cleland *et al.*, 2009, p.478). The psychological fidelity was further enhanced by the tasks and the modes of presentation. Lane and Stone (2006) stated that these kinds of high-fidelity tasks can easily be translated to expected performance in the real world. The tasks were designed in such a way that they could also occur on a real clinical ward, including referral letters and the opportunity to request for lab and radiology results. Finally, the act of observing can interfere with the level of authenticity. However, the candidates were never observed by assessors at unrealistic moments. For example, none of the assessors were present during the patient encounters.

The effect of our efforts to maximize authenticity was evaluated after the pilot assessments. On the evaluation forms both candidates and assessors were asked whether the pilot had a high level of authenticity. The candidates' mean score on this item was 4.3 (n=18, SD=0.66) and the assessors' mean score was 3.9 (n=20, SD=0.55) on a 5-point scale. Another question was whether the assessed competences were relevant for clinical practice. The mean score of the candidates on this item was 4.4 (n=20, SD=0.59) and of the assessors 4.15 (n=19, SD=1.3) on a 5-point scale.

The basic assumption is that all activities in the test domain are necessary for the effective dealing with unfamiliar clinical situations in the practice domain. This assumption makes it reasonable to expect that candidates who were successful in the assessment would also be successful in unfamiliar situations in reality. However, this cannot be taken as absolute evidence. When a candidate demonstrates that he or she is capable to take responsibility for unfamiliar situations during UHTRUST, this is no guarantee that this behavior will be manifested in real clinical settings, e.g., when a candidate lacks other skills or features (such as motor coordination) that were not included in the test. Such limitations of the validity of assessments must be accepted (Knight, 2002). Kane (2004) states that for most performance assessments this assumption tends to be stronger on the negative side than it is on the positive side. Even though not every aspect of the construct can be measured, it is reasonable to assume that a candidate who showed serious deficiencies in the test domain would also show deficiencies in the practice domain.

Assumption 3.2. There are no systematic errors that are likely to undermine extrapolation

As mentioned above, the assessment and scoring conditions were fixed for all candidates. Some of these standardization measures brought about an artificial aspect to UHTRUST and resulted in sources of systematic error that had to be identified and controlled. First, also trained SPs can be a source of construct irrelevant variance. Holmboe and Hawkins (2008) stated that even though little research on the subject exists, it is inevitable that differences between SPs and real patients occur. For this reason the acceptability of the performance of the SPs during the pilot assessments was checked on the evaluation form. Candidates gave the plausibility of the SPs' performances a mean score of 4.55 (n=20, SD=0.58) on a 5-point scale. Furthermore, based on the recordings of the encounters in the pilot study, the training for the SPs was improved.

Second, time pressure can yield invalid measures of proficiency, contributing to construct-irrelevant variance (Holmboe & Hawkins, 2008). The medical experts that were involved in the development process indicated that the time allotted for the candidates was short, but realistic. On the evaluation form, candidates of the pilot (n=20) were asked if they had had enough time available to complete the tasks in the individual phases of the assessment. On average, the candidates were quite satisfied with the amount of time they had for various phases. For all phases, the mean score was higher than 3 on a 5-point scale. During the evaluation of the pilot assessments, various candidates and assessors indicated that time pressure is often present in clinical practice, and therefore should not be seen as irrelevant variance.

Inference 4: Interpretation

From the level of skill in the target domain to the test interpretation

Assumption 4.1. All assumptions are defensible with accurate and plausible evidence

To make sure that the most crucial validity assumptions were critically considered and substantiated with accurate and plausible backing, the inferences and assumptions that underlie an argument for validity were made explicit in a theoretical framework in advance (Table 1). The aim of this study was to evaluate the most prominent aspects of a performance assessment: planning, standardization measures, scoring procedure, reliability of the scores, the authenticity level and the investigation of potential threats to validity (Holmboe & Hawkins, 2008).

Assumption 4.2. Data acquired by the assessment can be used for the intended purposes

UHTRUST is being developed to answer the question to what extent medical graduates can be entrusted with clinical tasks in unfamiliar situations. As most validity assumptions were defendable with accurate and often parallel lines of backing, UHTRUST can be used for the intended purpose: the formative assessment of the readiness for clinical practice of medical graduates.

Utility aspects

For educational practice, also utility aspects are important. The acceptability of UHTRUST among the participants was good, as all persons involved were positive about it. This can be derived from the results of the evaluation form. Concerning the feasibility and costs of UHTRUST can be said that implementation requires considerable investments in regard to time and effort of staff and assessors and, as a result, considerable finances. These are comparable with the investments needed for the implementation of an OSCE (Boursicot & Roberts, 2005). Whether this is too much is a difficult to answer question, as it depends on the interest one has in a valid outcome. Studies to establish predictive validity can help to support such decisions.

		bility coeffi- nt)	Percentage of variance accounted by candidate		Percentage of variance accounted by assessor	
Facets of competence	2	3	2	3	2	3
1. Scientific and empirical grounded method of working	0.88	0.88	71	70	3	4
2. Knowing and maintaining own personal bounds and possibilities	0.85	0.84	65	64	7	5
3. Teamwork and collegiality	0.79	0.73	56	47	13	11
4. Verbal communication with colleagues and supervisors	0.85	0.80	64	57	9	10
5. Responsibility	0.88	0.87	70	69	1	0
6. Safety and risk management	0.74	0.74	48	48	10	9
7. Active professional development	0.89	0.88	72	71	0	0

 TABLE 2 QUALITY MEASURES FOCS, GROUPED BY NUMBER OF ASSESSING CLINICIANS

	(Phi-	bility coeffi- nt)	of va accoui	Percentage of variance accounted by candidate		ntage riance nted by essor
Entrustable professional activities	2	3	2	3	2	3
1. Emergency assistance with acute cardiac failure	0.71	0.70	45	44	9	7
2. Handling a patient complaint	0.48	0.42	23	19	44	40
3. Pre-operative information and consent	0.50	0.49	25	24	37	36
4. Breaking bad news	0.27	0.36	11	16	56	56
5. Clinical reasoning under time pressure	0.71	0.68	45	42	17	14
6. Solving a management problem	0.62	0.63	36	36	36	35
7. Suspicion of self-induced disease	0.51	0.48	26	23	28	29
8. Handling of a seriously ill patient	0.63	0.61	36	34	40	41
9. Interaction with a consultant	0.71	0.72	45	46	13	12

 TABLE 3 QUALITY MEASURES EPAS, GROUPED BY NUMBER OF ASSESSING CLINICIANS

	F0C 1	FOC 2	FOC 3	FOC 4	FOC 5	FOC 6	F0C 7
EPA 1	0.686	0.605	0.471	0.552	0.570	0.581	0.545
EPA 2	0.476	0.363	0.418	0.407	0.458	0.466	0.375
EPA 3	0.502	0.507	0.423	0.393	0.495	0.497	0.429
EPA 4	0.407	0.317	0.362	0.311	0.389	0.369	0.262
EPA 5	0.628	0.587	0.445	0.537	0.605	0.615	0.540
EPA 6	0.553	0.512	0.391	0.451	0.530	0.458	0.485
EPA 7	0.507	0.464	0.408	0.379	0.486	0.485	0.471
EPA 8	0.564	0.494	0.427	0.441	0.516	0.448	0.524
EPA 9	0.558	0.493	0.465	0.516	0.561	0.502	0.455

Table 4 Pearson correlation coefficients between EPAs and FOCs. All correlations are significant at the 0.01 level (2-tailed)

		(Phi-o	Reliability (Phi-coeffi- cient)		Percentage of variance accounted by candidate		ntage riance nted by essor
Post-Patient Encounter For	ms	2	3	2	3	2	3
Description of problems	Case 1	0.63	0.64	37	37	28	23
	Case 2	0.84	0.86	63	66	15	10
	Case 3	0.39	0.42	18	19	28	23
	Case 4	0.73	0.75	48	50	20	16
	Case 5	0.68	0.66	42	39	2	8
Differential diagnosis	Case 1	0.85	0.77	66	53	8	15
	Case 2	0.85	0.86	65	68	9	9
	Case 3	0.67	0.75	40	50	17	6
	Case 4	0.90	0.89	76	74	4	2
	Case 5	0.59	0.48	32	24	0	0
Proposal for treatment	Case 1	0.85	0.79	65	55	0	13
	Case 2	0.78	0.68	54	41	5	9
	Case 3	0.84	0.86	63	67	0	0
	Case 4	0.84	0.80	64	57	19	22
	Case 5	0.71	0.71	45	45	0	0

TABLE 5 QUALITY MEASURES POST-PATIENT ENCOUNTER FORMS, GROUPED BY NUMBER OFASSESSING CLINICIANS

Discussion

UHTRUST is an authentic assessment procedure that intends to measure a broad and complex construct i.e. the extent to which medical graduates can be entrusted with (unfamiliar) critical clinical activities. In the current study Kane's argument-based approach to validity was used to write an argument to support the validity of UHTRUST. The construction of an argument for validity is an iterative process that should lead to continued improvement in the quality and defensibility of an assessment (Kane, 2006). It was found that most validity assumptions were defendable with accurate and often parallel lines of backing. Based on this argument we conclude that the assessment can be used for the intended purpose.

One of the weaker components of the validity argument of UHTRUST is the reliability of the EPA scores, which is moderate for some EPAs. This is the most novel part of the assessment procedure, as we not only asked to evaluate observed behavior, but also to determine trust in future behavior. We are not aware of other procedures that attempt to measure this construct. It can be assumed that uncertainty among assessors increases if they must make inferences to predict unobserved behavior. Sterkenburg *et al.* (2010) found that anesthesiology physicians value substantial acquaintance with trainees highly as a condition to trust them.

In addition, context variables, such as time of the day, hospital personnel and facilities available influence entrustment decisions (Dijksterhuis, 2009; Sterkenburg, 2010). This aspect needs attention in future applications of UHTRUST. Further improvement of the training or instructions for assessors and additional information about the candidates could increase the reliability of the EPA scores, as well as including context limitations.

To determine the utility of UHTRUST, the *utility equation* of assessments (Van der Vleuten, 1996) is useful. As stated above, the reliability and validity evidence is acceptable. Furthermore, the educational impact is good: candidates noted that the assessment gave them a lot of information about their strengths and weaknesses. They considered UHTRUST the "most comprehensive examination ever encountered" (quote by one candidate). Also the acceptability appears to be good: the assessors and candidates were very positive about this assessment. Despite the considerable investments regarding time and finances that are needed for the implementation of UHTRUST, which are comparable with an OSCE, it is not unfeasible. The results of the G-studies show the difference in reliability between assessors who observed the candidate

all day and those who only observed the candidate during the reporting phase, is small. This indicates that it is possible to reduce the required time investments of clinicians.

To provide a validity argument for UHTRUST, we followed Kane's argument-based approach for validation. This approach appeared to be useful to shed a light on the overall validity of complex assessments such as UHTRUST. It makes decisions explicit during the development and implementation of the assessment, and consequently the strengths and weaknesses. It remains disputable whether or not more validity evidence should have been gathered. For example, in our study, interviews with assessors could have given more insight in the quality of their cognitive processes and the effects of the assessor training. More discussion about how much and what kind of validity is needed should therefore be valuable. Kane (2006) stated that it is unlikely that all inferences can evaluated. The decision which inferences should and which should be not evaluated depends mainly on the purpose of the assessment.

In our opinion, UHTRUST can be implemented for different purposes. One possibility is to compare the readiness for clinical practice of medical graduates from different medical schools, in the context of research or to find out whether curricular change is needed. The assessment can also be used to judge the performance of individual graduates, for instance for the purpose of residency selection for postgraduate training programs. Further exploration of the possible implementations of UHTRUST, including conditions and consequences, is recommended.



Vertically integrated medical education and the readiness for practice of graduates

Chapter based on publication: Wijnen-Meijer, M., Ten Cate, Th.J., Van der Schaaf, M., Burgers, C., Borleffs, J.C.C. & Harendza, S. (Submitted for review and publication). Vertically integrated medical education and the readiness for practice of graduates.

Abstract

Background and aim

Medical curricula become more and more vertically integrated (VI) to prepare graduates better for clinical practice. VI curricula show early clinical education, integration of biomedical sciences and focus on increasing clinical responsibility levels for trainees. Results of earlier questionnaire-based studies indicate that the type of the curriculum can affect the perceived preparedness for work as perceived by students or supervisors. The aim of the present study is to determine difference in actual performance of graduates from VI and non-VI curricula.

Methods

We developed and implemented an authentic performance assessment based on different facets of competence for medical near-graduates in the role of beginning residents on a very busy day. Fifty-nine candidates participated: 30 VI (Utrecht, The Netherlands) and 29 non-VI (Hamburg, Germany). Each candidate was independently assessed on different facets of competence by two clinicians, a nurse and standardized patients. Afterwards, the clinicians indicated how much supervision they estimated each candidate would require on nine so called *Entrustable Professional Activities* (EPAs) unrelated to the observed scenarios.

Results

We did not find many differences between the two groups of candidates. Graduates from a VI curriculum received significantly higher scores by the clinicians for the facet of competence 'active professional development', with features like 'reflection' and 'asking for feedback'. In addition, we found some differences in the scores for the EPAs and the scores given by the nurses and standardized patients.

Conclusions

This study gives an impression of the actual performance of medical graduates from VI and non-VI curricula. The findings do not match the results of earlier questionnaire based studies regarding integrated curricula, which showed stronger differences. Possible causes of this discrepancy are discussed.

Introduction

Background

In the past decades, many medical school curricula have been revised so that they represent vertically integrated (VI) programmes. A fully vertically integrated undergraduate medical curriculum can be defined by the following four features (Wijnen-Meijer *et al.*, accepted for publication, *chapter 6*): provision of early clinical experience; integration of biomedical sciences and clinical cases; progressive increase of clinical responsibility longitudinally and extended clerkships in the final year of medical school (Harden *et al.*, 1984; Schmidt *et al.*, 1996; Vidic & Weilauf, 2002; Ten Cate, 2007). One specific aim of vertical integration is to facilitate the transition from medical school to clinical practice and postgraduate training (Vidic & Weilauf, 2002).

In earlier studies we found that, in comparison with those who have followed non-VI programmes, graduates of VI curricula appear to make definitive career choices earlier, need less time and fewer applications to obtain residency positions and feel more prepared for work and postgraduate training (Wijnen-Meijer *et al.*, 2009, *chapter 4*; Wijnen-Meijer *et al.*, 2010, *chapter 5*). Results of other studies indicate that the type of the curriculum can affect the perceived preparedness for work (Goldacre *et al.*, 2003; Watmough *et al.*, 2006; Cave *et al.*, 2009; Wijnen-Meijer *et al.*, accepted for publication, *chapter 6*). All of these studies are questionnaire-based analyses of student or faculty perceptions.

To substantiate these impressions there is a need to investigate the actual performance of graduates from VI and non-VI curricula. As known from previous research, two factors contribute to the development of (medical) performance: the amount of knowledge and the amount of experience in practice (Ericsson *et al.*, 2006). Keijers *et al.* (2009) found that students who had followed a VI curriculum had less knowledge of basic medical sciences at graduation in comparison with students from a more conventional, non-VI curriculum.

Little is known about the impact of the increased and qualitatively different clinical experience of students in VI curricula. By hearsay we know that the supervisors of postgraduate training programmes have much appreciation for the proactive clinical functioning and competences of VI graduates, such as coping with unfamiliar clinical situations. This could be explained by the fact that VI graduates have had more clinical experience during the first years of medical school and have had extended clerkships, giving them the opportunity to see more types of patients. Besides, students get more responsibility during their final clerkships. Numerous studies described the positive influence of these curriculum aspects on learning (Dornan & Bundy, 2004; Dunphy & Williamson, 2004; Cantillon & MacDermott, 2008).

Aim and research question

The purpose of the current study was to determine differences in readiness for clinical practice between graduates from a VI curriculum and those from a non-VI curriculum. More specifically, the research question was: Do graduates from a medical school with a VI curriculum differ from graduates from a medical school with a non-VI curriculum in their competences to cope with unfamiliar clinical situations?

Context

The study was carried out among near-graduates from the medical schools in Utrecht and Groningen (The Netherlands) and Hamburg (Germany). The medical schools in Utrecht and Groningen have a vertically integrated curriculum, according to the description above. An important part of the VI curriculum is the degree of responsibility that final year medical students take on. In their role as *semi-physicians* they are expected to bear responsibility for patient care comparable to the responsibility of junior residents (Ten Cate, 2007; Wijnen-Meijer *et al.*, 2010, *chapter* 5).

A recent Dutch national review committee judged positive about the increased responsibility in final year clinical clerkships that current medical programs show, based on interviews with clinicians involved in residency programs (Hillen, 2012). The reason to compare graduates from these Dutch schools with non-Dutch graduates was that other Dutch schools have developed of are developing VI curricula too. Differences that existed some years ago (Wijnen-Meijer *et al.*, 2010, *chapter 5*) are gradually disappearing. It was decided to compare with German graduates to maximize chance to find an effect.

The curriculum in Hamburg was not vertically integrated during the execution of this study (Nikendei *et al.*, 2009). In this curriculum there is more emphasis on the acquisition of knowledge, especially in the first two years of the programme.

In addition, in the medical schools at Utrecht and Groningen, more curriculum time is spent on training communication skills. In both countries, students enter medical school directly after finishing secondary education and in both countries undergraduate medical training lasts six years. We asked the Educational Commission for Foreign Medical Graduates (ECFMG) of the United States to provide us with comparative data of Dutch and German applicants over the period 2002–2011. These data showed a pass rate at first attempt for all German applicants for United States Medical Licensing Examination (USMLE) Step 1 Basic Sciences of 76.6% (n=1861) and 61.1% for Dutch applicants (n=193). For USMLE Step 2 Clinical Knowledge these figures were exactly similar at 85.5% (Germans, n=1413; Dutch, n=110); for USMLE Step 2 Clinical Skills they were 84.3% (Germans, n=1241) and 89.6% (Dutch, n=86).

The numbers are too low to draw firm conclusions, but they align with our prior impression that German curricula focus more on basic science knowledge and Dutch more on clinical skills.

Validity

In a separate study we elaborated a validity argument for the assessment procedure used in the current study, following Kane's approach (Kane, 1992; Kane, 2006). It is argued that the procedure is valid for the assessment of the readiness for clinical practice of medical graduates (Wijnen-Meijer *et al.*, conditionally accepted for publication, *chapter 9*).

Method

Procedure

We developed and implemented an authentic performance assessment for medical students near graduation to evaluate their readiness for clinical practice. The key question we wanted to address was whether the graduate can be entrusted with critical clinical activities. We defined these as situations that have not necessarily been encountered during clerkships, but require adequate coping by junior residents. The assessment was developed in Utrecht in collaboration with Hamburg. Consequently the assessment was called 'Utrecht Hamburg Trainee Responsibility for Unfamiliar Situation Test' (UHTRUST). Because the assessment had taken place in two different countries, all procedures were identically designed in two languages (Dutch and German). During the assessment, candidates were placed in the position of beginning residents on a very busy day: "This is your first day as a resident at a ward which is yet unknown to you. Unfortunately, your supervisor is called away. It is not possible to cancel the patient appointments, so you will be responsible for them, but you can call your supervisor for help whenever you feel the need to."

Phase	Activities	Duration	Assessors
	Briefing	30 minutes	
Phase 1	Short meeting with supervisor // consultation by five patients	1 hour	Standardized patients
	Walking to next location	10 minutes	
Phase 2	Collection of diagnostic information about five patients // seven distracting tasks // halfway meeting with supervisor // drawing up management plans // if needed: calls with supervisor	3 hours	Clinicians 1 and 2, Nurse
	Walking to next location	10 minutes	
Phase 3	Report and discuss examination plans and treatment plans	30 minutes	Clinicians 1 and 2
	Debriefing	30 minutes	

FIGURE 1 SCHEMATIC OVERVIEW OF UHTRUST

The assessment consisted of three phases. In the first phase (1 hour), the candidates saw five standardized patients at the outpatient department with unusual medical problems. During the second phase (3 hours) the candidate had time to gather information on the internet or in pocket books, and to request additional information, e.g. lab results. The candidates had to make a diagnosis for the five patients and to draw up an examination or treatment plan. During this phase the candidates also had to face seven realistic distracting tasks, like changes in one of the patient's condition, questions from nurses or junior students and an urgent organizational problem that had to be solved. The candidates had the opportunity to call their supervisors and halfway there was a meeting planned between candidate and supervisor to discuss the progress. In the third phase (30 minutes) the candidates reported their differential diagnoses and options for policy or treatment. Figure 1 provides a schematic overview of UHTRUST. For an example of patient case, see *appendix A*.

At both locations pilot assessments were organized a few months before the assessment days. One goal of the pilot was to improve the assessment procedure based on the experiences from the pilot. The other goal was to rehearse the complex organization.

Assessors

Each candidate was independently assessed by two physicians, a nurse and six standardized patients (SPs) on different facets of competence. One of the physicians also acted as the candidate's supervisor. The second physician was present all day and listened in to conversations between supervisor and candidate through speaker amplified cell phones and during the face-to-face meetings. The nurses observed the candidate during the second phase and they deliberately disturbed candidates with distracting tasks. After all observations, physicians were asked to individually indicate how much supervision they estimated this trainee would require on nine so called *Entrustable Professional Activities* (EPAs), unrelated to the observed scenarios. Prior to the assessment, all assessors received a frame-of-reference training (Holmboe & Hawkins, 2008). This training included explanations about the use of scoring forms and the impact of scoring biases.

Participants

During the assessment days in July and August of 2011, 60 candidates participated: 23 from Utrecht, 7 from Groningen (The Netherlands) and 29 from Hamburg (Germany). The candidates from Utrecht and Groningen participated in the assessment on the same day, which took place in Utrecht. All candidates had nearly graduated from medical school at the moment of the assessments. They participated voluntarily and they presented themselves in reaction to announcements. At both locations, ten physicians and four nurses were involved. Furthermore, in Utrecht six standardized patients (SPs) and in Hamburg 18 SPs participated. The difference between the numbers of SPs is caused by the fact that in Utrecht all 30 times the six roles were played by the same SPS, while in Hamburg for practical reasons every role rotated between three SPS. The physicians and nurses were invited to participate, based on their clinical experience and experience with supervising trainees.

Instruments

The physicians completed three kinds of scoring forms for each candidate. One scoring form included seven so called *facets of competence* (FOCs) that are a key in making entrustment decisions by supervisors about residents (see Table 1). These facets had been developed during a Delphi study among physician supervisors (Wijnen-Meijer *et al.*, accepted for publication, *chapter* 7; Wijnen-Meijer *et al.*, submitted, *chapter* 8). For each FOC, the candidates were scored on a 3-point Likert scale of 1 = 'weak' to 3 = 'good' for each of five different patient cases. Next to this, the assessors gave an overall score for each FOC on a 5-point Likert scale, from 1 = 'very weak' to 5 = 'very good'. The form can be found in *appendix B, page 227*.

The second questionnaire consisted of nine so called *Entrustable Professional Activities* (EPAs), tasks that are suitable to entrust to a trainee once sufficient capability is attained for unsupervised practice (Ten Cate, 2005) (see Table 2). The physicians were asked to indicate on a 5-point scale how much supervision they think that the candidate would need for these EPAs, which were different from the actually observed activities (1 = 'he/she is not able to do this', 2 = 'he/she is able to do this under direct supervision', 3 = 'he/she is able to do this if supervision is available', 4 = 'he/she is able to do this independently', 5 = 'he/she is able to supervise others in performing this activity') (Ten Cate & Scheele, 2007). The form can be found in *appendix B, page 228*.

The third form was a so called *Post-Patient Encounter Form* (PPEF), based on Durning's Post-Encounter Form (Durning *et al.*, 2012). The candidates summarised on this form for each patient case the most important problems, differential diagnoses and a proposal for treatment. The assessing physicians scored these aspects on a 5-point Likert scale from 1 = 'below expectations' to 5 = 'above expectations'. The form can be found in *appendix B*, *page 229*.

The fourth scoring form was completed by the nurses and contained six FOCs, similar to six of the seven FOCs scored by the clinicians (see Table 3). For each FOC, the candidates were scored by the nurse on a 3-point Likert scale of 1 = 'weak' to 3 = 'good' for their performance regarding five different disturbances. Additionally, the nurse gave an overall score for each FOC on a 5-point Likert scale, from 1 = 'very weak' to 5 = 'very good'. The form can be found in *appendix B, page 230*.

The SPs completed the so called CARE-questionnaire, a validated instrument consisting of 10 questions to measure consultation skills and empathy (Mercer *et al.*, 2004). The SPs scored the candidates on a 5-point scale (1 = 'poor', 2 = 'fair', 3 = 'good', 4 = 'very good', 5 = 'excellent'). The questionnaire included three questions about making action plans together with the patient. Because the candidates in this assessment did not make any action plans, only seven questions of the questionnaire were applicable in this study (see Table 4).

At the German site, a different 5-point scale was used to score the CARE-questionnaire (1 = 'not at all applicable', 2 = 'hardly applicable', 3 = 'partially applicable', 4 = 'largely applicable', 5 = 'fully applicable'). To correct the differences between the two scales, we transformed the scores into a 3-point scale. (Dutch questionnaire: $1 \rightarrow 1$, $2 \rightarrow 2$, $3 \rightarrow 3$, $4 \rightarrow 3$, $5 \rightarrow 3$; German questionnaire: $1 \rightarrow 1$, $2 \rightarrow 1$, $3 \rightarrow 2$, $4 \rightarrow 3$, $5 \rightarrow 3$). The form can be found in *appendix B*, *page 231*.

Finally, the candidates completed the NEO-FFI personality test (Costa & McCrae, 1992). This is a validated test containing 60 items on a 5-point Likert scale (from 1 = 'totally disagree' to 5 = 'totally agree'), measuring five personally dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness to experience.

Analyses of the data

Mann-Whitney tests were computed to compare the scores of the candidates from the VI and non-VI curriculum. We conducted Chisquare-tests to compare the results of the personality tests of the two groups of candidates. To compare the scores given by Dutch and German assessors, we used Mann-Whitney tests and we calculated the degree of consistency among the assessors with Jury alpha. Not all candidates had completed all PPEF forms. We calculated Pearson correlation coefficients between the number of completed PPEFs and the scores on FOCs and EPAs, as we suspected that the number of completed PPEFs could be indicative of the candidate's proficiency.

Ethical approval

Ethical approval for the Dutch part of the study was obtained from the NVMO Ethical Review Board. For the German part, ethical approval was obtained from the State of Hamburg Physicians' Ethics Board.

Results

Background information regarding the candidates

In the Dutch group 20 (67%) candidates and in the German group 22 (76%) candidates were female. These percentages approximately reflect the gender distribution of the total groups of medical graduates. Dutch candidates (M=24.4 year) were younger than the German candidates (M=26.0 year) based on independent samples 2-tailed *t*-test (t(57)=4.65, p=0.00, r=0.47). We found no significant correlations between age and scores, so we assumed that this difference in age did not affect the results.

Based on the NEO-FFI personality test, there were no differences between the two groups regarding the dimensions neuroticism, extraversion and conscientiousness. The Dutch candidates scored higher on the dimension 'openness to experience' ($\chi^2(8,59)=17.74$, p=0.02). The German candidates had a significantly higher score on the dimension agreeableness ($\chi^2(7,59)=15.02$, p=0.04).

Comparison groups of assessors

Because the Dutch candidates were assessed by Dutch assessors, and the German candidates by German assessors, we wanted to know whether there were systematic differences in scoring between the groups of assessors (Schijven et al., 2010). Prior to the actual assessment we showed both groups of clinicians and SPs the same video recording of one Dutch and one German candidate who had participated in the pilot of the UHTRUST assessment. Both candidates had been selected at random. We asked the assessors to judge both candidates on the FOC and EPA scoring forms to be used during the assessment days. We found no statistically significant differences between their mean scores (Mann-Whitney U). We also examined the consistency among the clinicians' scores by calculating Jury alpha, for both the total group of clinicians and the Dutch and German group separately. In all cases the internal consistency was very high: Jury alpha varied from 0.96 to 0.98. For the nurses, it was not possible to compare groups, because of the small population.

For five of seven aspects of the CARE questionnaire, the German SPs gave significantly higher scores in comparison with the Dutch SPs (p<0.05). This was the case for both candidates. The consistency between the scores of the SPs at either site was very high (Jury alpha is 0.98).

Facets of competence

Table 1 shows the mean scores of the clinicians on the seven FOCs for the two groups of candidates. Regarding the FOC 'active professional development' the Dutch candidates (M=3.55, SD=0.77) received higher scores than the German candidates (M=2.81, SD=1.14, U=278, z=-2.41, p=0.02, r=0.31).

- (*) Mann-Whitney test: <i>U</i> =278, <i>z</i> =-2.41, <i>p</i> =0.02, <i>r</i> =0.31	(VI curr	tch iculum) =30	German (non-VI curriculum) <i>n</i> =29		
Facets of competence	М	SD	М	SD	
1. Scientific and empirical grounded method of working	3.23	1.01	3.22	0.94	
2. Knowing and maintaining own personal bounds and possibilities	3.32	0.75	3.14	1.01	
3. Teamwork and collegiality	3.34	0.61	3.50	0.93	
4. Verbal communication with colleagues and supervisors	3.50	0.94	3.50	0.93	
5. Responsibility	3.38	0.75	3.28	1.12	
6. Safety and risk management	3.02	0.83	3.24	0.96	
7. Active professional development*	3.55	0.77	2.81	1.14	

 TABLE 1 SCORING FACETS OF COMPETENCE BY CLINICIANS (5-POINT SCALE; MEAN SCORE OVER TWO ASSESSORS)

Table 2 compares the mean scores for the two groups of candidates given by the nurses. There is a significant difference between the groups regarding the FOC for 'knowing and maintaining own personal bounds and possibilities' (Mann-Whitney U=487, z=2.08, p=0.04, r=0.28). German candidates (M=3.46, SD=0.71) had higher scores on this FOC than the Dutch candidates (M =3.13; SD=0.63). This is also the case for 'teamwork and collegiality' (Dutch candidates: M=3.17, SD=0.65; German candidates: M=3.58, SD=0.64, U=514, z=2.25, p=0.02, r=0.30).

- (*) Mann-Whitney test: <i>U</i> =487, <i>z</i> =2.08, <i>p</i> =0.04, <i>r</i> =0.28 (†) Mann-Whitney test: <i>U</i> =514, <i>z</i> =2.25, <i>p</i> =0.02, <i>r</i> =0.30	(VI curr	(VI curriculum) (non-VI cu		man urriculum) =29
Entrustable professional activities	М	SD	MS	
1. Scientific and empirical grounded method of working	3.10	0.85	3.08	0.63
2. Knowing and maintaining own personal bounds and possibilities*	3.13	0.63	3.46	0.71
3. Teamwork and collegiality†	3.17	0.65	3.58	0.64
4. Verbal communication with colleagues and supervisors	3.10	0.85	3.17	0.76
5. Responsibility	3.27	0.69	3.25	0.85
6. Safety and risk management	2.93	0.98	3.20	0.91

 TABLE 2 SCORING FACETS OF COMPETENCE BY NURSES (5-POINT SCALE)

Entrustable Professional Activities

Table 3 shows the mean scores on the EPAs for the two groups. As can be seen, we found statistically significant differences for two EPAs. For 'breaking bad news' the German group (M=3.10, SD=0.54) scored higher than the Dutch group (M=2.62, SD=0.70, Mann-Whitney U=617, z=2.85, p=0.00, r=0.37). In contrast, the Dutch candidates (M=3.70, SD=0.53) received higher scores for 'solving a management problem' than the German candidates (M=3.19, SD=0.75, U=265, z=2.67, p=0.01, r=0.35).

(*) Mann-Whitney test: U=617, z=2.85, p=0.00, r=0.37 (†) Mann-Whitney test: U=265, z=2.67, p=0.01, r=0.35	(VI curr	tch iculum) :30	(non-VI cu	man ırriculum) =29
Facets of competence	М	SD	М	SD
1. Emergency assistance with acute cardiac failure	2.33	0.53	2.36	0.57
2. Handling a patient complaint	2.70	0.92	2.86	0.63
3. Pre-operative information and consent	3.42	0.54	3.43	0.59
4. Breaking bad news*	2.62	0.70	3.10	0.54
5. Clinical reasoning under time pressure	2.38	0.57	2.38	0.56
6. Solving a management problem †	3.70	0.53	3.19	0.75
7. Suspicion of self-induced disease	2.93	0.57	2.91	0.44
8. Handling of a seriously ill patient	2.40	0.58	2.21	0.68
9. Interaction with a consultant	3.23	0.60	3.28	0.64

 TABLE 3 SCORING ENTRUSTABLE PROFESSIONAL ACTIVITIES BY CLINICIANS (5-POINT SCALE;

 MEAN SCORE OVER TWO ASSESSORS)

Post-Patient Encounter Forms

The candidates were instructed to complete a Post-Patient Encounter Form (PPEF) for each of the five patient cases prior to the reporting phase, with a summary of the most important problems, differential diagnoses and a proposal for treatment. Within the Dutch group 95% of the PPEFs were completed, the German group only completed 67%.

We calculated Pearson correlation coefficients between the number of completed PPEFs and the scores on FOCs and EPAs. We found significant correlations for five FOCs and two EPAs which led us to believe that the number of completed PPEFs is an indicator of proficiency. This would mean that within the German group the PPEF responders make a positively biased subgroup. Because of these findings we decided not to compare scores on the PPEFs between the two groups.

Standardized patients: CARE-questionnaire

The mean scores on the CARE-questionnaire are presented in Table 4. The German candidates received significantly higher scores for 'making you feel at ease' (M=2.86, SD=0.18, U=639.50, z=3.22, p=0.00, r=0.42) and 'really listening' (M=2.88, SD=0.12, U=573.50, z=2.21, p=0.03, r=0.29), in comparison with the Dutch candidates (M= 2.69, SD=0.21 and M=2.79, SD=0.17 respectively).

For the aspect 'being positive' the Dutch candidates had significantly higher scores (M=2.80, SD=0.19) than the German candidates (M=2.58, SD=0.36, U=279, z=-2.43, p=0.02, r=0.32).

(*) Mann-Whitney test: <i>U</i> =639.50, <i>z</i> =3.22, <i>p</i> =0.00, <i>r</i> =0.42 (†) Mann-Whitney test: <i>U</i> =573.50, <i>z</i> =2.21, <i>p</i> =0.03, <i>r</i> =0.29 (‡) Mann-Whitney test: <i>U</i> =279, <i>z</i> =-2.43, <i>p</i> =0.02, <i>r</i> =0.32	(VI curr	tch 'iculum) =30	(non-VI c	man urriculum) =29
How was the doctor at	М	SD	М	SD
1. Making you feel at ease*	2.69	0.21	2.86	0.18
2. Letting you tell your story	2.83	0.17	2.89	0.14
3. Really listening†	2.79	0.17	2.88	0.12
4. Being interested in you as a whole person	2.68	0.20	2.55	0.28
5. Fully understanding your concerns	2.72	0.26	2.74	0.23
6. Showing care and compassion	2.73	0.25	2.76	0.26
7. Being positive‡	2.80	0.19	2.58	0.36

 Table 4 Scoring statements CARE-questionnaire by standardized patients (3-point scale; mean score across six SPs)

Discussion

Our study was designed to test the differences in readiness for clinical practice between graduates from VI and non-VI curricula. Results of earlier studies indicate that the type of curriculum can affect the preparedness for practice (Goldacre, 2003; Watmough et al., 2006; Cave et al., 2009; Wijnen-Meijer et al., 2009, chapter 4; Wijnen-Meijer et al., 2010, *chapter* 5; Wijnen-Meijer *et al.*, accepted for publication, *chapter* 6). All of these studies are questionnaire-based analyses of student or supervisor perceptions. To substantiate these findings, we decided to investigate the actual performance of graduates from VI and non-VI curricula by means of an authentic assessment. In a separate study we provided a validity argument for the UHTRUST assessment procedure. The results of this parallel study indicate that most validity assumptions were defendable with accurate and often parallel lines of backing and that UHTRUST can be used to assess the readiness for clinical practice of medical graduates (Wijnen-Meijer et al., conditionally accepted for publication, *chapter* 9).

Contrary to our expectations, we found only few differences between the two groups of candidates, despite large differences in the curriculum they had followed. One finding is that, based on the judgments of the physicians, the candidates from a VI curriculum scored better on the facet of competence 'active professional development'. The complete description of this FOC is: "The physician aims for quality and professional development by means of a critical attitude towards himself and his environment, study, self-assessment, reflection, asking for feedback and setting and achieving learning goals. She/he reacts to criticism constructively and is aware of his/her own responsibility regarding his/her own abilities". This FOC is relevant for the continuing education of physicians, which requires that medical graduates are capable of setting their own learning goals, receiving feedback and reflection (Fraser & Greenhalgh, 2001).

Next to this, the VI candidates received higher scores for the EPA 'solving a management problem' and for 'being positive' at the CAREquestionnaire for standardized patients. On the other hand, the non-VI candidates were judged better on the FOCs 'teamwork and collegiality' and 'knowing and maintaining own personal bounds and possibilities' by the nurses and on the EPA 'breaking bad news' by the clinicians. They also had higher scores for the aspects 'making you feel at ease' and 'really listening' at the CARE-questionnaire. Because of the amount of tests, the found differences may be due to chance. Reaffirmation of the results in future studies is therefore desirable. Based on the NEO-FFI personality test, the Dutch and German candidates differed on the dimensions 'openness to experience' and 'agreeableness'. Based on a meta-analysis of comparison studies, Barrick *et al.* (2001) concluded that only the dimensions of neuroticism and conscientiousness can be used to predict job performance. Cave *et al.* (2009) found that junior doctors with high scores on the dimensions extraversion and conscientiousness feel better prepared for clinical practice, in contrast to junior doctors with high scores on neuroticism. Based on the results of these studies we assumed that the differences in personality we found between the two groups of candidates did not affect the results of our study.

Our study has a number of limitations. One limitation is that the two groups of candidates are from different countries with, possibly cultural, differences. Another limitation related to this, is the fact that the candidates were assessed by clinicians, nurses and standardized patients from their own country. We had no possibility to employ neutral assessors for all candidates. To reduce this disadvantage, we compared scoring standards of both groups of assessors, by letting them score a Dutch and a German candidate who had been video recorded during a pilot assessment. We found no differences between assessors from different countries. Unfortunately, we could not compare standards among nurses because of the low number of raters. German SPs gave higher scores than Dutch SPs on the videotaped examples, which makes it difficult to interpret the results of these parts of the study.

Finally, we did not have full control over the populations because the candidates signed up for participation voluntarily, which may have affected the results.

Several authors have argued that it is difficult to investigate the impact of different curricula on the readiness for practice of medical graduates by means of experimental research (Albano *et al.*, 1996; Norman, 2003; Albanese, 2009; Cook & Beckman, 2010). One argument for this statement is that the many components of a curriculum form a complex entity, which makes it difficult to determine the influence of the specific features which differ from each other. Additionally, there are many differences in background and characteristics among the students of the same medical school. Nevertheless, we expected to find differences between the performances of the VI and non-VI graduates, because the two curricula differ in many respects from each other. Especially the differences in amount of clinical experience and the responsibility for patient care of final year medical students are believed to have a big impact on the development to junior doctors (Dornan & Bundy, 2004; Ericsson *et al.*, 2006; Cantillon & MacDermott, 2008). In addition, ECFMG figures over the period 2002–2011 give some information to expect a difference, if carefully evaluated.

Despite the limitations, we think that the current study contributes to the existing literature. We spent maximum effort to construct a valid assessment procedure (Wijnen-Meijer *et al.*, conditionally accepted for publication, *chapter* 9) that should allow detecting differences, if they exist. This study has been unable to demonstrate that vertically integrated curricula prepare their graduates better for clinical practice. But there is no reason to jump to the conclusion that the trend towards vertical integration is not the right direction. First, because numerous educational theories about expertise development suggest that a positive effect of vertical integration is to be expected (Vygotsky, 1978; Schmidt *et al.*, 1990; Dreyfus, 2004; Ericsson, 2006; Dunphy & Williamson, 2004). Second, because other studies detected evidence for these theories (Goldacre, 2003; Watmough *et al.*, 2006; Cave *et al.*, 2009; Wijnen-Meijer *et al.*, 2009, *chapter 4*; Wijnen-Meijer *et al.*, 2010, *chapter 5*; Wijnen-Meijer *et al.*, accepted for publication, *chapter 6*).

There are several possible explanations for the fact that we found only few differences between graduates from VI and non-VI curricula. One possibility is that the populations are unequal and not random. In addition, it is possible that despite the preparatory training, assessors were focused on different aspects of the performance because of their personal experiences and background (Van der Vleuten *et al.*, 2007; Yeates *et al.*, 2012). To reduce these differences, the frame-of-reference training (Holmboe & Hawkins, 2008) possibly needs improvement or extension. The differences between the groups of assessors are conceivably strengthened because they are from different countries with different cultures at the clinical workplace. Therefore, similar studies, preferably within one country, need to be undertaken before the difference in performance of VI and non-VI graduates is more clearly understood.



General discussion



Introduction

Context of the thesis

In the first chapter we described the context and motivation for the research project. The purpose of medical education is to equip their trainees to work as physicians in the complex clinical context. Several authors have described the process of expertise development, with attention to the successive phases learners should pass through from novice to expert (Dreyfus & Dreyfus, 1986; Schmidt *et al.*, 1990).

There is a large volume of published studies describing the role of education in stimulating expertise development of learners. Several authors have noted the importance of deliberate practice (Ericsson et al., 1993; Duvivier et al., 2011). To become an expert in a specific domain, skills need to be practiced continually, over a longer period of time and at increasingly challenging levels and supported by feedback (Ericsson *et al.*, 1993; Duvivier *et al.*, 2011). Also other authors have emphasized the importance of learning from experience and assistance by others (Vygotsky 1978; Billet, 1995; Dornan et al., 2007; Tynjälä, 2008). Working together with peers and more competent colleagues makes it possible to learn by sharing information, observing others and performing tasks under supervision (Lave & Wenger, 1991; Dornan, 2007; Tynjälä, 2008; Li et al., 2009). In addition, to facilitate the obtaining of knowledge that is applicable in practice, it is essential that this knowledge is taught in conjunction with realistic situations (Schmidt et al., 1990; Bransford, 1999; Norman, 2009).

Ideally the phases of expertise development of trainees connect with the transitions that medical trainees make during their period of education. These include the transition from theoretical to clinical training and the transition from undergraduate to postgraduate training. Several studies have shown that many trainees do not feel ready for their new role with more responsibilities (Moss & McManus, 1992; Whitehouse *et al.*, 2002; Goldacre, 2003; Lempp, 2004; Van Hell *et al.*, 2008; Cave, 2009; Brennan, 2010).

Vertical integration

To improve the transition from theoretical learning to clinical education and from medical school to postgraduate training, many medical school curricula in Western countries are revised into a more vertically integrated approach. A vertically integrated curriculum can be represented as an "inverted triangle curriculum" or a Z-shape curriculum (Harden *et al.*, 1984; Ten Cate, 2007; Leinster, 2009). The more traditional curriculum, in which theory is programmed in the first years and clinical education in the final years, can be depicted as an H-shape (Ten Cate, 2007, see Figure 1).

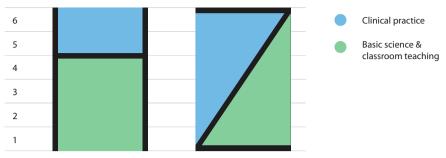


Figure 1 The traditional H-shaped medical curriculum is being replaced by a Z-shaped model

As described in chapter 1, in this thesis a fully vertically integrated curriculum at medical school is defined by the following five features:

- the delivery of basic science teaching in conjunction with a clinical context;
- 2. the occurrence of real patient contacts from the first year of training;
- 3. clinical clerkships from halfway or earlier during the curriculum;
- 4. a progressive increase of clinical responsibility of students over the course of all clerkships and
- 5. long clinical clerkships with an elective nature and substantial responsibility for patient care in the final year of the curriculum.

Vertical integration can improve the readiness for practice of medical graduates in several ways. The connection between basic science and clinical cases stimulates the building of relevant knowledge frameworks (Schmidt *et al.*, 1990; Bransford, 1999; Norman, 2009). Further, the increased role of workplace learning is important. Workplace learning gives trainees the opportunity to develop their competences by means of working under decreasing supervision, practicing clinical skills and experiencing multiple patient cases. As a result of early clinical experience, medical students discover earlier what it means to be a doctor and learn in an early stage to act as a professional (Dornan & Bundy, 2004; Kamalski *et al.*, 2007).

For the development into professionals, both internalization and socialization are needed. Workplace learning plays an important part in both processes. The internalization of knowledge means that various types of knowledge are enclosed in the trainee's knowledge framework. Important for this process of internalization is reflection on prior experiences and adaption of the knowledge framework (Schaap *et al.*,

2011). By means of socialization trainees become acquainted with the common norms and values and with the collective knowledge of a group of professionals. To get socialized, trainees need to participate as full members of a vocational community (Billet, 2001; Schaap *et al.*, 2011).

The progressive increase of clinical responsibility, in combination with proper supervision motivates students for learning. Furthermore, increasing responsibility for trainees during the final phase of medical school makes feel trainees better prepared for postgraduate training and possibly eases the transition (Brennan *et al.* 2010; Cantillon, 2008).

The combination of aspects of vertical integration mentioned above increases the possibilities for both low- and high-road transfer (Salomon & Perkins, 1989) Low-road transfer is the transfer of highly practiced skills almost automatically to a new context. High-road transfer concerns an intentional and conscious abstraction of something from one context and application in a new context. Both types of transfer are needed for medical trainees to become *adaptive experts* (Bereiter & Scardamalia, 2003; Mylopoulos & Regehr, 2009). Adaptive experts are able to adapt their way of working to the ever-changing clinical environment and to cope with unfamiliar medical situations in clinical practice.

Goals of the thesis

The main goal of this thesis was to determine whether a vertically integrated curriculum at medical school enhances the readiness for clinical practice and the transition to postgraduate training. Related goals were:

- Description of the existing structures of medical education around the world, including the transitions to a next phase or level.
- Studying the influence of different types of transitions.
- Description of the development of an assessment instrument to measure the readiness for practice of medical graduates, based on the entrustability approach.

Related to these four goals we formulated research questions which were researched in chapters 2 to 10. The following section of this discussion summarizes the main findings of the research questions, related to the four goals. Next, we will discuss the strengths and limitations of the studies in this thesis. Whereas specific strengths and limitations of the studies have been discussed in the independent chapters, in this chapter we focus on overarching aspects of the thesis in total. Then, we present implications for the practice of medical education. Finally, we recommend suggestions for further research.

Main findings of the thesis

In chapter 2 we give an overview of general models of medical education around the world. We conclude that there are many differences in structure and length of medical training. Also the point that unrestricted practice is allowed and the types of transitions vary between countries. It appears that names of stages and degrees are not very meaningful with regard to the level of the medical trainees. This is even the case for countries that mutually recognize each other's diplomas, such as the countries of the European Union.

In the study described in chapter 3 we focus on the influence of the type of transition on the perceived readiness of medical trainees to be entrusted with professional tasks. In some countries, the completion of medical school corresponds with formal licensure for the medical profession; in other countries (e.g. the UK) these two transitions are separated by a year or more. This situation makes it possible to study the differential influence of a transition to an increased responsibility in comparison with a transition to formal licensure as a physician. We find that the transition to an increased level of responsibility has more impact on the perceived competence development of trainees than the transition to the status of a registered physician. This is the case from the perspectives of both the trainees and the supervisors. It seems that the transition from student to doctor is very important for the expertise development of trainees. This finding may have to do with the fact that after graduation from medical school, medical trainees are part of the community of doctors, which possibly stimulates the trainees' feelings of commitment and the motivation for learning (Ryan & Deci, 2000; Koens *et al.*, 2005). In terms of the theory of communities of practice, finishing medical school possibly stimulates trainees to take the next step towards becoming core participants of the community (Lave & Wenger, 1991; Li *et al.*, 2009).

Another remarkable finding of this study is that the trainees and the supervisors consistently disagree about the amount of necessary supervision, or external regulation, according to Vygotsky's theory about the Zone of Proximal Development (Vygotstky, 1978; Chaiklin, 2003). In comparison with the judgment of the supervisors, the trainees assess themselves as more capable to carry out tasks independently. This difference of opinion between trainees and supervisors may be explained by different frames of reference: supervisors have better understanding of the possible complexities and risks of clinical situations. This interpretation is consistent with the finding that the disagreement between trainees and supervisors about the necessary

amount of supervision is smaller for the more experienced trainees. This finding also accords with the results of a study by Sterkenburg *et al.* (2010). In their study about decisions of supervisors regarding the tasks they entrust to trainees they found discrepancies between what anesthesiology trainees are expected to do as judged by their supervisors, what they actually do and what they think they are able to do.

Chapters 4 to 6 describe studies with regard to the main goal of this thesis: the influence of a vertically integrated curriculum at medical school. Based on educational theories about expertise development and transfer, a positive effect of vertical integration was to be expected (Vygotstky, 1978; Dreyfus & Dreyfus, 1986; Salomon & Perkins, 1989; Schmidt et al., 1990; Ericsson et al., 1993; Dornan & Bundy, 2004; Dunphy & Williamson, 2004; Kamalski *et al.*, 2004; Cantillon, 2008; Norman, 2009; Brennan, 2010). In general, our findings support the idea that vertical integrated curricula affect the readiness for practice and the transition to postgraduate training in a positive way. The results of two questionnaire studies among medical graduates (chapter 4 and 5) indicate that medical graduates who followed vertically integrated curricula at medical school make at an earlier stage their definite choice for a specialization and need less time and fewer applications to obtain positions for postgraduate training programs, in comparison with graduates from non-vertically integrated curricula. In addition, graduates from vertically integrated curricula find themselves better prepared for clinical work and postgraduate training.

Furthermore, also the results of a questionnaire study among supervisors of postgraduate training programs in Utrecht (vertically integrated curriculum at medical school) and Hamburg (non-vertically integrated curriculum) show several differences in performance between the two groups of medical graduates (chapter 6). According to supervisors, graduates from a vertically integrated curriculum are more capable to work independently, to solve medical problems, to manage unfamiliar medical situations, to prioritize their tasks, to collaborate with other people, to estimate when they have to consult their supervisors and to reflect on their activities.

Also the finding that the condition in which trainees get more responsibility affects the competence development positively, can be seen as an underpinning of the effectiveness of vertical integration (chapter 3). As defined in chapter 1, a progressive increase of clinical responsibility is one of the components of a fully vertically integrated undergraduate medical curriculum. Chapters 7, 8 and 9 focus on the development of an authentic performance assessment, called UHTRUST (Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test). Chapter 7 describes a Delphi study among experienced clinical educators. This study resulted in a list of 25 so called facets of competence. The participating experts agreed that these facets of competence are important for the entrustment of critical clinical tasks to starting residents. In addition, they agreed about the formulation of the facets of competence. The consensus we found about important facets of competence of medical graduates can be relevant for clinical practice. For instance, this list can be used for a frame of reference training for clinical educators (Holmboe & Hawkins, 2008), who must frequently make entrustment decisions.

The next step was to investigate whether clinical educators from two different countries (the Netherlands and Germany) agree on the importance of the individual facets of competence by ranking them (chapter 8). An important finding is that there was consensus between the two groups of assessors, despite substantial differences in medical curricula in these two countries. This result suggests that the findings are generalizable to other medical education cultures as well and that it is possible to develop assessments procedures based on the most important facets of competence, which is widely applicable.

In the study described in chapter 9 a validity argument for UHTRUST is written, according the Kane's argument-based approach to validity (Kane, 1992). It is found that most validity assumptions were defendable, so the conclusion is that UHTRUST can be used for the intended purpose: the formative assessment of the readiness for clinical practice of medical graduates.

Chapter 10 describes a study to compare the capability to cope with unfamiliar clinical situations of medical graduates who followed either a vertically integrated or a non-vertically integrated curriculum. Near-graduates from Utrecht/Groningen (vertically integrated) and Hamburg (non-vertically integrated) participated in UHTRUST. Despite the large differences in the curriculum the two groups of graduates had followed, we found only few differences in the assessment scores. The most important finding is that candidates from a vertically integrated curriculum get better scores on the facet of competence 'active professional activities', which includes reflection, asking for feedback and setting learning goals. These features are relevant for the continuing education of physicians (Fraser & Greenhalgh, 2001).

Strengths and limitations

An important strength of this thesis is the combination of research methods, e.g. questionnaires studies, Delphi study, developmental research. Also the fact that the issues of readiness for clinical practice and transitions in medical education are examined from the perspective of both the medical trainees and clinical educators can be considered as a strength. This combination of research methods and perspectives helps to gain a complete view at the topics investigated in this thesis. In addition, the studies described in this thesis contribute both to the theory as the practice of medical education (Albert *et al.*, 2006). Other strong points are the attention for the international context of medical education and the fact that the research project was cross-institutional (Schuwirth et al., 2011). Especially the consensus about the ranking of the facets of competence in different countries and different educational climates (chapter 8) enhances the external validity of UHTRUST. This is also the case for the positive evaluation of the implementation of UHTRUST (chapter 9) by both medical students and supervisors in different countries.

A possible issue for debate is the fact that we used the same implementation of UHTRUST for both collecting evidence for the validity argument and comparing the performances of two groups of medical graduates. In our opinion, the advantage of the combination of gathering evidence for validity and using the data for the intended purpose is that context is the same (chapters 9 and 10). If an instrument is validated in one context and implemented in another context, it may be difficult to determine whether the instrument is also valid in this context (Schuwirth *et al.*, 2011).

One limitation of this thesis is that in the comparison studies (chapters 4, 5, 6 and 10) just a part of the total population was involved and that all participants participated voluntarily, which is a possible threat for internal validity. Due to the voluntary participation it is possible that the sample consisted especially of a specific group of trainees and supervisors. In all studies this was the case for both groups and therefore we assume that a possible bias stemming from this situation was the same for both groups and that this has no effect on the differences found between the groups. The fact that we asked medical graduates about their preparedness for work and postgraduate training in retrospect could possibly affect the internal validity of the study (chapters 4 and 5). Again this was the same for both groups. Therefore, we assume that this situation did not influence the findings of these studies.

In addition, the fact that the near-graduates who participated in UHTRUST (chapter 10) were assessed by clinicians, nurses and standardized patients from their own country (i.e. the Netherlands and Germany) is a limitation. Unfortunately we had no possibility to employ independent assessors for all candidates, because of the differences in language and the distance between the two locations. To reduce this weakness, we compared scoring standards of both groups of assessors, by letting them judge a Dutch and a German candidate who had been video recorded during a pilot assessment. These checks provide useful information for interpreting the results of the study.

Another limitation is that we only studied the effect of a fully vertically integrated curriculum in total (chapters 4, 5, 6 and 10). As a result we cannot say what the effect is of the separate components of vertical integration (Norman, 2003). On the other hand, the study described in chapter 3 provides useful information about the possible influence of an increase in responsibility on the competence development of medical trainees.

A final limitation of this thesis is that most studies are based on opinions and experiences of (groups) of individuals, e.g. medical graduates, clinical educators and standardized patients. Because the essence of clinical practice lies in working with patients and colleagues, it is not possible to measure this complex construct of readiness for clinical practice without subjective judgments. Although the judgments of experts will probably be biased because of the personal experiences and characteristics of the assessors, they are able to assess the performance of medical trainees precisely because of their experiences (Govaerts *et al.*, 2007; Bakker, 2008; Sterkenburg *et al.*, 2010). In addition, as known from published research, assessments based on subjectivity are not automatically unreliable (Van der Vleuten *et al.*, 2010). In the studies described in this paper we also found mostly acceptable reliability and agreement measures (chapters 7–10).

Suggestions for the practice of medical education

The results of the studies described in this thesis lead to suggestions for medical education in general.

Vertical integration of undergraduate medical curricula seems to have a positive influence on the readiness for clinical practice of medical trainees, although not all studies confirm this statement as evident as we had expected. Both the medical graduates and supervisors of postgraduate training programs indicate that vertically integrated curricula prepared the students better for working as physicians (chapters 5 and 6).

Looking at our definition of vertical integration, especially two features are unusual in current medical education and deserve more attention from curriculum developers, i.e. 'a progressive increase of clinical responsibility of students over the course of all clerkships' (feature 4) and 'long clinical clerkships with an elective nature and substantial responsibility for patient care in the final year of the curriculum' (feature 5). As shown by the results of the study in chapter 3, an increase in responsibility can stimulate the competence development of trainees.

Related to the amount of responsibility of trainees, is the recurring difficulty in medical education of monitoring the balance between giving the trainees responsibility. Supervisors have the tasks to stimulate the development of their competences and, at the same time, to deliver safe patient care. Supervisors have to make continual judgments about the amount of supervision their trainees need to carry out specific tasks in the clinical practice. Therefore, the use of Entrustable Professional Activities may be helpful for the development of a competency based medical curriculum (Ten Cate, 2005; Ten Cate & Scheele, 2007; Boyce *et al.*, 2011; Jones, 2011).

Also the results of the Delphi study described in chapter 7 can be useful for the practice of medical education. The identified features of medical trainees that enable supervisors to entrust the trainees with clinical tasks, can be used for the development of assessment methods or the training of clinical supervisors.

To conclude, the design of UHTRUST can be implemented in medical education, for instance for the purpose of residency selection for postgraduate training programs. It is possible to adapt the design for specific medical specialisms by changing the content of the patient scenarios and disturbances. Further exploration of the possible implementations of UHTRUST, including conditions and consequences for the students and the organization, is recommended.

Implications for further research

The results presented in this thesis lead to suggestions for further research on this topic.

First, a complete picture of the readiness for clinical practice of specific trainees who followed different types of curricula is still missing. Therefore future research needs to make a direct comparison between the sense of preparedness for clinical practice of a trainee, the assessment by the supervisor about particular trainees and their actual performance in clinical practice.

Second, more research is needed to identify the impact of the separate features of a fully vertically integrated curriculum on the competency development of trainees, also in relation to patient safety.

Next, further research is needed into the specification under which these curriculum effects occur. Besides more research into vertical integrated curricula, a further study with more focus on the effects of longitudinal integrated clerkships is suggested.

In addition, an extension of the evaluation of the identified facets of competence in other countries and different educational climates will be important for the purpose of generalizability of our findings.

Recently, there is increasing attention for competency-based education and the approach of entrustability (Carraccio & Burke, 2010; Frank *et al.*, 2010; Ten Cate *et al.*, 2010; Boyce *et al.*, 2011 Jones *et al.* 2011), which probably leads to a growing demand for assessment procedures that fit into this development. Therefore, further research into UHTRUST is desirable. For developmental research it is recommended that a series of studies are conducted to improve the design and make is useful for educational practice (Schuwirth, 2011). If the assessment procedure of UHTRUST will be used in other contexts or for other purposes (for instance, for the selection of residencies) it is needed that it will be validated again for these specific contexts and purposes.

To conclude, to confirm the results we found, UHTRUST should be used again to compare the readiness for practice of graduates from vertically integrated and non-vertically integrated curricula, preferably within one country to reduce cultural bias.

In conclusion

The research project described in this thesis aims to contribute to the literature about the influence of vertically integrated curricula and different types of transitions on the readiness for clinical practice of medical graduates. Several studies in this thesis improve the understanding of these topics. In addition, this thesis provides information about the existing structures of medical education around the world. Finally, we developed a new assessment instrument to measure the readiness for clinical practice. To validate this instrument for the intended purpose, evidence was collected to support the arguments for validity.



Summary

Summary

The introduction to this thesis in chapter 1 discusses the context of the research project. It summarizes what is already known about expertise development in medical education. Also the difficulties around transitions in medical education are explored, related to theories about transfer of learning. The next section of the general introduction deals with several methods to measure the readiness for practice and different types of curricula at medical schools, i.e. discipline based, horizontal integrated and vertically integrated curricula. The expected influence of these types of curricula on the readiness for clinical practice is elaborated.

Zusammenfassung

In der Einleitung zu dieser Arbeit wird in Kapitel 1 der Kontext des Forschungsproiektes dargestellt. Die Einleitung beginnt mit einem Fachliteraturüberblick über die jüngsten Entwicklung im Bereich der medizinischen Ausbildung. Die Schwierigkeiten bei den Übergängen in der medizinischen Ausbildung werden ebenfalls exploriert, auch in Bezug auf Theorien zum Transfer von Gelerntem. Der nächste Abschnitt der allgemeinen Einleitung beschäftigt sich mit verschiedenen Methoden, um die Praxisreife und verschiedene Arten von medizinischen Curricula zu messen, z.B. Fächer-basierte, horizontal oder vertikal integrierte Curricula. Der vermutete Einfluss dieser Arten von Curricula auf die Praxisreife wird ausgeführt.

Ein vollkommen vertikal integriertes Curriculum für das Medizinstudium wird in Kapitel 1 folgender Maßen definiert:

(1) die Vermittlung von naturwissenschaftlichen Grundlagen im klinischem Kontext;

Samenvatting

In hoofdstuk 1, de inleiding op dit proefschrift, wordt ingegaan op de context van het onderzoeksproject. Er wordt samengevat wat er tot op heden bekend is over de ontwikkeling van expertise gedurende de medische opleiding. Ook de problemen rond transities in medisch onderwiis worden verkend, in relatie tot theorieën over transfer van leren. In het volgende deel van de algemene inleiding wordt ingegaan op verschillende methoden om te bepalen in hoeverre aeneeskundestudenten ziin voorbereid op het werken in de praktijk. Ook worden verschillende soorten curricula van aeneeskunde-opleidinaen besproken, zoals horizontaal en verticaal geïntegreerde curricula en curricula geordend op basis van vakgebieden (discipline georiënteerd). Vervolgens wordt het te verwachten effect van de verschillende soorten curricula op de mate van voorbereiding op (het werken in) de klinische praktijk in kaart gebracht.

Een volledig verticaal geïntegreerd curriculum van de opleiding geneeskunde (leiA fully vertically integrated undergraduate medical curriculum is defined in chapter 1 as follows:

- the delivery of basic science teaching in conjunction with a clinical context;
- 2. the occurrence of real patient contacts from the first year of training;
- 3. clinical clerkships from halfway or earlier during the curriculum;
- 4. a progressive increase of clinical responsibility of students over the course of all clerkships and
- 5. long clinical clerkships with an elective nature and substantial responsibility for patient care in the final year of the curriculum.

(2) das Ermöglichen von echten Patientenkontakten vom ersten Jahr des Studiums an;

(3) klinische Famulaturen ab der zweiten Hälfte des Curriculums oder früher;

(4) eine schrittweise Zunahme von klinischer Verantwortung durch Studierende während des Verlaufs aller Famulaturen und

(5) lange klinische Famulaturen mit einem Wahlfach-Charakter und substantieller Verantwortung für Patienten im letzten Studienjahr des Curriculums.

Außerdem beschreibt das erste Kapitel die Ziele und die Forschungsfragen für diese Arbeit. Das Hauptziel ist, festzustellen, ob ein vertikal integriertes Curriculum im Medizinstudium den Übergang in die ärztliche Weiterbildung verbessert. Verknüpft mit diesem Hauptziel sind drei andere Themen. Erstens beschreibt diese Arbeit die bestehenden curricularen Strukturen des Medizinstudiums in aller Welt. Zweitens wird der Einfluss von verschiedenen Übergängen untersucht. Weiterhin beschreiben wir ein Prüfungsinstrument, dass wir entwickelt haben, um die Praxisreife von dend tot het artsdiploma) wordt in hoofdstuk 1 als volgt gedefinieerd:

(1) het aanbieden van biomedische kennis gekoppeld aan een klinische context;

(2) authentiek patiëntencontact vanaf het eerste jaar van de opleiding;

(3) klinische coschappen startend halverwege het curriculum of eerder;

(4) een voortschrijdende toename van klinische verantwoordelijkheid van studenten in de loop van alle coschappen;

(5) lange, klinische keuzecoschappen en substantiële verantwoordelijkheid voor patiëntenzorg in het laatste jaar van het curriculum.

Daarnaast worden in hoofdstuk 1 de doelen en de onderzoeksvragen van het proefschrift beschreven. Het hoofddoel is vaststellen of een verticaal geïntegreerd curriculum in de geneeskunde-opleiding de transitie naar de vervolgopleiding vergemakkelijkt. Gerelateerd aan dit hoofddoel gaat dit proefschrift in op drie andere onderwerpen. Ten eerste beschrijft het proefschrift de verschillende structuren van medisch onderwijs die wereldwijd bestaan, Further, the first chapter describes the goals and research questions of the thesis. The main goal is to determine whether a vertically integrated curriculum at medical school enhances the transition to postgraduate training. Related to this main goal this thesis addresses three other issues. First, this thesis describes the existing structures of medical education around the world, including the transitions between successive phases. Second, the influence of different types of transitions was studied. Further, we describe an assessment instrument that we developed to determine the readiness for practice of medical graduates, based on the entrustability approach.

In connection with these four goals, the following research questions are addressed in this thesis:

Absolventen eines Medizinstudiums basierend auf einem Zutrauensansatz.

In Verbindung mit diesen vier Zielen wurden die folgenden Forschungsfragen gestellt:

Strukturen von medizinischen Curricula in aller Welt

• Welche Modelle für das Medizinstudium gibt es, inklusive der Übergänge zu einer nächsten Phase oder Stufe?

• Zu welchem Zeitpunkt wird die uneingeschränkte Tätigkeit erlaubt?

• Welcher akademische Abschluss wird am Ende des Medizinstudiums gewährt?

Einfluss von verschiedenen Arten von Übergängen bezüglich der wahrgenommenen Kompetenzentwicklung und der Praxisreife

• Wie schätzen Personen in der ärztlichen Aus- und Weiterbildung ihre Praxisreife bezüglich ausgewählter klinischer Aufgaben in drei Phasen der Ausbildung ein?

•Wie schätzen klinische Vorgesetzte die Praxisreife von Personen in der ärztlichen Aus- und Weiterbildung bezüglich ausgeinclusief de transities tussen de opeenvolgende fasen. Ten tweede werd de invloed van verschillende soorten transities onderzocht. Tot slot beschrijven we een assessmentinstrument dat we ontwikkelden om in kaart te brengen in hoeverre basisartsen voorbereid zijn op het werken in de praktijk. Dit instrument is gebaseerd op het idee dat supervisoren gefundeerd moeten beslissen in hoeverre ze bepaalde activiteiten kunnen toevertrouwen aan artsen in opleiding.

In aansluiting op deze vier doelen worden in dit proefschrift de volgende onderzoeksvragen beantwoord:

Structuren van medisch onderwijs wereldwijd

• Welke modellen, inclusief transities naar een volgende fase of een volgend niveau, bestaan er binnen het medisch onderwijs?

- Op welk moment is zelfstandige beroepsuitoefening toegestaan?
- Welke titels worden toegekend na de geneeskunde-opleiding?

Structures of medical education around the world:

- What are existing models of medical education, including transitions to a next phase or level?
- > At which moment is unrestricted practice allowed?
- Which degrees are granted after medical school?

Influence of different types of transition on perceived competence development and readiness for practice:

- How do medical trainees evaluate their readiness to be entrusted with selected clinical tasks at three stages of training?
- How do clinical supervisors evaluate the readiness to be entrusted with selected clinical tasks of medical trainees at three stages of training?

wählter klinischer Aufgaben in drei Phasen der Ausbildung ein?

•Welchen unterschiedlichen Einfluss haben eine Verantwortungsübergabe und eine Examensübergabe auf die Kompetenz-Selbsteinschätzung und auf die Wahrnehmung von Vorgesetzen?

Einfluss eines vertikal integrierten Curriculums im Medizinstudium auf den Übergang zur ärztlichen Weiterbildung im Vergleich mit nicht vertikal integrierten Curricula

• In welchem Zeitraum wurde die definitive Fachrichtung für die weitere Karriere gewählt?

• Welchen Einfluss hatten die Erfahrungen während des letzten Jahres des Medizinstudiums auf den Prozess der Karriereentscheidung?

• Wie lang war der Zeitraum zwischen Abschluss des Medizinstudiums und Beginn der ärztlichen Weiterbildung?

• Wie viele Bewerbungen waren erforderlich um eine ärztliche Weiterbildungsstelle zu erhalten?

• Wie ist die Gesamtzufriedenheit von Absolventen eines Medizinstudiums mit ihrer

Invloed van verschillende soorten transities op de ervaren competentieontwikkeling en de mate waarin artsen in opleiding toegerust zijn voor de praktijk

 In welke mate kunnen medische studenten/artsen in drie opeenvolgende fasen van de opleiding geselecteerde activiteiten worden toevertrouwd, volgens henzelf?

 In welke mate kunnen studenten/artsen in drie opeenvolgende fasen van de opleiding geselecteerde activiteiten worden toevertrouwd, volgens hun supervisoren?

• Wat is het verschil in effect tussen een transitie in verantwoordelijkheid en een transitie in formele licentiestatus op de competentie van artsen in opleiding, volgens henzelf en hun supervisoren?

Invloed van verticaal geïntegreerde curricula in de geneeskunde-opleiding op de overgang naar de vervolgopleiding, in vergelijking met niet-verticaal geïntegreerde curricula

• In welke periode werd de definitieve keuze voor een specialisatie gemaakt?

• Wat was de invloed van ervaringen in het laatste jaar van de geneeskunde-opleiding op het proces van het maken van een carrièrekeuze? What is the differential impact of a responsibility transition and a licensure transition on self-perceived competence and supervisor perceptions?

Influence of vertically integrated curricula at medical school on the transition to postgraduate training, by means of comparison with non-vertically integrated curricula:

- During which time period was the definite specialty career choice made?
- What was the influence of experiences in the final year of medical school on the process of making a career choice?
- How long was the period between graduating medical school and starting residency?
- How many applications were required to obtain a residency position?

Karriere seit dem Abschluss des Medizinstudiums?

• In welchem Ausmaß fühlen sich Absolventen eines Medizinstudiums auf ihre Tätigkeit und auf die ärztliche Weiterbildung nach Ende des Medizinstudiums vorbereitet?

• Welche Unterschiede bestehen zwischen Absolventen von Medizinstudiengängen mit vertikal integriertem bzw. nicht integriertem Curriculum aus der Sicht von Vorgesetzten?

• Welche Unterschiede bestehen zwischen Absolventen von Medizinstudiengängen mit vertikal integriertem bzw. nicht integriertem Curriculum im Hinblick auf deren Kompetenz zur Bewältigung unbekannter klinischer Situationen aus der Sicht von Vorgesetzten?

Entwicklung einer authentischen Leistungsprüfung zur Messung der klinischen Praxisreife von Absolventen eines Medizinstudiums

• Welche Facetten klinischer Kompetenz halten klinische Vorgesetzte für essentiell um einen Arzt in der Weiterbildung mit kritischen klinischen Aufgaben zu betrauen? • Hoe lang was de periode tussen het afstuderen en de start van de medische vervolgopleiding?

• Hoe veel sollicitaties waren er nodig om een plaats in de vervolgopleiding te verkrijgen?

 Hoe is de algemene tevredenheid van basisartsen met hun carrière na het afronden van de initiële geneeskunde-opleiding?

• In welke mate voelen afgestudeerden zich voorbereid op het werk en de vervolgopleiding na het afronden van de initiële geneeskunde-opleiding?

• Welke verschillen zijn er met betrekking tot het toegerust zijn voor het werk, medische kennis en de bekwaamheid om om te gaan met belangrijke, specifieke onderdelen van het werk als arts, tussen basisartsen die een geneeskunde-opleiding met een verticaal geïntegreerd curriculum hebben gevolgd en basisartsen die een niet-verticaal geïntegreerd curriculum hebben gevolgd, volgens supervisoren in de vervolgopleidingen?

• Wat zijn de verschillen in competentie tussen basisartsen die een verticaal geïntegreerd geneeskunde curriculum hebben

- How generally satisfied do graduates feel with their careers since leaving medical school?
- To what extent do graduates feel prepared for work and postgraduate training after finishing medical school?
- What are the differences regarding preparedness for work, medical knowledge and capabilities to manage important specific parts of the work as a physician between starting residents graduated from a medical school with a vertically integrated curriculum and from a medical school with a non-vertically curriculum, according to supervisors in postgraduate training programs?
- What are the differences between graduates from a medical school with a vertically integrated curriculum and from a medical school with a non-vertically integrated curriculum in their competences to cope with unfamiliar clinical situations?

• Wie gut ist die Übereinstimmung bezüglich dieser Facetten zwischen klinischen Vorgesetzten aus verschiedenen Ausbildungsmodellen?

• Wie kann Validität in der Leistungsprüfung für Praxisreife erreicht werden?

Um diese Forschungsfragen zu beantworten, wurde eine Vielfalt von Methoden verwendet, insbesondere quantitative und qualitative Fragebogen, Dokumentanalysen und Design-basierter Experimenten.

Kapitel 2 beschäftigt sich mit verschiedenen Modellen der medizinischen Ausbildung in aller Welt. In einer Welt, die zunehmend dem internationalen Austausch von Informationen über medizinische Ausbildung dient, erfahren Studierende, Ärzte und Ausbilder sind einer grossen Anzahl von oft widersprechenden Varianten von medizinischen Ausbildungsmöglichkeiten ausgesetzt.

Studierende, Ärzte und Lehrende viele, oft verwirrende Varianten von medizinischen Ausbildungsmöglichkeiten. Um eine Übersicht über Kursstrukturen und Terminologie in medizinischer Ausbildung zu ermöggevolgd en basisartsen die een niet-verticaal geïntegreerd curriculum hebben gevolgd met betrekking tot het omgaan met onbekende klinische situaties?

Ontwikkeling van een authentiek performance assessment om in kaart te brengen in hoeverre afgestudeerde basisartsen toegerust zijn voor het werk in de klinische praktijk

- Wat zijn volgens ervaren klinische opleiders essentiële competentiefacetten die bepalend zijn voor de beslissing om een arts in opleiding kritische klinische taken toe te vertrouwen?
- In hoeverre is er overeenstemming in de oordelen over deze competentiefacetten tussen klinische opleiders in verschillende opleidingsklimaten?
- Hoe kan een valide beoordeling van de mate van toegerustheid voor de praktijk worden bereikt?

Om deze onderzoeksvragen te kunnen beantwoorden zijn verschillende methoden gebruikt, namelijk kwantitatieve en kwalitatieve vragenlijstonderzoeken, documentanalyses en onderwijskundig ontwikkelingsonderzoek. Development of an authentic performance assessment to measure the readiness for clinical practice of medical graduates:

- What do experienced clinical educators consider as essential facets of competence that determine decisions to entrust a trainee with critical clinical tasks?
- How well do clinical educators in different educational climates agree on these facets?
- How can validity in the assessment of readiness for practice be attained?

To answer these research questions, a variety of methods are used, namely quantitative and qualitative questionnaire studies, document analyses and design-based research.

lichen, führten wir eine Fragebogenstudie unter Experten in 40 Ländern durch. Thematisiert wurde: Struktur der medizinischen Ausbildung inklusive Abschnitte und Übergänge, Zeitpunkt der unabhängigen Ausübung ärztlicher Tätigkeit, Namen der Abschlüsse, die vergeben werden, und relevante Terminologie für das medizinische Ausbildungssystem. Außerdem suchten wir nach Literatur zur Beschreibung von Länder-spezifischen Informationen. Die Ergebnisse dieser Studie zeigen, dass eine Vielzahl von Unterschieden zwischen den Ländern besteht. Eine weitere wichtige Schlussfolgerung ist, dass die Namen der Ausbildungsabschnitte und der vergebenen Abschlüsse nicht sehr informativ sind bezüglich des Inhalts der genossenen Ausbildung und bezüglich des Niveaus des medizinischen Absolventen. Weltweit bestehen sechs unterschiedliche Modelle. Bei detaillierterer Betrachtung variieren die Modelle in ihrer Struktur, der Länge der Ausbildungsphasen und bezüglich des Zeitpunkts für die Zulassung als Arzt.

In Kapitel 3 werden zwei Übergänge und dessen Auswirkung auf die wahrgenom-

De focus in hoofdstuk 2 ligt op verschillende modellen van medisch onderwijs wereldwijd. Er vindt steeds meer internationale uitwisseling plaats van informatie, medische studenten en artsen. Hierbij hebben studenten, artsen en opleiders reaelmatig te kampen met, vaak verwarrende, verschillen in het medisch onderwijs. Om een overzicht te kunnen opstellen van de structuur van het medisch onderwijs en de terminologie die daarbii wordt gebruikt. voerden we een vragenlijstonderzoek uit onder deskundigen uit 40 landen. Onderwerpen van de vragenlijst waren de structuur van het medisch onderwijs, inclusief fasen en transities, het moment waarop zelfstandige beroepsuitoefening is toegestaan, namen van titels die worden toegekend en relevante terminologie in het medisch onderwijssysteem. Daarnaast zochten we in de literatuur naar beschriivingen van systemen van medisch onderwijs in verschillende landen. Uit de resultaten van dit onderzoek blijkt dat er veel verschillen zijn tussen landen. Een andere belangrijke bevinding is dat de namen van fasen en titels weinig informatief zijn met betrekking tot de gevolgde opleiding en Chapter 2 focuses on different models of medical education around the world. In a world that increasingly serves the international exchange of information on medical training, many students, physicians and educators encounter numerous, often confusing, variations of medical training. To provide an overview of course structures and terminology occurring in medical education, we conducted a questionnaire study among representatives of 40 countries. Topics of the questionnaire were: structure of medical education, including stages and transitions, moment that unrestricted practice is allowed, names of degrees that are granted and relevant terminology of the medical education system. In addition, we searched the literature for descriptions of countryspecific information. The results of this study make clear that there are many differences between countries. Another important conclusion is that names of stages and degrees are not very informative about the education received and the level of the medical trainee. There appear to be six general models of medical education around the world. At more

mene Kompetenzentwicklung erforscht. Der erste Übergang betrifft die Zulassung und der zweite den Zuwachs von Verantwortung. Die Untersuchung wurde in Großbritannien durchgeführt, wo diese beiden Übergänge ein Jahr auseinander liegen. Der Übergang vom letzten Jahr des Medizinstudiums zum ersten Jahr in einem ärztlichen Weiterbildungsprogramm (FY1) ist ein Verantwortungsübergang. Der Übergang vom ersten zum zweiten Jahr im ärztlichen Weiterbildungsprogramm (FY1 zu FY2) basiert auf einer formalen Zulassung, d.h. auf der vollen Beurkundung beim General Medical Council. Wir führten eine Fragebogenstudie unter Studierenden im letzten Jahr ihres Medizinstudiums an der Medical School der Universität Leeds und mit Teilnehmern im ersten und zweiten Jahr der ärztlichen Weiterbildung an der Leeds Foundation Training School und deren Supervisoren durch. Der Fragebogen bestand aus 16 sogenannten "Entrustable Professional Activities" (EPAs), um das Kompetenzniveau in verschiedenen Stadien der Ausbildung zu messen, wie es von Lernenden und Supervisoren wahrgenommen wird. Wie erwartet waren het niveau van de arts in opleiding. Er blijken wereldwijd zes algemene modellen te bestaan waarmee het medisch onderwijs kan worden beschreven. Op een meer gedetailleerd niveau variëren deze modellen in structuur, lengte van de fasen van de opleiding en moment van registratie als arts.

Het doel van hoofdstuk 3 was het onderzoeken van de invloed van zowel transities in formele bevoegdheid (registratie) als transities in de mate van verantwoordelijkheid op de ervaren competentieontwikkelina. Dit onderzoek werd uitaevoerd in het Verenigd Koninkrijk, waar een jaar tussen deze transities zit. De overaana van het laatste jaar van de geneeskundeopleiding naar het eerste jaar van het 'foundation programme' (tweejarige opleiding tussen de initiële geneeskunde-opleiding en de opleiding tot medisch specialist) is een transitie in verantwoordelijkheid. De transitie van het eerste naar het tweede jaar van het foundation program (FY1 naar FY2) is gebaseerd op het behalen van een formele licentie, namelijk volledige registratie bij de 'General Medical Council'. We voerden een vragenlijstonderzoek uit onder studenten detailed level the models vary in structure and length of the phase of training and point of full registration as a physician.

Chapter 3 aimed to investigate how transitions in licensure and increased responsibility affect the perceived competence development. This study was carried out in the UK, where these two transitions are separated by a year. The transition from the final year of medical school to the first year of the foundation program (FY1) is a responsibility transition. The transition from the first to the second year of the foundation program (FY1 to FY2) is based in formal licensure, i.e. full registration at the General Medical Council. We conducted a questionnaire study among students of the final year of Medical School of the University of Leeds, first and second year trainees of the Leeds Foundation Training School and their supervisors. The questionnaire consisted of 16 Entrustable Professional Activities (EPAS), to determine the competency levels

die mittleren wahrgenommenen Kompetenzwerte am niedrigsten für Medizinstudierende im letzten Studienjahr und am höchsten für FY2 Teilnehmer, was nahe legt, dass die Kompetenz von Auszubildenden insgesamt als mit der Zeit steigend angesehen wird. Außerdem fanden wir für alle Niveaus, das die mittleren Kompetenzwerte, die von den Lernenden vergeben wurden, signifikant über denen lagen, die von Supervisoren vergeben wurden. Die Unterschiede zwischen den Bewertungen der Lernenden und der Supervisoren waren am kleinsten für das Ausbildungsjahr FY2. Basierend auf den Wahrnehmungen von Lernenden und Supervisoren ist die Auswirkung des Verantwortungsübergangs, also vom Medizinstudium zu FY1, bezogen auf die Kompetenzentwicklung größer als beim Übergang zur formalen Zulassung, also von FY1 zu FY2. Damit scheint der Übergang zur formalen Übernahme von Verantwortung mit der Zulassung zur eigenständigen medizinischen Tätigkeit eine geringere Bedeutung zu haben, für das Zutrauen von Lernenden und ihren Supervisoren in die Ausführung von kritischen klinischen Aktivitäten, als der in het laatste jaar van de geneeskundeopleiding van de Universiteit van Leeds, eerste en tweedejaars artsen in opleiding van de Leeds Foundation Training School en hun supervisoren. De vragenlijst bestond uit 16 'Entrustable Professional Activities' (EPAs), om het competentieniveau van artsen in verschillende stadia van de opleiding, volgens henzelf en hun supervisoren in kaart te brengen. Zoals verwacht waren de aemiddelde toeaekende competentiescores voor zowel de klinische als de meer algemene activiteiten het laagste voor de studenten in het laatste jaar van de geneeskunde-opleiding en het hoogste voor FY2 trainees. Dit wiist erop dat het algemene competentieniveau van de artsen in opleiding in de loop van de tijd toeneemt. Daarnaast bleek dat de gemiddelde scores toegekend door artsen in opleiding voor alle niveaus significant hoger waren dan de gemiddelde scores toegekend door supervisoren. De verschillen tussen de scores van trainees en supervisoren waren het kleinste voor de FY₂ studenten. Op basis van de inschatting van zowel artsen in opleiding als supervisoren, is het effect van een transitie in verantwoordelijkheid op de at different stages of training, perceived by trainees and supervisors. As was expected, the mean perceived competence scores for both clinical activities and general activities were lowest for final year medical school students and the highest for FY2 trainees, suggesting that the trainee competence in general is perceived as increasing over time. In addition we found that for all levels the mean scores given by trainees were significantly higher than the mean scores given by supervisors. The differences between the scores of trainees and supervisors are the smallest for the FY2 trainees. Based on the perceptions of both trainees and supervisors, the impact of a responsibility transition, i.e. from medical school to FY1, on the competence development is larger than that of a licensure transition, i.e. from FY1 to FY2. This appears to suggest that the transition to formal responsibility with a license to practice medicine seems to have less impact on confidence of trainees and their supervisors to ex-

Übergang von informeller, aber signifikant höherer Verantwortung.

In Kapitel 4 und 5 werden zwei Fragebogenstudien unter Absolventen eines Medizinstudiums beschrieben. Ziel dieser Studien war es herauszufinden, ob ein vertikal integriertes Curriculum im Medizinstudium den Übergang zur ärztlichen Weiterbildung in positiver Weise beeinflussen kann. Die in Kapitel 4 erläuterte Studie vergleicht die Blickwinkel von Absolventen zweier aufeinanderfolgender Studierendenkohohrten an der Utrecht Medical School. Die Absolventen der einen Kohorte wurden in einem vertikal integrierten Curriculum unterrichtet und die Absolventen der anderen Kohorte folgten einem traditionellen Curriculum. Themenfelder des Fragebogens waren: (1) Aktivitäten seit Abschluss des Medizinstudiums, (2) benötigte Zeit für und Anzahl von Bewerbungen, um einen Platz für die ärztliche Weiterbildung zu erhalten, (3) der Prozess, um eine Karriereentscheidung zu treffen und (4) die Gesamtzufriedenheit mit der eigenen Karriere seit Abschluss des Medizinstudiums. Wir fanden, dass im Vergleich competentieontwikkeling, zoals van de initiële geneeskundeopleiding naar FY1, groter dan die van een transitie in de vorm van een formele registratie, zoals van FY1 naar FY2. Dit lijkt erop te wijzen dat de transitie naar formele verantwoordelijkheid op basis van een licentie minder invloed heeft op het vertrouwen van artsen in opleiding en supervisoren dat de artsen in opleiding kritische taken zelfstandig kunnen uitvoeren, dan een transitie naar informele, maar significant grotere verantwoordelijkheid.

In de hoofdstukken 4 en 5 worden twee vragenlijstonderzoeken onder basisartsen beschreven. Het doel van deze onderzoeken was achterhalen of een verticaal geintegreerd curriculum in de geneeskundeopleiding een positieve invloed heeft op de transitie naar de (medische) vervolgopleiding. In het onderzoek beschreven in hoofdstuk 4 werden de perspectieven van afgestudeerden van twee opeenvolgende cohorten van het UMC Utrecht vergeleken. De afgestudeerden van het ene cohort volgden een verticaal geïntegreerd curriculum en die van het andere cohort volgden een traditioneel curriculum. Onderwerpen ecute critical activities than the transition to informal but significantly higher responsibility.

In Chapter 4 and 5 two questionnaire studies among medical graduates are described. The aim of these studies was to find out whether a vertically integrated curriculum at medical school affects the transition to postgraduate training in a positive way. The study described in chapter 4 compared the perspectives of graduates of two consecutive cohorts of the Utrecht Medical School. The graduates of one cohort followed a vertically integrated curriculum and the other graduates a traditional curriculum. Topics of the questionnaire were: (1) activities since medical school, (2) required amount of time and number of applications to get admission to residency, (3) the process of making career choices and (4) general satisfaction with their careers since medical school. We found that, compared with graduates from a traditional curriculum, graduates

mit Absolventen eines traditionellen Curriculums die Absolventen, die in einem vertikal integrierten Curriculum ausgebildet worden waren, ihre definitive Karriereentscheidung früher getroffen hatten. Außerdem benötigten sie weniger Zeit und eine geringere Anzahl von Bewerbungen, um eine Stelle für die ärztliche Weiterbildung zu erhalten. Wir fanden keine statistisch signifikanten Unterschiede bezüglich der Zufriedenheit mit der eigenen Karriere.

In Kapitel 5 wurden die Antworten von Absolventen eines Medizinstudiums derselben Kohorte an sechs verschiedenen medizinischen Hochschulen, die entweder in einem vertikal integrierten Curriculum oder in einem vertikal nicht integrierten Curriculum studiert hatten, miteinander verglichen. Wir fanden die gleichen Ergebnisse wie in der früheren, in Kapitel 4, dargestellten Studie. Die Studienteilnehmer wurden außerdem gefragt, bis zu welchem Grad ihr Medizinstudium sie für ihre Arbeit als Ärzte in der ärztlichen Weiterbildung vorbereitet hat. Die Absolventen, die einem vertikal integrierten Curriculum gefolgt waren, äußerten sich positiver im

van de vragenlijst waren: (1) activiteiten sinds (de afronding van) de geneeskundeopleiding, (2) de hoeveelheid tijd en het aantal sollicitaties die basisarts nodig had om een opleidingsplaats in een vervolgopleidina te verkrijaen, (3) het carrièrekeuzeproces en (4) de algemene tevredenheid met de eigen carrière na het afronden van de geneeskunde-opleiding. Basisartsen die een verticaal geïntegreerd curriculum hadden aevolad bleken, in veraeliikina met basisartsen die een traditioneel curriculum hadden gevolgd, eerder een definitieve carrièrekeuze te hebben gemaakt. Daarnaast hadden zij minder tijd en minder sollicitaties nodig om een opleidingsplaats te verkrijgen. We vonden geen significante verschillen met betrekking tot de tevredenheid over de eiaen carrière.

In het onderzoek waarover wordt gerapporteerd in hoofdstuk 5, vergeleken we de antwoorden van basisartsen uit hetzelfde cohort van zes verschillende geneeskunde-opleidingen, die òf een verticaal geïntegreerd curriculum òf een niet-verticaal geïntegreerd curriculum hadden doorlopen. We vonden dezelfde resultaten als who followed a vertically integrated curriculum had made their definite career choices earlier. In addition, they needed less time and fewer applications to obtain residency positions. We did find not statistically significant differences regarding the satisfaction about their careers.

In the study presented in chapter 5, we compared the answers of medical graduates of the same cohort from six different medical schools, who had followed either a vertically integrated curriculum or a non-vertically integrated curriculum. We found the same results as in the earlier study presented in chapter 4. The respondents were also asked to what extent their training at medical school had prepared them for work as a physician and postgraduate training. The graduates who had followed a vertically integrated curriculum were more positive about their preparedness.

The aim of the study described in chapter 6 was to determine whether medical graduates who followed vertically integrated curricula meet

Hinblick auf ihre Vorbereitung auf die ärztliche Tätigkeit.

In Kapitel 6 wurde erforscht, ob Absolventen eines Medizinstudiums mit vertikal integrierten Curriculum besser die Erwartungen von Supervisoren von Programmen für ärztliche Weiterbildung erfüllten. als Absolventen eines vertikal nicht integrierten Curriculums. Um diese Frage zu beantworten führten wir eine Fragebogenstudie unter Supervisoren für die ärztliche Weiterbildung in Utrecht (Niederlande; vertikal integriertes Curriculum für das Medizinstudium) und in Hamburg (Deutschland: vertikal nicht integriertes Curriculum für das Medizinstudium) durch. Die Supervisoren wurden aufgefordert eine Einschätzung bezüglich der Vorbereitung ihrer jeweiligen Absolventen auf die ärztliche Tätigkeit, über deren Kenntnisse und über deren Vermögen, einige spezifische Aufgaben in ihrer Arbeit als Ärzte zu managen abzugeben. Die beiden Gruppen der Supervisoren unterschieden sich nicht bezüglich der Auffassungen über ihre Absolventen hinsichtlich ihrer Vorbereitung für die ärztliche Tätigkeit, ihres biomediziin het eerdere onderzoek, beschreven in hoofdstuk 4. De respondenten werd ook gevraagd in welke mate hun opleiding hen had voorbereid op het werk als arts en op de (medische) vervolgopleiding. De basisartsen die een verticaal geïntegreerd curriculum hadden gevolgd waren positiever over de voorbereiding op werk en vervolgopleiding.

Het doel van het onderzoek beschreven in hoofdstuk 6 was nagaan of basisartsen die een verticaal geïntegreerd curriculum hebben doorlopen beter voldoen aan de verwachtingen van supervisoren van vervolaopleidinaen dan basisartsen die een niet-verticaal geïntegreerd curriculum hebben doorlopen. Om deze vraag te kunnen beantwoorden voerden we een vragenlijstonderzoek uit onder supervisoren van medische vervolgopleidingen in Utrecht (Nederland; geneeskunde-opleiding met een verticaal geïntegreerd curriculum) en Hamburg (Duitsland; geneeskunde-opleiding met een niet-verticaal geïntegreerd curriculum). De supervisoren werd gevraagd een oordeel te geven over de mate waarin basisartsen toegerust zijn voor het werk,

expectations of supervisors of postgraduate training programmes better than graduates from a non-vertically integrated curriculum. To answer this question, we conducted a questionnaire study among supervisors of postgraduate training programmes in Utrecht (the Netherlands; vertically integrated curriculum at medical school) and Hamburg (Germany; non-vertically integrated curriculum at medical school). The supervisors were asked about their medical graduates' preparedness for work, knowledge and capabilities to manage some specific parts of the work as a physician. The two groups of supervisors did not differ in their judgment of the graduates regarding preparedness for work, amount of biomedical knowledge, amount of clinical knowledge and amount of pathophysiological knowledge about diseases. However, graduates from a vertically integrated curriculum were evaluated higher by their supervisors with respect to the capability to work independently, to solve medical problems, to manage unfamiliar medical situations, to

nischen und klinischen Wissens und ihrer pathophysiologischen Krankheitskenntnisse. Allerdings wurden Absolventen eines vertikal integrierten Curriculums von ihren Supervisoren höher eingeschätzt hinsichtlich ihrer Fähigkeiten, unabhängig zu arbeiten, medizinische Probleme zu lösen, unbekannte medizinische Situationen zu managen, Aufgaben zu priorisieren, mit anderen Personen zusammenzuarbeiten, einzuschätzen, wann sie ihren Supervisor konsultieren sollten, und ihre Tätigkeiten zu reflektieren.

Kapitel 7, 8 und 9 beschreiben die Entwicklung einer authentischen Leistungsprüfung für die Einschätzung der Praxisreife von Absolventen eines Medizinstudiums. Diese Prüfung wurde in Utrecht in Zusammenarbeit mit der Medizinischen Fakultät der Universität Hamburg entwickelt und "Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test" (UHTRUST) genannt.

Zur Entwicklung der Prüfung ist es relevant, die generellen Eigenschaften von Auszubildenden zu kennen, die es ihren Supervisoren ermöglichen, ihnen die

over hun kennis en over hun capaciteiten om bepaalde relevante taken van een arts uit te voeren. Wat betreft de mate van toeaerustheid voor het werk. de hoeveelheid biomedische kennis, klinische kennis en pathofysioloaische ziektekennis deden zich aeen verschillen voor in de oordelen van de twee groepen supervisoren. Afgestudeerden die een verticaal geïntegreerd curriculum hadden doorlopen werden door hun supervisoren echter hoaer beoordeeld met betrekking tot het vermogen om zelfstandig te werken, medische problemen op te lossen, om te gaan met onbekende medische situaties, taken te prioriteren, samen te werken met anderen, in te schatten wanneer ze hun supervisor moeten raadplegen en te reflecteren op hun eigen werk.

De hoofdstukken 7, 8 en 9 beschrijven de ontwikkeling van een authentiek performance assessment voor het beoordelen van de mate waarin basisartsen toegerust zijn voor het werk in de klinische praktijk. Het assessment werd in Utrecht ontwikkeld, in samenwerking met de geneeskunde-opleiding in Hamburg. Het assessment kreeg de naam "Utrecht Hamburg Trainee prioritize tasks, to collaborate with other people, to estimate when they have to consult their supervisors and to reflect on their activities.

Chapters 7, 8 and 9 describe the development of an authentic performance assessment for the assessment of readiness for clinical practice of medical graduates. The assessment was developed in Utrecht in collaboration with the medical school in Hamburg and the assessment was called 'Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test'(UHTRUST).

For the development of the assessment, it is relevant to know which general features of trainees facilitate supervisors' trust in trainees to perform critical clinical tasks. The aim of the study presented in chapter 7 was to uncover such essential *facets of competence*(FOCs), based on the opinion of experienced physician educators. First, we developed a draft questionnaire with a preliminary list of FOCs, drawn from a literature

Durchführung kritischer klinischer Aufgaben anzuvertrauen. Das Ziel der Studie, die in Kapitel 7 dargestellt wird, war es deshalb, solche essentiellen "Kompetenzfacetten" (FOCs) auf der Basis der Ansicht von erfahrenen ärztlichen Weiterbildern zu ermitteln. Zuerst entwarfen wir einen Fragebogen mit einer vorläufigen Liste von FOCs aus einer Literaturübersicht zu ausgewählten Kompetenzrahmen und relevanten Publikationen. Basierend auf diesen Ouellen erstellten wir eine Liste von 24 FOCs. Im nächsten Schritt führten wir eine Delphistudie mit zwei Runden unter erfahrenen ärztlichen Ausbildern in den Niederlanden durch. Die Experten wurden gebeten, die FOCs hinsichtlich zwei Aspekten zu bewerten: (1) Eindeutigkeit und Vollständigkeit der Formulierung, (2) Wichtigkeit für die Entscheidung, einem Arzt, der die ärztliche Weiterbildung beginnt, eine kritische klinische Aufgabe anzuvertrauen. Die Delphistudie erbrachte eine Liste mit 25 FOCs. Wesentlich ist, dass unter den Experten darüber Einigkeit herrschte, dass diese FOCs wichtig oder sehr wichtig für die Entscheidung sind, einem Arzt, der die *Responsibility for Unfamiliar Situations Test" (UHTRUST).*

Voor de ontwikkeling van een dergelijk assessment is het relevant om te weten welke alaemene kenmerken van artsen in opleiding maken dat supervisoren hen kritische klinische taken toevertrouwen. Het doel van het onderzoek waarover gerapporteerd wordt in hoofdstuk 7 was het achterhalen van zulke essentiële 'competentiefacetten', gebaseerd op de meningen van ervaren artsopleiders. Als eerste stap ontwikkelden we een concept vragenlijst met een voorlopige lijst van competentiefacetten, die voortkwam uit een analvse van geselecteerde competentie-raamwerken en relevante publicaties. Op basis van deze bronnen stelden we een lijst op met 24 competentiefacetten. Vervolgens voerden we een Delphi-onderzoek met twee ronden uit onder ervaren Nederlandse artsopleiders. Deze experts werd gevraagd om elk van de competentiefacetten op twee punten te beoordelen: (1) duidelijkheid en compleetheid van de formulering, (2) belang voor het toevertrouwen van kritische klinische taken aan beginnende specialisreview of selected competency frameworks and relevant publications. Based on the sources, we constructed a list of 24 FOCs. Next, we conducted a Delphi study of two rounds among experienced physician educators in the Netherlands. The experts were asked to score each of the FOCs on two aspects: (1) Clarity and completeness of formulation, (2) Importance for the entrustment of critical clinical tasks to starting residents. The Delphi study resulted in a list with 25 FOCs. The most important finding was that the experts agreed that these FOCs are important or very important for the entrustment of critical clinical tasks to starting residents.

The aim of the study described in chapter 8 was to determine the external validity, i.e. the generalizability of the facets of competence resulting from the Delphi study presented in chapter 7. Our approach was to ask physician educators in two countries to rank these FOCs and

ärztliche Weiterbildung beginnt, kritische klinische Aufgaben anzuvertrauen.

Das Ziel der Studie, die in Kapitel 8 beschrieben wird, war die Festlegung der externen Validität, d.h. die Generalisierbarkeit der Kompetenzfacetten, die sich aus der Delphistudie, die in Kapitel 7 dargestellt wurde, ergeben hatten. Unser Ansatz war, Ausbilder in der ärztlichen Weiterbildung aus zwei Ländern zu bitten, diese FOCs in eine Rangliste zu bringen und anschließend den Grad der Übereinstimmung zwischen den beiden Ländern bezüglich der am höchsten bewerteten FOC zu überprüfen. Die Studie wurde unter Mitwirkung von niederländischen und deutschen Ausbildern in der ärztlichen Weiterbildung durchgeführt. Insgesamt kann konstatiert werden, dass niederländische medizinische Hochschulen "moderne medizinische Curricula" aufweisen und die meisten deutschen medizinischen Fakultäten vorwiegend "traditionelle medizinische Curricula" haben. Die niederländischen und deutschen Ausbilder bewerteten die Relevanz von jedem einzelnen FOC. Ein Ranglistenvergleich zeigte, dass ten in opleiding. Het Delphi-onderzoek resulteerde in een lijst met 25 competentiefacetten. De belangrijkste uitkomst was dat de experts het erover eens waren dat deze competentiefacetten belangrijk of zeer belangrijk zijn voor het toevertrouwen van kritische klinische taken aan basisartsen aan het begin van de vervolgopleiding.

Het doel van het onderzoek beschreven in hoofdstuk 8 was om een uitspraak te kunnen doen over de externe validiteit. oftewel de generaliseerbaarheid van de competentiefacetten die voortkwamen uit het Delphi-onderzoek dat is beschreven in hoofdstuk 7. Onze benaderina was om artsopleiders uit twee landen te vragen een rangorde aan te brengen in deze competentiefacetten en om over de competentiefacetten met de hoogste scores de mate van overeenstemmina te bepalen tussen beide landen. Het onderzoek werd uitgevoerd onder Nederlandse en Duitse artsopleiders. In het algemeen kan worden gesteld dat Nederlandse geneeskundeopleidingen 'moderne medische curricula' hebben, terwijl de meeste Duitse geneeskunde-opleidingen overwegend traditioto determine the level of agreement among the countries about the highest scoring FOCs. The studied was carried out among Dutch and German medical educators. In general can be said that Dutch medical schools have modern medical curricula and most German medical schools have predominantly traditional medical curricula. The Dutch and German educators scored the relevance of each FOC. A rank-order comparison showed that there was almost full agreement about the top 10 FOCs. Among those were: 'scientific and empirical grounded method of working', 'knowing and maintaining own personal bounds and possibilities', 'active professional development', 'teamwork and collegiality', 'active listing to patients' and 'verbal communication with colleagues and supervisors'.

In chapter 9 we provided a validity argument for UHTRUST, an authentic simulation procedure to assess whether medical trainees are ready to be

fast vollständige Übereinstimmung bezüglich der ersten 10 Rangplätze bestand. Unter diesen waren: "wissenschaftlich und empirisch begründete Arbeitsmethoden", "Kennen und Handhaben von persönlichen Grenzen und Möglichkeiten", "aktive professionelle Entwicklung", "Teamwork und Kollegialität", "aktives Zuhören gegenüber Patienten" und "verbale Kommunikation mit Kollegen und Vorgesetzten."

In Kapitel 9 liefern wir ein Validitätsargument für UHTRUST, eine authentische Simulation um einzuschätzen, ob Absolventen eines Medizinstudiums reif sind, mit unbekannten klinischen Aufgaben betraut zu werden. In dieser Prüfung werden Medizinstudenten kurz vor Abschluss ihres Studiums in die Rolle von Ärzten zu Beginn ihrer ärztlichen Weiterbildung an einem sehr arbeitsreichen Tag versetzt. Sie wurden von Oberärzten, Schwestern und standardisierten Patienten bewertet. Um die Validität von UHTRUST zu evaluieren, folgten wir Kanes begründungsbasiertem Ansatz für Validierung. Ein theoretischer Rahmen für die Validitätsbegründung wurde auf verschiedenen theoretischen

neel zijn. De Nederlandse en Duitse opleiders bepaalden van elk competentiefacet de mate van relevantie voor het toevertrouwen van kritische taken aan artsen in opleiding. Op basis van een vergelijking van de ranaordes, bleek dat de top 10 van de Duitse en Nederlandse experts bijna volledig overeenkwam. Deze top 10 bevat onder andere de competentiefacetten: "wetenschappelijk en empirisch gefundeerde werkwiize", "kennen en handhaven van de eigen grenzen en mogelijkheden", "actieve professionele ontwikkeling", "teamwork en collegialiteit", "actief luisteren naar de patiënt" en "mondelinge communicatie met collega's en supervisoren".

In hoofdstuk 9 we beschrijven de validiteit van UHTRUST, een authentiek assessment waarmee kan worden bepaald of artsen in opleiding bekwaam genoeg zijn om hen onbekende klinische taken toe te vertrouwen. Bij dit assessment kregen bijna-afgestudeerde artsen de rol van artsen in het begin van de medische vervolgopleiding op een hele drukke dag. Ze werden beoordeeld door artsen, verpleegkundigen en simulatiepatiënten. Om de validiteit van entrusted with unfamiliar clinical tasks. In this assessment near-graduates are placed in the role of beginning residents on a very busy day. They were assessed by clinicians, nurses and standardized patients. To evaluate the validity of UHTRUST we followed Kane's argument-based approach for validation. A theoretical framework for the validity argument was developed based on various theoretical and empirical studies related to the argument-based approach to validity. The framework illustrates the following major inferences that are associated with an argumentbased approach to validity and their underlying assumptions:

- Scoring: from the observed performance to the observed score.
- Generalization: from the observed score to the expected universe score.
- Extrapolation: from the universe score to the expected level of skill in the target domain.

und empirischen Studien zum Thema, begründungsbasierten Ansatz für Validität' basiert. Das Rahmenwerk illustriert die folgenden Hauptdeduktionen, welche mit dem begründungsbasierten Ansatz für Validität unter folgenden zugrundeliegenden Annahmen assoziiert sind:

• Bewertung: von der beobachteten Leitung zur beobachteten Bewertung.

• Generalisierbarkeit: von der beobachteten Bewertung zur erwarteten Gesamtbewertung.

• Hochrechnung: von der Gesamtbewertung zur erwarteten Fertigkeitsebene in der Zieldomäne.

• Interpretation: von der Fertigkeitsebene in der Zieldomäne zur Testinterpretation.

Alle verfügbaren Daten wurden verwendet, um eine kohärente und plausible Begründung darzulegen. Ein beträchtlicher Datensatz wurde während der Entwicklung der Prüfung zusammengestellt. Außerdem wurde ein Generalisierbarkeitsstudie durchgeführt, um die Reliabilität der Bewertungen der Prüfer zu evaluieren und das Verhältnis der durch die Prüflinge und UHTRUST te kunnen beoordelen gebruikten we Kane's 'argument-based approach for validation'. Op basis van literatuuronderzoek hebben we een theoretisch raamwerk opgesteld voor het validiteitsargument. Dit raamwerk expliciteert de volgende gevolgtrekkingen (inferenties), waarvoor vervolgens validiteitsbewijs wordt verzameld:

- Score: van de geobserveerde prestatie naar de geobserveerde score.
- Generalisatie: van de geobserveerde score naar de verwachte score over het testdomein.

• Extrapolatie: van het testdomein naar het verwachte functioneren in het praktijkdomein.

 Interpretatie: van het verwachte functioneren in het praktijkdomein naar de interpretatie van het assessment en het bepalen van de geschiktheid van het assessment.

Alle beschikbare data werden gebruikt om een samenhangend en aannemelijk validiteitsargument op te stellen. Tijdens de ontwikkeling van de assessmentprocedure werd een aanzienlijke hoeveelheid data verzameld. Daarnaast is een gene• Interpretation: from the level of skill in the target domain to the test interpretation.

All available data were used to design a coherent and plausible argument. Considerable data was collected during the development of the assessment procedure. In addition, a generalizability study was conducted to evaluate the reliability of the scores given by assessors and to determine the proportion of variance accounted by candidates and assessors. It was found that most of Kane's validity assumptions were defendable with accurate and often parallel lines of backing. Therefore, UHTRUST can be used to compare the readiness for clinical practice of medical graduates.

Chapter 10 focuses on the differences in readiness for clinical practice between graduates from a vertically integrated curriculum and from a non-vertically integrated curriculum. In particular we wanted to know

Prüfer bedingten Varianz zu bestimmen. Es wurde festgestellt, das die meisten Validitätsannahmen gemäß Kane mit akkuraten und oft parallelen Back-up Linien gemacht werden konnten. Daher kann UHTRUST verwendet werden, um die Praxisreife von Absolventen eines Medizinstudiums zu vergleichen.

Kapitel 10 beschäftigt sich mit den Unterschieden in der klinischen Praxisreife zwischen Absolventen eine vertikal integrierten und eines vertikal nicht integrierten Curriculums. Insbesondere wollten wir herausfinden, welche Unterschiede zwischen diesen beiden Gruppen in ihren Kompetenzen bestehen, mit unbekannten klinischen Situationen umzugehen. Um diese Frage zu beantworten, nahmen Studierende kurz vor Abschluss ihres Medizinstudiums aus Utrecht/Groningen (vertikal integriertes Curriculum) und Hamburg (vertikal nicht integriertes Curriculum) an UHTRUST teil. Jeder Kandidat wurde bezüglich verschiedener Kompetenzfacetten unabhängig voneinander von zwei Oberärzten, einer Pflegekraft und sechs standardisierten Patienten eingeschätzt. raliseerbaarheidsstudie uitgevoerd om de betrouwbaarheid van de toegekende scores te bepalen en om vast te stellen welk percentage van de variantie kan worden toegeschreven aan de kandidaten en de beoordelaars. Het bleek dat de meeste veronderstellingen in het valideitsargument verdedigbaar waren en konden worden onderbouwd met bewijzen. UHTRUST kan daarom worden gebruikt om te bepalen in welke mate basisartsen toegerust zijn voor het werk in de klinische praktijk.

Hoofdstuk 10 gaat in op de vraag in hoeverre basisartsen die een verticaal geïntegreerd geneeskundecurriculum hebben gevolgd en basisartsen die een niet-verticaal geïntegreerd curriculum hebben gevolgd van elkaar verschillen in de mate waarin zij zijn voorbereid op het werk in de klinische praktijk. In het bijzonder wilden we weten wat de verschillen zijn tussen deze twee groepen in hun vermogen om om te gaan met onbekende medische situaties. Om deze vraag te beantwoorden hebben bijna-afgestudeerden van de geneeskunde-opleidingen in Utrecht/Groningen (verticaal geïntegreerd curriculum) what the differences are between these two groups in their competences to cope with unfamiliar clinical situations. To answer this question, near-graduates from Utrecht/Groningen (vertically integrated curriculum) and Hamburg (non-vertically integrated curriculum) participated in UHTRUST. Each candidate was independently assessed on different facets of competence by two clinicians, one nurse and six standardized patients. Afterwards, the clinicians indicated how much supervision they estimated each candidate would require on nine Entrustable Professional Activities (EPAs) unrelated to the observed cases. We did not find many differences between the two groups of candidates. Graduates from a vertically integrated curriculum received significantly higher scores by the clinicians for the facet of competence 'active professional development', with features as 'reflection' and 'asking for feedback'. In addition, we found some differences in the scores for the EPAs and the scores given by the nurses and standardized patients.

Anschließend gaben die Oberärzte ihre Einschätzung ab, wie viel Supervision jeder Kandidat für neun "entrustable professional activities" (EPAs), die ohne Bezug zu den beobachteten Fällen waren, benötigen würde. Wir fanden nicht viele Unterschiede zwischen den beiden Kandidatengruppen. Absolventen des vertikal integrierten Curriculums erhielten durch die Oberärzte signifikant höhere Bewertungen für die Kompetenzfacetten "aktive professionelle Entwicklung", die Eigenschaften wie "Reflektion" und "Feedback einfordern" beinhaltete. Außerdem fanden wir einige Unterschiede in den EPA Bewertungen und bei den Bewertungen, die von den Pflegekräften und den standardisierten Patienten abgegeben wurden.

Kapitel 11 fasst die Hauptergebnisse der vorhergehenden Kapitel zusammen, beschreibt die Stärken und Schwächen des Forschungsprojekts und präsentiert Vorschläge für weitere Forschung und Praxis in der medizinischen Ausbildung. Die neun Studien in dieser Arbeit erweitern das Verständnis des Einflusses von vertikal integrierten Curricula auf die verschiede-

en Hamburg (niet-verticaal geïntegreerd curriculum) aan het UHTRUST assessment deelgenomen. Elke kandidaat werd onafhankeliik beoordeeld door twee artsen. een verpleegkundige en zes simulatiepatienten op verschillende competentiefacetten. Daarna hebben de artsen aanaeaeven hoeveel supervisie de kandidaten volgens hen nodig zouden hebben bij het uitvoeren van negen 'entrustable professional activities' (EPAs), die niet overeenkwamen met de geobserveerde casussen. We vonden niet veel verschillen tussen de twee groepen. Basisartsen die een verticaal geïntegreerd curriculum hadden gevolgd werden door de artsen sianificant beter beoordeeld op de competentie 'actieve professionele ontwikkeling', met kenmerken als 'reflectie' en 'vraaen om feedback'. Daarnaast vonden we enkele verschillen in de EPA-scores en de scores die de verpleegkundigen en simulatiepatiënten hebben gegeven.

In hoofdstuk 11 worden de belangrijkste bevindingen uit de voorgaande hoofdstukken samengevat, sterke en zwakke punten van het onderzoeksproject beschreven en aanbevelingen gedaan voor verder onChapter 11 summarizes the main findings of the previous chapters, describes strengths and weaknesses of the research project and presents suggestions for further research and the practice of medical education. The nine studies in this thesis improve the understanding of the influence of vertically integrated curricula and different types of transitions on the readiness for clinical practice of medical graduates. Based on the results of the distinctive studies can be concluded that vertical integration of undergraduate medical curricula seems to have a positive influence on the readiness for clinical practice of medical trainees, although not all studies confirm this statement as evident as we had expected. Both the medical graduates and supervisors of postgraduate training programmes indicate that vertically integrated curricula prepare the students better for working as physicians. Also the finding that the condition in which trainees get more responsibility affects the competence development positively, can be seen as an

nen Arten von Übergängen bezüglich der Praxisreife von Absolventen eines Medizinstudiums. Basierend auf den Ergebnissen dieser markanten Studien kann geschlussfolgert werden, dass die vertikale Integration von Curricula im Medizinstudium einen positiven Einfluss auf die Praxisreife von Absolventen hat, obwohl nicht alle Studien diese Aussage mit solcher Evidenz bestätigen, wie wir erwartet hatten. Sowohl die Absolventen, als auch die Supervisoren der ärztlichen Weiterbildungsprogrammen, gaben an, dass vertikal integrierte Curricula Studenten besser auf ihre Arbeit als Ärzte vorbereiteten. Auch der Befund. dass die Bedingungen zu der Auszubildende mehr Verantwortung erhalten, die Kompetenzentwicklung positiv beeinflusst, kann als Untermauerung der Effektivität der vertikalen Integration angesehen werden. Außerdem liefert diese Arbeit eine Übersicht über die existierenden Strukturen von medizinischer Ausbildung in aller Welt. Schlussendlich entwickelten wir ein neues Prüfungsinstrument, um die klinische Praxisreife zu messen. Um das Instrument für den intendierten Zweck zu validieren, wurden Evidenzen gesammelt um

derzoek en de praktijk van het medisch onderwijs. De negen onderzoeken in dit proefschrift dragen bij aan kennis over het effect van verticaal geïntegreerde curricula en verschillende soorten transities op de mate waarin basisartsen voorbereid ziin op de klinische praktijk. Op basis van de resultaten van de verschillende onderzoeken kan worden geconcludeerd dat verticale integratie een positieve invloed lijkt te hebben op de mate waarin basisartsen zijn toegerust voor het werk in de klinische praktijk, hoewel niet alle onderzoeken dit even duidelijk bevestigden als we hadden verwacht. Zowel de basisartsen als supervisoren van medische vervolgopleidingen geven aan dat verticaal geïntegreerde curricula de studenten beter voorbereiden op het werk als arts. Ook de bevinding dat de situatie waarin artsen in opleiding meer verantwoordelijkheid krijgen hun competentieontwikkeling positief beïnvloedt, kan worden gezien als argument voor de effectiviteit van verticale integratie. Daarnaast verschaft dit proefschrift informatie over structuren van medisch onderwijs die wereldwijd bestaan. Tot slot ontwikkelden we een nieuw assessmentinstrument waarunderpinning of the effectiveness of vertical integration. In addition this thesis provides information about the existing structures of medical education around the world. Finally, we developed a new assessment instrument to determine the readiness for clinical practice. To validate this instrument for the intended purpose, evidence was collected to support the arguments for validity. The strengths of the studies we performed are the combination of research methods and perspectives and the fact that the research project was cross-institutional and international. Important limitations of the studies were that in the comparison studies just a part of the total population was involved and that all participants participated voluntarily. In addition, the fact that the students who participated in UHTRUST were assessed by clinicians, nurses and standardized patients from their own country is a limitation. Chapter 11 also describes implications for educational practice. Especially two features of vertical integration deserve more attention

die Validitätsbegründung zu unterstützen. Die Stärken der von uns durchgeführten Studien bestehen in der Kombination der Forschungsmethoden und der Tatsache, dass das Forschungsprojekt sich über verschiedenen Institutionen und Länder erstreckte. Das an den Vergleichsstudien nur ein Teil der Gesamtpopulation beteiligt war und dass die Teilnehmer freiwillig mitwirkten, muss bei der Interpretation der Eraebnisse berücktlichtig werden. Außerdem ist die Tatsache, dass die an UHTRUST teilnehmenden Studierenden von Oberärzten, Pflegekräften und standardisierten Patienten ihrer eigenen Länder bewertet wurden, eine Studienschwäche. Kapitel 11 beschreibt darüber hinaus die Konsequenzen für die Ausbildungspraxis. Insbesondere zwei Eigenschaften der vertikalen Integration verdienen mehr Aufmerksamkeit von Curriculumsentwicklern, d.h. "schrittweise Zunahme der klinischen Verantwortung von Studierenden über alle Kurse und Famulaturen" und "lange klinische Praktika mit einer Wahlfachnatur und substantieller Verantwortung in der Patientenversorgung im letzten Jahr des Curriculums." Weiterhin kann das Design mee kan worden bepaald in welke mate basisartsen zijn toegerust voor de klinische praktijk. Om dit instrument te valideren voor het beoogde doel, zijn bewijzen verzameld om het validiteitsargument te ondersteunen.

Sterke punten van de onderzoeken die we uitvoerden zijn de combinatie van onderzoeksmethoden, de verschillende perspectieven en het feit dat het onderzoeksproject internationaal van opzet was en in meerdere instellingen is uitgevoerd. Belangrijke beperkingen van de onderzoeken zijn dat slechts een deel van de totale populatie deelnam aan de vergelijkende onderzoeken en dat alle participanten op vrijwillige basis deelnamen. Ook het feit dat de studenten die deelnamen aan UHTRUST werden beoordeeld door artsen, verpleegkundigen en simulatiepatiënten uit hun eigen land is een beperking.

In hoofdstuk 11 worden ook consequenties van de onderzoeksresultaten voor de onderwijspraktijk beschreven. Twee kenmerken van verticale integratie verdienen in het bijzonder meer aandacht van curriculumontwerpers, te weten "een voortschrijfrom curriculum developers, i.e. 'a progressive increase of clinical responsibility of students over the course of all clerkships' and 'long clinical clerkships with an elective nature and substantial responsibility for patient care in the final year of the curriculum'.

Further, the design of UHTRUST can be implemented in medical education, for instance for the purpose of residency selection for postgraduate training programmes. Chapter 11 concludes with suggestions for further research. Future research needs to make a direct comparison between the sense of preparedness for clinical practice of a trainee, the assessment by the supervisor about particular trainees and their actual performance in clinical practice. Also further exploration of the possible implementations of UHTRUST is recommended.

von UHTRUST in der medizinischen Ausbildung implementiert werden, z.B. zum Zweck der Auswahl von Absolventen für die ärztlichen Weiterbildungsprogramme. Kapitel 11 schlieβt mit Vorschlägen für weiterführende Forschung. Weitere Forschungsprojekte sollten einen direkten Vergleich zwischen der von Absolventen wahrgenommenen klinischen Praxisreife und der Einschätzung von Supervisoren von einzelnen Auszubildenden und ihrer tatsächlichen Leistung in der klinischen Praxis anstreben. Außerdem wird eine weitere Ausarbeitung der möglichen Implementierung von UHTRUST empfohlen. dende klinische verantwoordelijkheid van studenten in de loop van alle coschappen" en "lange klinische keuzecoschappen met substantiële verantwoordelijkheid voor patiëntenzorg in het laatste jaar van het curriculum". Verder kan de opzet van UHTRUST worden toegepast in het medisch onderwijs, bijvoorbeeld ten behoeve van de selectie voor de medische vervolgopleidingen.

Hoofdstuk 11 sluit af met aanbevelingen voor verder onderzoek. Het is belangrijk dat in toekomstig onderzoek een directe vergelijking wordt gemaakt tussen de mate waarin basisartsen zichzelf voorbereid voelen op het werk in de klinische praktijk, de inschatting van de mate voor voorbereiding door supervisoren en het feitelijke functioneren van basisartsen in de praktijk. Ook wordt aanbevolen om de toepassingsmogelijkheden van UHTRUST verder te onderzoeken.



Appendix A

Example of a UHTRUST Case

Description for assessors

Patient	details
---------	---------

Name patient	Bert van der Veen
Working title of scenario	53-year old man with progressive fatigue and haemoptysis (coughing up blood)
Final diagnosis	Wegener's granulomatosis
Background	No particulars, no allergies, all vaccinations according to schedule.
Medication	None

Medical history

The patient has been working as a missionary in Zaire for years. He is now back for a short visit to his family. Family members noticed that he is very tired and he has had episodes of coughing up blood. His family pressed him to attend the emergency care unit. In hospital, saturation turned out to be low (91%) and because of that the patient was hospitalized in the medium care ward of the internal medicine unit.

He exercises now and then (running), but not excessively. He does not use alcohol, does not smoke and does not use drugs or medication.

The history reveals:

- The patient is hospitalized in the medium care ward of the internal medicine unit.
- The patient (Dutch) works as a missionary in Zaire and is now in The Netherlands for a family visit
- The patient has been very tired for some months and has been coughing up some blood a couple of times, back in Zaire. In between no fits of coughing.
- Yesterday evening it happened again and urgently pressed by his family, the patient attended the hospital.
- Due to severe patient pressure at the emergency unit examination was limited to some lab tests and a chest x-ray. The radiograph revealed a veiling density of the left lower lobe.
- Because of low saturation (91%) the patient received oxygen and was transferred to the medium care ward of the internal medicine unit.
- The patient himself thinks it is all overexaggerated and wants to be discharged.

- He lost 5 kilograms of weight over the last two months. He blames this to the stress caused by the project in Zaire. Besides, his loss of appetite is conspicuous.
- Not nauseous, no vomiting.
- He has had several spontaneous nosebleeds as well. Very strange, because he never had those before. For a while now he has suffered from a common cold and a runny nose.
- He has a fever, circa 38°C, but because of the high temperatures in Zaire he didn't notice.
- He has been a little dyspnoeic for 3 days now, especially during exertion.
- No palpitations, no chest pain, no swollen legs.
- No stomach ache, has had no diarrhoea.

Urogenital tract:

• He urinates slightly less than he normally does; otherwise n.p.

Neurological:

▶ N.p.

General:

- No night sweats
- Vaccinated against Yellow Fever and Diphtheria Tetanus Polio
- Does not use anti-malarial drugs
- Until now, no Mantoux test carried out

Which information will be given to candidates beforehand?

Short emergency unit admission letter

Which information can be requested by candidates?

- Physical examination (for results see annex)
- Additional tests (for results of blood tests see separate form, for other additional tests see annex)
- > The CXR made at the Emergency Unit
- Consult (it will be possible to request for a consult with a specialist, but this doctor will obviously not come and see the candidate within the time of the assessment)

Role of supervisor

This case is hard to solve. Wegener's granulomatosis is hard to diagnose and candidates don't get many clues from this case. They could be misled by the patient's work in Zaire (think TB or malaria) and because of the lung problems that stand out (malignance, infection). However, in this assessment, it is very important that you also observe and assess how candidates handle a difficult case like this.

Therefore, provide no or almost no substantive guidance to candidates (let them structure the complaints and symptoms by themselves and let them formulate their own diagnostic and treatment strategy).

If necessary, you are allowed to spend a little more time on this case during the short progress meeting. Only ask candidates how far they've proceeded till now and guide them on the organisational level (which tests do they want to request and how to prioritize in the light of the remaining self-study time. Tests that take more time have to be requested earlier on; results will then hopefully be available in time).

Of course, candidates are allowed to call their supervisor to deliberate about the tests they consider to request.

End product of candidates

- A general medical history should be taken and at least the respiratory tract should be sufficiently inquired into (because of haemoptysis and low saturation).
- Furthermore, the fatigue should be sufficiently investigated and attention should be paid to this gentleman's stay in Zaire (increased risk of TB, malaria and other tropical diseases).
- It is crucial for the candidate to link the veiling density in the lung, the nosebleeds and chronic colds and the decreased kidney function. Only when the candidate notices these things, he/she will think beyond the obvious and will consider the rarer diseases.
- A differential diagnosis has been formulated in which autoimmune diseases are present at least (out of the box thinking with the given complaints).
- There has to be a diagnostic strategy (partly implemented already) based on which one is able to differentiate between the different diagnoses in the DD.
- The treatment strategy has to be started with. If this is not the case yet, ask for the treatment of the most likely diagnosis (if they haven't reached the conclusion of Wegener's granulomatosis, ask them for the most likely group of diseases: autoimmune. Which therapy is often used in treating autoimmune diseases:

immunosuppressive drugs, like prednisone. Wegener's granulomatosis is further treated with a complex combination of steroids, plasmapheresis and immunoglobulins.)

During the reporting phase questions about content may of course be asked. However, the most important question in this assessment is: how did candidates reach a differential and probability diagnosis and do they know how to act subsequently (right history, diagnostics and treatment strategy). Don't forget that this case is difficult to solve and that the candidates' differential diagnosis and treatment strategy will be less extensive.

Physical examination Bert van der Veen

General impression:

- Ectomorphic habitus (1.78 m; 63 kg), tired, dyspnoeic
- ▶ Temperature 38.2°C
- Blood pressure 125/80 mmHg
- Saturation 91%

Examination of thorax:

- Inspection: increased expiratory flow
- > Percussion: impaired resonance over left lower lobe
- Auscultation: lungs: diminished breath sounds over left lower lobe, no adventitious breath sounds. Heart: S1 S2 no souffles.

Examination of abdomen:

- Inspection: No visible abnormalities
- Auscultation: normal peristalsis, no abdominal bruits
- Percussion: no abnormalities
- Palpation: no pain, no abnormal tension, no costovertebral tenderness.

Neurological examination:

- Exploratory neurological examination: no abnormalities. (Possibilities: Pupillary light reflex, examination of eye movement, examination of visual fields, face (show teeth, close eyes tightly, sensibility)
- Muscle strength (keeping arms outstretched for one minute, pushing arms against something, squeezing hands forcefully, walking on heels and toes)
- Coordination (walk on a straight line, finger-to-nose test)
- Sensibility, reflexes.

ENT examination:

- Ears: no abnormalities
- Nose: has a cold. Stuffy nose, crusting inside. Pain when pressing on sinus cavities.
- Throat: no abnormalities, no swollen lymph nodes in neck, axilla or groins.

Skin and extremities:

- No oedema in the legs. No visible abnormalities of the skin, no peripheral cyanosis.
- > Joints: no redness, swelling, left-right differences. Movement unrestricted and not painful.
- Capillary refill < 5 sec.

Overview of additional examinations

Candidates may be asking for the following additional examinations:

- Lab
 - Inflammatory parameters
 - Hb, ht, мсv
 - Liver function
 - Kidney function
 - Electrolytes
 - PT, APTT, thrombin time
 - ANA/ANCA
- Urine sediment
- Thick blood smear for Malaria
- Sputum culture
- Peripheral blood smear
- CXR
- Renal ultrasound
- Biopsy sample from the inflamed area (in this case kidney or lung). Results will not be in time during assessment.

Results:

- Lab, see table (c-ANCA pos).
- Urine, see table (dysmorphic red cells, acanthocytosis).
- Thick blood smear for malaria. Result given to candidates: No parasites present, over 200 fields examined.

- Sputum culture. Result given to candidates: No growth of pathological bacteria.
- Peripheral blood smear. Result given to candidates: No spherocytes.
 Fragmentocytes in the smear.
- CXR. Result given to candidates: Normal position diaphragm, veiling density left lower lobe, no lymphadenopathy.
- Renal ultrasound. Result given to candidates: both sides normal size, normal echotexture, no signs of hydronephrosis.

Description for standardized patient

Patient details

Name patient	Bert van der Veen
Working title of scenario	53-year old man with progressive fatigue and haemoptysis (coughing up blood)

Guiding question

You want to go back to Zaire as soon as possible, so your problem has to be solved as quickly as possible. Besides, you do not quite recognize that you are having a problem. You don't feel ill enough to be in hospital and therefore don't wish to stay.

Current situation

You are hospitalized in the medium care ward of the internal medicine unit because of coughing up blood and low saturation in the blood. Now you are talking to the attending physician of the ward. He/she will take up your story, carry out / request for additional tests and will come up with a treatment strategy.

General information/psycho-social situation

You are Bert van der Veen and you are 53 years old. You are not married. As a student you came in contact with Zaire during an internship. After that experience you have committed yourself to the care of people and the improvement of living conditions in Zaire. After working from The Netherlands for several years, you preferred to work at the 'core'. Therefore, you have now been working as a missionary for ten years, paying regular visits to the camp Mugunga. There you assist in the care for orphaned children and, among other things, are involved in the distribution of food and the general organisation of the camp. You have been in Zaire for 6 months, without intervals. You went back to The Netherlands to prepare for a new project and decided to pay a visit to your family.

You exercise now and then (running), but not excessive. You don't use alcohol. You do not smoke, do not use drugs or medication. At most an occasional aspirin when you have a headache.

Background

Back in Zaire, you have been coughing up blood a couple of times and you have been somewhat more tired for a couple of months now. You never paid much attention to these complaints though.

Yesterday evening, you have been coughing up blood in the presence of your family. This scared them so much that they told you they have serious concerns about your health. Because your family insisted upon it, yesterday evening you attended the emergency unit of this hospital. Due to a large traffic accident, patient pressure on the emergency ward was very high. Therefore they only draw some blood and took a CXR and you have been admitted directly afterwards. They told you that your saturation was too low (91%, while 95–100% is normal in adults) and that is why you were admitted to the medium care ward of the internal medicine unit. To tell the truth, you do not understand this decision and think it's all rubbish.

Complaints

You suffer from the following complaints. Note that these complaints already existed in Zaire:

- You have been coughing up blood several times, about 5 times over the last month. Not a lot of blood, bright red, just marks in the sputum that you are coughing up. You don't cough otherwise, no fits of coughing in between or anything like that.
- You feel tired for some months. You noticed that you often have to take a nap in the afternoon and at night, you go to bed early. You used to sleep for just 6 hours a night on average. Now you sleep at least 10 hours a day.
- You feel dyspnoeic more quickly. You notice that carrying out some of your tasks, like the distribution of food, gets increasingly difficult and takes more energy. It seems as if you are exhausted more quickly. And for 3 days now, you really feel a bit dyspnoeic. You are panting after just a small exertion. This is not something you are familiar with. For example: "I feel very old—as if everything I do requires extra energy, even breathing".
- Weight loss. You lost 5 kilograms of weight over the last two months. You blame this to the stress caused by the project in Zaire, though. Besides, you noticed that you lost your appetite. You are not nauseous, but you just have no appetite.
- Continuing colds and nosebleeds. You had several spontaneous nosebleeds (circa 3 times over the last 2 months). You find this very strange because you never had nosebleeds before. For about 4 months now, you have a common cold and a runny nose as well.

You don't seem to get over it. Nasal drainage is clear, not green or anything like that.

Fever: 38°C. Because of the temperature in Zaire you never really noticed that you had a fever. But being here in The Netherlands, you noticed that you felt rather hot all the time, and that's why you measured your temperature a few times. Every time your temperature was around 38°C.

Additional medical information, only available on request of the doctor:

- No palpitations, no chest pain, no swollen legs.
- No abdominal pain of diarrhoea. No nausea or vomiting.

Urogenital (bladder and kidneys):

- > You urinate slightly less than normal
- No pain urinating, urine doesn't smell or look differently
- No weak urine stream, no dribbling.
- No flank pain

Neurological:

- No limitations in mobility in arms and/or legs
- No sensibility changes
- No headache

Circulatory (heart and blood vessels):

- No palpitations
- No chest pain
- No swollen legs
- You never had heart problems, and as far as you know heart diseases don't run in your family

Gastro-intestinal (stomach and intestines):

- > You have had no stomach ache
- No diarrhoea
- No nausea or vomiting
- A loss of appetite though. You just have no appetite

Musculoskeletal system and skin:

- No joint pain
- No abnormalities on skin noticed

General:

No night sweats

- Vaccinated against Yellow Fever and Diphtheria Tetanus Polio
- You do not use anti-malarial drugs

Background

You never had any unusual illness and there are no unusual illnesses that run in your family. There are no known allergies. You do not know if you have encountered TB patients. A Mantoux-test has not been carried out.

The way you behave in the experimental situation

You think it's rubbish that you have to stay in hospital. You would rather go back to your family and then back to Zaire as soon as possible.

You are 53 years old and normally very healthy. However, although you are trying to conceal it, you are certainly rather ill at the moment. Try therefore to adopt a role in which you act as if nothing is the matter, although you are actually quite ill. This is reflected by you being less cheerful and healthy than you normally are and by being a little short of breath now and then (a bit of wheezing in between sentences).

Try to enact this role as best as possible. You are absolutely not supposed to invent any additional medical information or add any hospital experience that you may possess to the scenario. This to prevent the medical scenario from becoming more complicated than originally intended. When asked for symptoms or complaints not described in this scenario your answer will be 'no' or 'I really wouldn't know'. When questioned about your past you are free to describe the common children's diseases that you contracted yourself.

Letter from emergeny unit

Mr. van der Veen, Bert

Date of birth: 25-09-1957

On August 25 we saw the above patient on the emergency unit.

Past Medical History (РМН)/ No particulars

Medications/ none

History of the present illness (HPI)/ the patient has been coughing up blood several times over the past few months. This happened again yesterday and that is why he is now attending the emergency unit. Subfebrile temperature (37.5–38.0), not feeling ill, though very tired for some time. Because of severe patient pressure at the emergency unit due to a large traffic accident and admission indication haemoptysis, medical history and physical and supplementary examination were kept very limited.

The patient works in Zaire as a missionary and is in the Netherlands for a 1 week visit to his family.

Physical examination/ healthy man, looks a little dyspnoeic.

CV: S1, S2, no souffles

Resp: CTA bilaterally, no r/r/w

Vital signs: T (Temperature): 37,5; BP (Blood pressure): 120/70; P (Pulse): 80/min; O₂ Sats (Oxygen Saturation): 91%

Additional tests/ Laboratory tests: inflammatory markers, results not available yet. CXR will be assessed by radiologist.

Policy/ after consultation with supervisor 1.5L Oxygen directly and rapid transfer to medium care because of dyspnoea and haemoptysis. Further medical examination and diagnostic procedure will take place there.

Results additional examinations

Peripheral blood smear

No spherocytes. Fragmentocytes in the smear.

Thick blood smear

No parasites present, over 200 fields examined.

Sputum culture

No growth of pathological bacteria.

Renal ultrasound Bert van der Veen

Ultrasound right kidney: normal size, normal echotexture, no signs of hydronephrosis.

Ultrasound left kidney: normal size, normal echotexture, no signs of hydronephrosis.



CXR Bert van der Veen

Normal position diaphragm, veiling density left lower lobe, no lymphadenopathy.



Results of tests	Value	Normal reference ranges	Avail
Blood			
Erythrocyte sedimentation rate (mm/h)	100	0–20	1
C-reactive protein (mg/L)	150	0.2-3.0	1
White blood cell count (× 10³/mm³)	10.5	3.5–9.1	1
Differential Cell Count (× 10³/mm³)			1
Eosinophils	0.3	0-0.6	1
Basophils	0.2	0-0.2	1
Band forms (/µL)	500	<700	1
Neutrophils	4.2	1.4–4.6	1
Lymphocytes	3.3	0.7–4.6	1
Monocytes	0.6	0.2–0.7	1
Hemoglobin (g/dL)	10.8	12–15.8	1
Hematocrit (percent)	42.0	35.4–44.4	1
Mean corpuscular volume ($ imes$ m ³)	90	79–93	1
Alanine transaminase (U/L)	15	7–41	1
Aspartate transaminase (U/L)	18	12–38	1
Gamma-glutamyl transpeptidase (U/L)	12	9–58	1
Alkaline phosphatase (U/L)	45	33–96	1
Sodium (mEq/L)	134	136–146	1
Potassium (mEq/L)	5.1	3.5-5.0	1
Creatinine (mg/dL)	2.12	0.5–0.9	1
Urea nitrogen (mg/dL)	70	7–20	1
Glomerular filtration rate (GFR) (mL/min)	93	106–132	1
Platelet (× 10 ⁹ /L)	600	165–415	1
Lactate dehydrogenase (U/L)	420	115–221	1
Bilirubin, total (mg/dL)		0.3–1.3	1
Bilirubin, conjugated (mg/dL)		0.1–0.4	1
Blood Glucose test (mg/dL)		65–100	1
Troponin I (ng/mL)		0-0.8	1
Thyroid-stimulating hormone (μlU/mL)		0.34-4.25	1
Thyroxine, free (ng/dL)		0.8–2.7	1
Prothrombin time (sec)	14.1	12.7–15.4	2
Partial thromboplastin time, activated (sec)	35	26.3–39.4	2
Thrombin time (sec)	15.6	14–16	2
D-dimer (mg/L)	0.4	<0.5	2
Plasma fibrinogen (g/L)		1.5–3	2
Antigliadin antibodies			2

Laboratory results patient Bert van der Veen

Results of tests	Value	Normal reference ranges	Avail.
ANA	Neg.		2
ANCA			4
Rheumatoid factors			2
Acetylcholine receptor antibody			2
Anti-MuSK			2
Ca 125 (carbohydrate antigen 125)			2
CEA (carcinoembryonic Antigen)			2
Urine			
Protein (mg/dL)	200	Neg	1
RBC's (/µL)	35	<10	1
WBC's (/µL)	42	<25	1
Glucose	Neg	Neg	1
Ketones	Neg	Neg	1
рН	5.1	4.5-8.0	1
Nitrite	Neg	neg	1
Dysmorphic red cells	Pos		2
Blood gas			
рН	7.39	7.38–7.42	3
pO ₂ (mmHg)	64	70–100	3
pCO ₂ (mmHg)	51	38–42	3
Bicarbonate (mEq/L)	28	22–26	3
O ₂ -saturation (fraction)	91	96–100	3

Legend availability

- 1. In 'real-life' within approximately 1 hour (during assessment, because of time constraints, after about 30–45 minutes).
- 2. Takes a long time, therefore candidates receive these results near the end of the study period, after circa 2-2.5 hours.
- Candidates may request this later on and it will take some time before someone arrives who is able to perform the test (most doctors do this themselves). The results will be available within 5 minutes after blood taking (during assessment, because of time/ organisation constraints, after about 30–45 minutes).
- 4. c-ANCA: pos. This result will not be given to the candidates during the assessment, due to the time the test takes.

Role descriptions disturbances for this case

New episode of coughing up blood

Description of disturbance. The nurse enters and asks for the candidate. Mr. Van der Veen has just been coughing up blood again. This happended 5 minutes ago. It was a trace of blood in the mucus he has been coughing up.

Solution. The candidate can act in a couple of ways:

- He/she wants to see the patient again (not possible, the patient is away for further examinations).
- He/she wants to consult the supervisor.
- He/she needs a moment to look up what he/she could / is supposed to do.
- He/she immediately states there not much he/she can do about it right now.

The nurse can tell the patient is feeling well for the rest. The patient is away now for further examinations, so it won't be possible to see him at the moment.

It should be possible for the candidate to consult his/her supervisor. The supervisor indicates that nothing can be done at the moment, but the tests have to be pushed ahead.

In case of candidate looking up information: the nurse asks when she will get an answer?

If he/she states there is not much that can be done, the nurse asks: it looks quite serious though. Are the test results in yet?

Duration. Including possible consult or looking up information max. 10–15 minutes.

Role Nurse. you enter the study room and ask the candidate to come with you. Outside, you tell that mister Van der Veen has just been coughing up some blood again. He was having a kind of coughing fit and produced only mucus at first. But, at a certain point, there was a trace of blood in the mucus as well (the total amount of blood after several coughing fits was about a teespoon full). Should something be done? The patient is now calm again. Vital signs are all good, blood pressure 130/85 mmHg, pulse 85/min regular, saturation has been low for a moment during the fit of coughing (89% with 1.5L O_2) but is now back at 95% (with 1.5L O_2).

The patient is away for further examinations (before, take a moment to check which tests the candidate requested for this patient), therefore it is not possible for the candidate to see him at the moment.

The aim is for the candidate to feel some more pressure to resolve the case of this patient. He/she doesn't need to act on the coughing up of the blood itself.

Try to direct the candidate to find out the cause of the haemoptysis quickly, by making remarks like: "gosh, it should be nice if we would know what is wrong with this patient, shouldn't it?" etc.

If the candidate tells you that he will consult his/her supervisor, you are fine with that. Ask when he/she will get back to you with an answer though, to build up the time pressure a little more.

Role Supervisor. Possibly, the candidate will call you to ask if he/ she should act on the coughing up of blood. Because of the amount (relatively little, doesn't seem to be an active bleeding) no action is required. It is important though for the nurses to monitor this patient closely and to ensure that all results will be in quickly, to be able to formulate a plan of action.

Disturbance by clerk

Description of disturbance. A clerk (junior student) enters and asks you to confer with you on something. She would like to give a presentation about the patient with haemoptysis (Bert van der Veen), who is hospitalized at the Medium Care. Do you have some more information already about the DD for this patient and some additional literature?

Possible solutions. The candidate gives an update on this case and refers to websites where the clerk could find additional information. He/she asks the clerk if she has been thinking about the case herself and if she has any suggestions.

Duration. Including full briefing, about 5–10 minutes.

Role clerk. Before, read the role description Bert van der Veen, the patient with haemoptysis.

You walk into the study room, holding pencil and paper, and ask your resident (candidate) to confer for a moment. It is important to ask him/her to walk out of the room with you, to make sure the other candidates won't overhear your conversation. Next Friday you have got a day of education and then, you tell, you are to give a short talk about an interesting case on the ward. You would like to examine in detail the case of the patient with haemoptysis. Would it be possible for the candidate to inform you about the current situation? Maybe he/she has found some additional literature that you could use as well?

This case is about a patient with Wegener's granulomatosis. This is a rare vasculitis manifesting itself in the small blood vessels especially. The disease is characterized by clumps of inflammatory cells (granulomas) that are mainly located in nose, lungs and kidneys. The skin, joints and the nervous system could be affected as well.

This is a difficult case with a difficult DD. Therefore, the candidates will have difficulties solving this case and they might consult you.

But you are only a third-year medical student and you have got just basic knowledge about differential diagnoses.

Therefore, you might suggest:

- Couldn't it be TB? We are taught that haemoptysis is one of the symptoms.
- Or maybe some other tropical disease? He has been in Zaire for quite some time, hasn't he?
- Or maybe a bleeding disorder? Because he is very tired and coughing up blood? (You are not able to give examples of bleeding disorders yet. Oh yes... isn't haemophilia one of them??)
- And a lung tumor? Has that been ruled out yet? That could cause coughing up blood and fatigue as well, couldn't it?

You take notes of the things the candidate comes up with.

You promise that you will search the literature as well and that you will contact him/her directly if you find out something interesting that could be relevant to this case.

Could the candidate suggest some websites to look at?

In case the candidate suggests that he/she is too busy to talk to you now, you start pushing a little. Tell him/her that is really has to be only a short update and that it would be very helpful to you (sweeping you aside is really not an option. At worst, offer to come back later on, but of course before the end of the self-study period).

Later on, you give the notes you have taken to the person watching over the study room, who will pass them on to the supervisor of the candidate.



Appendix B

Scoring forms UHTRUST

Facets of competences (physicians)

Overall score of this candidate on this Facet of Competence Scoring® © Very weak © Weak © Acceptable © Good © Very Good	Patient 5 A 36-year old man with rheumatoid arthritis and fever. Scoring Weak Acceptable Good	Patient 4 A 65-year old woman with difficulties to speak, swallow and chew. Scoring Image: Weak Acceptable Good	Patient 3 A 58-year old woman with abdominal pain. Scoring* Image: Weak Acceptable Good	Patient 2 A 53-year old man with progressive fatigue and haemoptysis. Scoring Weak Acceptable Good	Patient 1 A 5-year-old girl with weariness and abdominal pain. Scoring Image: Weak Acceptable Image: Good	Facet of Competence 1: Scientific and empirical grounded method of working The physician uses evidence-based procedures whenever possible and relies on scientific knowledge. He searches actively and purposefully for evidence and consults high quality resources. He uses his scientific knowledge critcally and carefully in his work.	Your name	Identification of according to the state
o f Competence ble	l arthritis and fever.	es to speak, swallow and chew.	nal pain.	fatigue and haemoptysis.		rical grounded method of wor Ihenever possible and relies on sci He uses his scientific knowledge I	Your role Supervisor Assessor-entire assessment Sessor-reporting phase	
					(Optional) Explanation of the score you gave	king entific knowledge. He searches actively ar ritcally and carefully in his work.	Candidate"	
					ave	nd purposefully	Please note Field marked with an are mandatory	

The header of the form identifies the candidate the assessor and his or her role.

development. and riskmanagement, and supervisors, (5) and maintaining own of working, (2) Knowing cal grounded method of Competence: (7) Active professional Responsibility, (6) Safety tion with colleagues (4) Verbal communicawork and collegiality, possibilities, (3) Teampersonal bounds and (1) Scientific and empirirepeated for each Facet This part of the form is

Identification of ass Your name	Identification of assessor and candidate Your name	Your role	Candidate	Please note	The header of the form identifies the candidate.
		 Supervisor Assessor-entire assessment Assessor-reporting phase]	Field marked with an are mandatory	the assessor and his or her role.
Please rate each cand	Please rate each candidate for each op the EPA's on a scale of 1–5	in a scale of 1–5			Instruction
EPA 1: Emergency assistance with Case: A 74 year-old female patient, gastritis. At 02.00 AM she complain the last hour, her pulse is 120 bpm.	EPA 1: Emergency assistance with acute cardiac failure Case: A 74 year-old female patient, known with diabetes ty gastritis. At 02.00 AM she complains of chest pain and dizzi the last hour, her pulse is 120 bpm.	EPA 1: Emergency assistance with acute cardiac failure Case: A 74 year-old female patient, known with diabetes type I and ischemic chest pain history is admitted because of suspected gastritis. At 02.00 AM she complains of chest pain and dizziness. The blood pressure has dropped from 140/90 to 95/60 mmHg within the last hour, her pulse is 120 bpm.	ain history is admitted becaus has dropped from 140/90 to 9	e of suspected 5/60 mmHg within	This part of the form is repeated for each EPA: (1) Emergency assis- tance with acute cardiac failure, (2) Handling a pratient complaint (3)
Scoring	 He/she is not able to do this He/she is able to do this und 	 He/she is not able to do this He/she is able to do this under direct supervision 			Pre-operative informa- tion and consent, (4) Breaking bad news, (5)
	 He/she is able to do this He/she is able to do this 	 He/sne is able to do this if supervision is available He/she is able to do this independently 			Clinical reasoning with acute infection, (6) Solve
	© He/she is able to superv	to supervise others in performing this activity			a management prob- lem, (7) Suspicion of self-induced disease, (8) Handling of a seriously ill patient, (9) Interaction with a consultant.
					٦

Entrustable Professional Activities (physicians)

Post Patient Encounter Form (physicians)

Identification of candidat	e and patient Date	Please note Field marked with an are mandatory	Score by rater below expectations meets expectations exceeds expectations
1 Give a summary state- ment about the patient in one sentence (most important features of the patient).			
2 List 3 problems (complaints or symptoms).			
3 Give a three-option differential diagnosis, with most likely on top. Each diagnosis should explain the most im- portant problems.			
4 Select four key his- tory and/or physical examination facts that support the most likely diagnosis.			
5 Give three policy or treatment options, with most suitable on top.			
6 Give three arguments for the most suitable policy or treatment option.			
Identification rater Rater	Space for additiona	l remarks.	Overall score

Identification of Your name	Identification of assessor and candidate Your name	Candidate	Please note Field marked with an [†] are mandatory	The header of the form identifies the candidate, the assessor and his or her role.
Facet of Compet The physician uses for evidence and c	ence 1: Scientific and empir evidence-based procedures w onsults high quality resources.	Facet of Competence 1: Scientific and empirical grounded method of working The physician uses evidence-based procedures whenever possible and relies on scientifi for evidence and consults high quality resources. He uses his scientific knowledge critica	Facet of Competence 1: Scientific and empirical grounded method of working The physician uses evidence-based procedures whenever possible and relies on scientific knowledge. He searches actively and purposefully for evidence and consults high quality resources. He uses his scientific knowledge critcally and carefully in his work.	This part of the form is repeated for each Facet of Competence:
Situation 1 36- Scoring • V	36-year old man: part 1 Weak 2 Acceptable 3 Good	q	(Optional) Explanation of the score you gave	(1) Scientific and empiri- cal grounded method
Situation 2 Pot Scoring 0 V	Potassium level • Weak 2 Acceptable 3 Good	q		of working, (2) Knowing and maintaining own personal bounds and
Situation 3 53- Scoring • V	53-year old man coughing up blood • Weak • Acceptable • Good	pood		possibilities, (3) Team- work and collegiality, (4)
Situation 4 Pat Scoring 1 V	Patient with finger bandage • Weak 2 Acceptable 3 Good	q		Verbal communication with colleagues and supervisors, (5) Respon-
Situation 5 36- Scoring [*] ⁽¹⁾ V	36-year old man: part 2 • Weak © Acceptable © Good	q		sibility, (6) Safety and riskmanagement.
Overall score of t Scoring [*] [®] Very	Overall score of this candidate on this Facet of Competence Scoring © Very weak © Weak © Acceptable © Good © Very Good	f Competence ole © Good © Very Good		

Facets of competences (nurses)

CARE Questionnaire (standardized patients)



References

A

- Aarkrog, V. (2005). Learning in the workplace and the significance of school-based education: a study of learning in a Danish vocational education and training programme. *International Journal of Lifelong Education*, 24, 137–147.
- Albanese, M. A. (2011). Measurement and interpretation challenges in comparing student performance outcomes from different medical schools. *Academic Medicine*, 86, 1073–1075.
- Albanese, M. (2009). Life is tough for curriculum researchers. *Medical Education*, 43, 199–201.
- Albanese, M. (2000). Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. *Medical Education*, *34*, 729–738.
- Albanese, M., Mejicano, G., Mullan, P., Kokotailo, P., & Gruppen, L. (2008). Defining characteristics of educational competencies. *Medical Education*, 42, 248–255.
- Albano, M. G., Cavallo, F., Hoogenboom, R., Magni, F., Majoor, G., Manenti, F., ... Van der Vleuten, C. (1996). An international comparison of knowledge levels of medical students: the Maastricht Progress Test. *Medical Education*, 30, 239–245.
- Albert, M., Hodges, B., & Regehr, G. (2007). Research in medical education: balancing service and science. *Advances in Health Sciences Education*, 12, 103–115.
- Amin, Z., Merrylees, N., Hanif, A., & Talukder, H. K. (2008). Medical education in Bangladesh. *Medical Teacher*, 30, 243–247.
- Arnold, L. (2002). Assessing Professional Behavior: Yesterday, Today and Tomorrow. Academic Medicine, 77, 502–515.

В

- Baartman, L. K. J., & De Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review*, 6, 125–134.
- Babbot, S. (2010). Watching closely at a distance: key tensions in supervising resident physicians. *Academic Medicine*, 85, 1399–1400.
- Bakker, M. (2008). Design and evaluation of video portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teachers. Leiden, The Netherlands: Mostert & Van Onderen.
- Barnsley, L., Lyon, P. M., Ralston, S. J., Hibbert, E. J., Cunningham, I., Gordon, F. C., & Field, M. J. (2004). Clinical skills in junior medical officers: a comparison of selfreported confidence and observed competence. *Medical Education*, 38, 358–367.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9–30.
- Barton J. R, Corbett, S., & Van der Vleuten, C. P. (2012). The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. *Gastrointestinal Endoscopy*, 75, 591–597.
- Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. De Corte, L. Verschaffel, N. Entwistle, & J. van Merriënboer (Eds.), Unravelling basic components and dimensions of powerful learning environments. EARLI Advances in learning and instruction series. Oxford, England: Elsevier Science.

- Billett, S. (2001). Knowing in practice: Re-conceptualizing vocational expertise. *Learning and Instruction*, 11, 431–452.
- Billet, S. (1995). Workplace learning: its potential and limitations. *Education and Training*, 37 (5), 20–27.
- Birenbaum, M. & Dochy, F. (Eds.). (1996). Alternatives in assessment of achievement, learning processes and prior knowledge. Boston, MA: Kluwer Academic.
- Bleakley, A., & Brennan, N. (2011). Does undergraduate curriculum design make a difference to readiness to practice as a junior doctor? *Medical Teacher*, 33, 459–467.
- Bleker, O. P., & Blijham, G. H. (1999). Too old, too smart and too expensive. *Medisch Contact*, 54, 1201–1203 [in Dutch].
- Bleker, O. P., Ten Cate, Th. J., & Holdrinet, R. S. G. (2004). General competencies of the future medical specialist. *Dutch Journal of Medical Education*, 23, 4–14 [in Dutch].
- Bligh, J. & Brice, J. (2010). Course Design. In P. Cantillon, & D. Wood (Eds.), ABC of learning and teaching in medicine. Chichester, United Kingdom: Wiley-Blackwell.
- Bligh, J., Prideaux, D., & Parsell, G. (2001). PRISMS: new educational strategies for medical education. *Medical Education*, 35, 520–521.
- Boulet, J. R., Norcini, J. J., Whelan, G.P., Hallock, J.A., & Seeling, S. S. (2006). The international medical graduate pipeline; recent trends in certification and residency training. *Health Affairs*, 25, 469–477.
- Boursicot, K., & Roberts, T. (2005). How to set up an OSCE. The Clinical Teacher, 2, 16–20.
- Boyce, P., Spratt, C., Davies, M., & McEvoy, P. (2011). Using entrustable professional activities to guide curriculum development in psychiatry training. *BMC Medical Education*, 11, 96.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: a simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100.
- Brennan, N., Corrigan, O., Allard, J., Archer, J., Barnes, R., Bleakley, A., ... Regan de Bere, S. (2010). The transition from medical student to junior doctor: today's experiences of Tomorrow's Doctors. *Medical Education*, 44, 449–458.
- Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 1–16).
 Westport, CT: American Council on Education and Praeger.

С

- Cantillon, P., & Macdermott, M. (2008). Does responsibility drive learning? Lessons from intern rotations in general practice. *Medical Teacher*, 30, 254–259.
- Carraccio, C., & Burke, A. E. (2010). Beyond competencies and milestones: adding meaning through context. *Journal of Graduate Medical Education*, 2, 419–422.
- Carraccio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: from Flexner to competencies. *Academic Medicine*, 77, 361–367.
- Cave, J., Goldrace, M., Lambert, T., Woolf, K., Jones, A., & Dacre, J. (2007). Newly qualified doctors' views about whether their medical school had trained them well: Questionnaire surveys. *BMC Medical Education*, 7 (38), 1–6.

- Cave, J., Woolf, K., Jones, A., & Dacre, J. (2009). Easing the transition from student to doctor: how can medical schools help prepare their graduates for starting work? *Medical Teacher*, 31, 403–408.
- Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), Vygotsky's educational theory in cultural context. Cambridge, NY: Cambridge University Press.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29, 3–13.
- Chittenden, E. H., Henry, D., Saxena, V., Loeser, H., & O'Sullivan, P. (2009). Transitional clerkship: an experiental course based on workplace learning theory. *Academic Medicine*, 84, 872–876.
- Christensen, L. (2004). The Bologna process and medical education. *Medical Teacher*, *26*, 625–629.
- Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision making tasks in education. *Journal of Educational Psychology*, 17, 373–386.
- Cleland, J. A., Abe, K., & Rethans, J. (2009). The use of simulated patients in medical education: AMEE Guide no. 42. *Medical Teacher*, 31, 477–486.
- Cohen, A. S., & Wollack, J. A. (2006). Test administration, security, scoring and reporting. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 17–64). Westport, CT: American Council on Education and Praeger.
- Cohen-Schotanus, J., Muijtjens, A. M., Schönrock-Adema, J., Geertsma, J., & Van der Vleuten, C. P. (2008). Effects of conventional and problem-based learning on clinical and general competencies and career development. *Medical Education*, 42, 256–265.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.
- Cook, D. A., & Beckman, T. J. (2010). Reflections on experimental research in medical education. *Advances in Health Sciences Education*, 15, 455–464.
- Cooke, M., Irby, D. M., & O'Brien, B. (2010). Educating physicians: A call for reform of medical school and residency. San Francisco, CA: Jossey-Bass.
- Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.
- Cronbach, L. J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 443–507). Washington, DC: American Council on Education.
- Crossley, J., Johnson, G., Booth, J., & Wade, W. (2011). Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Medical Education*, 45, 560–569.
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational Measurement* (pp. 621–694). Washington, DC: American Council on Education.

Custers, E. J. F. M., & Ten Cate, O. T. J. (2002). Medical students' attitudes towards and perception of the basic sciences: a comparison between students in the old and the new curriculum at the University Medical Centre Utrecht, the Netherlands. *Medical Education*, 36, 1142–1150.

D

- Dahle, L. O., Brynhildsen, J., Behrbohm Fallsberg, M., Rundquist, I., & Hammar, M. (2002). Pros and cons of vertical integration between clinical medicine and basic science within a problem-based undergraduate medical curriculum: Examples and experiences from Linköping, Sweden. *Medical Teacher*, 24, 280–285.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *Journal of the American Medical Association*, 296, 1094–1102.
- Dawes, J. (2007). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50 (1), 61–77.
- De Loe, R. C. (1995). Exploring complex policy questions using policy Delphi. A multiround, interactive survey method. *Applied Geography*, 15, 53–68.
- Dekker-Groen, A. M., Van der Schaaf, M. F., Stokking, K. M. (2011). Teacher competences required for developing reflection skills of nursing students. *Journal of Advanced Nursing*, 67, 1568–1579.
- Dent, J. A., Angell-Preece, H. M, Mei-Ling Ball, H., & Ker, J.S. (2001). Using the Ambulatory Care Teaching Centre to develop opportunities for integrated learning. *Medical Teacher*, 23, 171–175.
- Dijksterhuis, M. G. K., Teunissen, P. W., Voorhuis, M., Schuwirth, L. W. T., Ten Cate, T. J., Braat, D. D. M., & Scheele, F. (2009). Determining competence and progressive independence in postgraduate clinical training. *Medical Education*, 43, 1156–1165.
- Distlehorst, L. H., Dawson, B. K., Klamen, D. L. (2009). Supervisor and self-ratings of graduates from a medical school with a problem-based learning and standard curriculum track. *Teaching and learning in medicine*, 21, 291–298.
- Dornan, T., Boshuizen, H. P., King, N., & Scherpbier, A. J. (2007). Experience-based learning: A model linking the processes and outcomes of medical students' workplace learning. *Medical Education*, 41, 84–91.
- Dornan, T., & Bundy, C. (2004). Learning in practice: What can experience add to early medical education? Consensus survey. *British Medical Journal*, 329, 834–837.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* Oxford, England: Basil Blackwell.
- Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science Technology & Society*, 24, 177–181.
- Driessen, E., Van Tartwijk, J., Vermunt, J. D., & Van der Vleuten, C. (2003). Use of portfolio's in early undergraduate medical training. *Medical Teacher*, 25, 18–23.
- Dunn, W. R., Hamilton, D. D., & Harden, R. M. (1985). Techniques of identifying competences needed of doctors. *Medical Teacher*, 7, 15–25.
- Dunphy, B. C., & Williamson, S. L. (2004). In pursuit of expertise. Toward an educational model for expertise development. Advances in Health Sciences Education, 9, 107–127.

- Durning, S. J., Artino, A., Boulet, J., La Rochelle, J., Van der Vleuten, C., Arze, B., & Schuwirth, L. (2012). The feasibility, reliability and validity of a post-encounter form for evaluating clinical reasoning. *Medical Teacher*, 34, 30–37.
- Dutch Federation of University Medical Centers (2009). *Blueprint of objectives for medical schools*. Utrecht, The Netherlands: NFU [in Dutch].
- Duvivier, R. J., Van Geel, K., Van Dalen, J., Scherpbier, A. J. J. A., & Van der Vleuten, C. P.
 M. (2012). Learning physical examination skills outside timetabled training sessions: what happens and why? *Advances in Health Sciences Education*, 17 (3), 339–355.
- Duvivier, R. J., Van Dalen, J., Muijtjens, A. M., Moulaert, V., Van der Vleuten, C., & Scherpbier, A. (2011). The role of deliberate practice in the acquisition of clinical skills. *BMC Medical Education*, 11, 101.
- Dwyer, C. A. (1995). Criteria for performance-based teacher assessments: validity, standards and issues. In A. J. Shinkfield, & D. Stufflebeam (Eds.), *Teacher evaluation:* guide to effective practice (pp. 62–80). Boston, MA: Kluwer Academic.

Ε

- Epstein, R. M. (2007). Assessment in Medical Education. New England Journal of Medicine, 356, 387–396.
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, *2*, 226–235.
- Eraut, M. (1994). *Developing professional knowledge and competence*. Abingdon, United Kingdom: RoutledgeFalmer.
- Ericsson, K. A., Charness, N., Hoffman, R. R., & Feltovich, P. J. (2006). *The Cambridge handbook of expertise and expert performance*. New York, NY: Cambridge University Press.
- Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Eva, K. W., & Regehr, G. (2011). Exploring the divergence between self-assessment and self-monitoring. *Advances in Health Sciences Education*, 16, 311–329.
- Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: a new conception of selfassessment ability. Academic Medicine, 82, 81–84.

F

- Fallon, D., Lahar, C. J., & Susman, D. (2009). Taking the high road to transfer: building bridges between English and psychology. *Teaching English in the Two-Year College*, 37, 41–55.
- Foley, R. P., Polson, A. L., & Vance, J. M. (1997). Review of the literature on PBL in the clinical setting. *Teaching and Learning in Medicine*, 9, 4–9.
- Frank, J. R., Snell, L. S., Ten Cate, O., Holmboe, E., Carraccio, C., Swing, S., ... Harris, K.A. (2010). Competency-based medical education: theory to practice. *Medical Teacher*, 32, 638–645.
- Frank, J. R. (Ed.) (2005). The CanMEDS 2005 Physician Competency Framework. Better standards. Better physicians. Better care. Ottawa, Canada: The Royal College of Physicians and Surgeons of Canada.

- Frank, J. R., Jabbour, M., & Tugwell, P. (1996). Skills for the new millennium: Report of the societal needs working group, CanMEDS 2000 Project. Annual Report, College of Physicians & Surgeons of Canada, 29, 206–216.
- Fraser, S. W., & Greenhalgh, T. (2001). Coping with complexity: educating for capability. *British Medical Journal*, 323, 799–803.
- Freidson, E. (1970). *Profession of medicine: A study of the sociology of applied knowledge*. New York, NY: Dodd, Mead & Company.

G

- Garcia-Perez, M. A., Amaya, C., & Otero, A. (2007). Physicians' migration in Europe: an overview of the current situation. *BMC Health Services Research*, 7, 201.
- General Medical Council (2009). *The new doctor: guidance on foundation training*. London, United Kingdom: General Medical Council.
- Gick, M., & Holyoak, K. J. (1983). Schema introduction and analogical transfer. *Cognitive* psychology, 15, 1–35.
- Gingerich, A., Regehr, G., & Eva, K. W. (2011). Rater-based assessments as social judgments: rethinking the etiology of rater errors. Academic Medicine, 86, 1–7.
- Ginsburg, S. (2011). Respecting the expertise of clinician assessors: construct alignment is one good answer. *Medical Education*, 45, 546–548.
- Ginsburg, S., McIlroy, J., Oulanova, O., Eva, K., & Regehr, G. (2010). Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Academic Medicine*, 85, 780–786.
- Gipps, C. V. (1994). Beyond testing. Towards a theory of educational assessment. London, England: RoutledgeFalmer.
- Goldacre, M. J., Lambert, T., Evans, J., & Turner, G. (2003). Preregistration house officers' views on whether their experience at medical school prepared them well for their jobs: national questionnaire survey. *British Medical Journal*, 326, 1011–1012.
- Govaerts, M. J., Van der Vleuten, C. P. M., Schuwirth, L. W. T., & Muijtjens, A. M. M. (2007). Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. Advances in Health Sciences Education, 12, 239–260.

Н

- Haller, G., Myles, P. S., Taffé, P., Perneger, T. V., & Wu, C. L. (2009). Rate of undesirable events at beginning of academic year: retrospective cohort-study. *British Medical Journal*, 339, b3974.
- Hallock, J. A., McKinley, D. W., & Boulet, J. R. (2007). Migration of doctors for undergraduate medical education. *Medical Teacher*, 29, 98–105.
- Hammond, K. R., Kern, F., Crow, W. J., Githens, J. H., Groesbeck, B., Gyr, J. W., & Saunders, L.H. (1959). *Teaching comprehensive medical care: A psychological study of a change in medical education*. Cambridge, MA: Harvard University Press.
- Hannafin, M. J. (1984). Guidelines for using locus of instructional control in the design of computer-assisted instruction. *Journal of Instructional Development*, 7, 6–10.
- Harden, R. M. (2006). International medical education and future directions: a global perspective. *Academic Medicine*, 81, 22–29.

- Harden, R. M. & Gleeson F. A. (1979). Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*, 13, 41–54.
- Harden, R. M., Sowden, S., & Dunn, W. R. (1984). Educational strategies in curriculum development: the SPICES model. *Medical Education*, 18, 284–297.
- Harendza, S. (2011). "HUS" diary of a German nephrologist during the current EHEC outbreak in Europe. *Kidney International*, 80, 687–689.
- Hawkins, R. E, Katsufrakis, P. J., Holtman, M. C., & Clauser, B. E. (2009). Assessment of medical professionalism: who, what, when, where, how, and ... why? *Medical Teacher*, 31, 348–361.
- Hayes, K., Feather, A., Hall, A., Sedgwick, P., Wannan, G., Wessier-Smith, A., ... McCrorie, P. (2004). Anxiety in medical students: is preparation for full-time clinical attachments more dependent upon differences in maturity or on educational programmes for undergraduate and graduate entry students? *Medical Education*, 38, 1154–1163.
- Herling, R. W. (2000). Operational Definitions of Expertise and Competence. Advances in Developing Human Resources, 2 (1), 8–21.
- Hillen, H.F.P. (Ed.) (2012). Medical Education in The Netherlands—State of the Art 2012. Utrecht, The Netherlands: Quality Assurance Netherlands Universities [in Dutch, draft version].
- Hirsh, D., Gaufberg, E., Ogur, B., Cohen, P., Krupat, E., Cox, M., ... Bor, D. (2012).
 Educational Outcomes of the Harvard Medical School-Cambridge Integrated
 Clerkship: A way forward for medical education. *Academic Medicine*, 87, 643–650.
- Holmboe, E. S., & Hawkins, R. E. (Eds.) (2008). *Practical guide to the evaluation of clinical competence*. Philadelphia, PA: Mosby-Elsevier.
- Holmboe, E. S., Ward, D. S., Reznick, R. K., Katsufrakis, P. J., Leslie, K. M., Patel, V. L., ... Nelson, E. A. (2011). Faculty development in assessment: the missing link in competency-based medical education. *Academic Medicine*, 86, 460–467.
- Howley, L. D. (2004). Performance assessment in medical education: where we've been and where we're going. *Evaluation & the Health Professions*, 27, 285–301.

I

Ibrahim, M. (2007). Medical education in Nigeria. *Medical Teacher*, 29, 901–905.

Ineson, S., & Seeling, S. S. (2005). The medical passport. *The Journal of Continuing Education in the Health Professions*, 25, 30–33.

J

- Jen, M. H., Bottle, A., Majeed, A., Bell, D., & Aylin, P. (2009). Early in-hospital mortality following trainee doctors' first day at work. *Public Library of Sciences One*, 4 (9), e7103.
- Jones, M. D., Rosenberg, A., Gilhooly, J., Carraccio, C. (2011). Competencies, outcomes controversy. *Academic Medicine*, 86, 161–165.

Κ

Kamalski, D. M. A., Ter Braak, E. W. M. T., Ten Cate, Th. J., & Borleffs, J. C. C. (2007). Early clerkships. *Medical Teacher*, 29, 915–920.

- Kane, M. (2006). Validation. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp.17–64). Westport, CT: American Council on Education and Praeger.
- Kane, M. (2004). Certification testing as an illustration of argument-based validation. Measurement: Interdisciplinary Research & Perspective, 2, 135–170.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535.
- Karle, H. (2008). International recognition of basic medical education programmes. Medical Education, 422, 12–17.
- Karle, H. (2008). World Federation for Medical Education Policy on international recognition of medical schools' programme. *Annals Academy of Medicine Singapore*, 37, 1041–1043.
- Kaufman, D. M., & Mann, K. V. (1999). Achievement of students in a conventional and problem-based learning (PBL) curriculum. Advances in Health Sciences Education, 4, 245–260.
- Kearney, R. A. (2005). Defining professionalism in anaesthesiology. *Medical Education*, 39, 769–776.
- Keijsers, C. J. P. W., Custers, E. J. F. M., & Ten Cate, O. Th. J. (2009). A new, problem oriented medicine curriculum in Utrecht: less basic science knowledge. *Nederlands Tijdschrift voor Geneeskunde*, 153, 1654–1659 [in Dutch].
- Kennedy, T. J. T., Regehr, G., Baker, G. R., & Lingard, L. (2008). Point-of-care assessment of medical trainee competence for independent clinical work. *Academic Medicine*, 83 (Suppl. 10), 89–92.
- Kennedy, T. J., Regehr, G., Baker, G. R., & Lingard, L. (2009). Preserving professional credibility: grounded theory study of medical trainees' requests for clinical support. *British Medical Journal*, 338, b128.
- Kern, D. E., Thomas, P. A., Howard D. M., & Bass, E. B. (1998). Curriculum development for medical education: a six-step approach. Baltimore, MD: The John Hopkins University Press.
- Kilminster, S., Cottrell, D., Grand, J., & Jolly, B. (2007). AMEE Guide No. 27: Effective educational and clinical supervision. *Medical Teacher*, 29, 2–19.
- Kilminster, S., Zukas, M., Quinton, N., & Roberts, T. (2010). Learning practice? Exploring the links between transitions and medical performance. *Journal of Health* Organization and Management, 24, 556–570.
- Kinzie, M. B., Sullivan, H. J., & Berdel, R. L. (1988). Learner control and achievement in science computer-assisted instruction. *Journal of Educational Psychology*, 80, 299–303.
- Knight, P. T. (2002). The Achilles' heel of quality: the assessment of student learning. Quality in Higher Education, 8, 107–115.
- Koens, F., Mann, K. V., Custers, E. J. F. M., & Ten Cate, O. Th. J. (2005). Analysing the concept of context in medical education. *Medical Education*, 39, 1243–1249.
- Kogan, J. R., Holmboe, E. S., Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees. A systematic review. *Journal of the American Medical Association*, 302, 1316–1326.
- Kreiter, C. D., & Bergus, G. (2008). The validity of performance-based measures of clinical reasoning and alternative approaches. *Medical Education*, 43, 320–325.

Kurdak, H., Altinsaş, D., & Doran, F. (2008). Medical education in Turkey: past to future. *Medical Teacher*, 30, 768–773.

L

- Lane, S., & Stone, C. A. (2006). Performance assessment. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 387–432). Westport, CT: American Council on Education and Praeger.
- Lave, J., & Wenger, E. (1991). Situated learning: legimate peripheral participation. Cambridge, NY: Cambridge University Press.
- Leinster, S. (2009). The undergraduate curriculum. In J.A. Dent, & R.M. Harden (Eds.), A practical guide for medical teachers (3rd ed.) (pp. 17–22). Philadelphia, PA: Elsevier.
- Lempp, H., Cochrane, M., Seabrook, M., & Rees, J. (2004). Impact of educational preparation on medical students in transition from final year to PRHO year: a qualitative evaluation of final-year training following the introduction of a new Year 5 curriculum in a London medical school. *Medical Teacher*, 26, 276–278.
- Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009).
 Evolution of Wenger's concept of community of practice. *Implementation Science*, 4 (11), DOI: 10.1186/1748-5908-4-11.
- Lie, N. (1995). Traditional and non-traditional curricula. Definitions and terminology. *Tidsskrift For Den Norske Laegeforening*, 115, 1067–1071.
- Linstone, H. A., Turoff, M. (Eds.) (1977). *The Delphi Method: Techniques and applications*. London, England: Addison-Wesley Publishing Company.
- Littlewood, S., Ypinazar, V., Margolis, S. A., Scherpbier, A., & Spencer, J. (2005). Early practical experiences and the social responsiveness of clinical education: systematic review. *British Medical Journal*, 331, 387–391.
- Lurie, S. J. (2012). History and practice of competency-based assessment. *Medical Education*, 46, 49–57.
- Lurie, S. J., & Mooney, C. J. (2010). Relationship between clinical assessment and examination scores in determining clerkship grade. *Medical Education*, 44, 177–183.
- Lurie, S. J., Mooney, C. J., & Lyness, J. M. (2009). Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Academic Medicine*, 84 (3), 301–309.
- Lurie, S. J., Mooney, C. J., & Lyness, J. M. (2011). Commentary: pitfalls in assessment of competency-based educational objectives. *Academic Medicine*, 86, 412–414.

Μ

- Margetson, D. B. (1999). The relation between understanding and practice in problembased medical education. *Medical Education*, 33, 359–364.
- Marzano, R. (2000). *A new era of school reform: going where the research takes us*. Aurora, CO: Mid-continent Research for Education and Learning.
- Mazotti, L., O'Brien, B., Tong, L., & Hauer, K. E. (2011). Perceptions of evaluation in longitudinal versus traditional clerkships. *Medical Education*, 45, 464–470.

- Mercer, S. W., Maxwell, M., Heaney, D., & Watt, G. C. M. (2004). The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice*, 21, 699–705.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp.13–104). New York, NY: American Council on Education and Macmillan.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65, 63–67.
- Modernising Medical Careers. Retrieved from *http://www.mmc.nhs.uk/* (November 22, 2011)
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, 23, 5–12.
- Moss, F., & McManus, I. C. (1992). The anxieties of new clinical students. *Medical Education*, 26, 17–20.
- Mylopoulos, M. & Regehr, G. (2009). How student models of expertise and innovation impact the development of adaptive expertise in medicine. *Medical Education*, 43, 127–132.

Ν

- Newble, D. (2004). Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education*, 38, 199–203.
- Nijveldt, M. (2007). Validity in teacher assessment. An exploration of the judgement processes of assessors. Enschede, The Netherlands: Gildeprint.
- Nikendei, C., Weyrich, P., Jünger, J., & Schrauth, M. (2009). Medical education in Germany. *Medical Teacher*, 31, 591–600.
- Norcini, J. J., Blank, L. L., Arnold, G. K., & Kimball, H. R. (1995). The Mini-CEX (Clinical Evaluation Exercise): A preliminary investigation. *Annals of Internal Medicine*, 123, 795–799.
- Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*, 29, 855–871.
- Norcini, J. J., Holmboe, E. S., & Hawkins, R. E. (2008). Evaluation challenges in the era of outcomes-based education. In Holmboe, E. S., & Hawkins, R.E. (Eds.), *Practical guide to the evaluation of clinical competence* (1st ed.) (pp. 1–9). Philadelphia, PA: Mosby Elsevier.
- Norman, G. (2010). Likert scales. Levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, 15, 625–632.
- Norman, G. (2009). Teaching basic science to optimize transfer. *Medical Teacher*, 31, 807–811.
- Norman, G. (2003). RCT = results confounded and trivial: the perils of grand educational experiments. *Medical Education*, 37, 582–584.
- Norman, G. R. & Schmidt, H. G. (1992). The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine*, 67, 557–565.

0

Ogur, B., Hirsch, D., Krupat, E., & Bor, D. (2007). The Harvard Medical School-Cambridge Integrated Clerkship: an innovative model of clinical education. *Academic Medicine*, 82, 397–404.

Ρ

- Palés, J., & Gual, A. (2008). Medical education in Spain: current status and new challenges. *Medical Teacher*, 30, 365–369.
- Papa, F.J., & Harasym, P. H. (1999). Medical curriculum reform in North America, 1765 to the present: a cognitive science perspective. *Academic Medicine*, 74, 154–164.
- Patricio, M., De Burbure, C., Costa, M. J., Schirlo, C., & Ten Cate, O. (2012). Bologna in medicine anno 2012: Experiences of European medical schools that implemented a Bologna two-cycle curriculum—An AMEE Medine2 survey. *Medical Teacher*, 34, 821–832.
- Patricio, M., Den Engelsen, C., Tseng, D., & Ten Cate, O. (2008). Implementation of the Bologna two-cycle system in medical education: where do we stand in 2007? Results of an AMEE-MEDINE survey. *Medical Teacher*, 30, 597–605.
- Poncelet, A., Bokser, S. Calton, B., Hauer, K. E., Kirsch, H., Jones, T., ... Robertson, P. (2011). Development of a longitudinal integrated clerkship at an academic medical center. *Medical Education Online*, 16, 5939.
- Prideaux, D. (2009). Integrated learning. In J. A. Dent, & R. M. Harden (Eds.), A practical guide for medical teachers (3rd ed.) (pp. 181–186). Philadelphia, PA: Elsevier.
- Prideaux, D. (2009). Medical education in Australia: Much has changed but what remains? *Medical Teacher*, 31, 96–100.
- Prince, K. J. A. H., Van de Wiel, M. W. J., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., & Boshuizen, H. P. A. (2000). A qualitative analysis of the transition from theory to practice in undergraduate training in a PBL medical school. *Advances in Health Sciences Education*, 5, 105–116.

R

- Regehr, G. & Norman, G. R. (1996). Issues in cognitive psychology: Implications for professional education. *Academic Medicine*, 71, 988–1001.
- Reis, S., Borkan, J. M., Weingarten, M. (2009). The current state of basic medical education in Israel: Implications for a new medical school. *Medical Teacher*, 31, 984–989.
- Ren, X., Yin, J., Wang, B., Schwarz, M. R. (2008). A descriptive analysis of medical education in China. *Medical Teacher*, 30, 667–672.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation on intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.

S

- Sackman, H. (1975). *Delphi critique. Expert opinion, forecasting, and group process.* Massachusetts: Lexington Books.
- Salomon, G. & Perkins, D. N. (1989). Rocky roads to transfer: rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24, 113–142.

- Schaap, H., Van der Schaaf, M. F., & De Bruijn, E. (2011). Development of students' personal professional theories in senior secondary vocational education. *Evaluation* & *Research in Education*, 24, 81–103.
- Scheerens, J. (1997). Conceptual models and theory-embedded principles on effective schooling. *School Effectiveness and School Improvement*, 8, 269–310.
- Schijven, M. P., Reznick, R.K., Ten Cate, O. Th. J., Grantcharov, T.P., Regehr, G., Satterthwaite, L., ... MacRae, H. M. (2010). Transatlantic comparison of the competence of surgeons at the start of their professional career. *British Journal of Surgery*, 97, 443–449.
- Schmidt, H. G., Cohen-Schotanus, J., & Arends, L. R. (2009). Impact of problem-based, active learning on graduation rates for 10 generations of Dutch medical students. *Medical Education*, 43, 211–218.
- Schmidt, H. G., Machiels-Bongaerts, M., Hermans, H., Ten Cate, T. J., Venekamp, R., & Boshuizen, H. P. A. (1996). The development of a diagnostic competence: comparison of a problem-based, an integrated, and a conventional medical curriculum. *Academic Medicine*, 71, 658–664.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: theory and implications. *Academic Medicine*, 65, 611–621.
- Schmidt, H. G., Vermeulen, L., & Van der Molen, H. T. (2006). Longterm effects of problem-based learning: a comparison of competencies acquired by graduates of a problem-based and a conventional medical school. *Medical Education*, 40, 562–567.
- Schuwirth, L., Colliver, J., Gruppen, L., Kreiter, C., Mennin, S., Onishi, H., ... Wagner-Menghin, M. (2011). Research in assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, 33, 224–233.
- Schwarz, M. R. (2001). Globalization and medical education. *Medical Teacher*, 23, 533–534.
- Schwarz, M. R., & Wojtczak, A. (2002). Global minimum essential requirements; a road towards competence-oriented medical education. *Medical Teacher*, 24, 125–129.
- Schwarz, M. R., Wojtczak, A., & Zhou, T. (2004). Medical education in China's leading medical schools. *Medical Teacher*, 26, 215–222.
- Scicluna, H. A., Grimm, M. C., O'Sullivan, A. J., Harris, P., Pilotto, L. S., Jones, P. D., & McNeil, H. P. (2012). Clinical capabilities of graduates of an outcomebased integrated medical program. *BMC Medical Education*, 12, DOI:10.1186/1472-6920-12-23
- Seguin, C., Jouquan, J., Hodges, B., Bréchat, P. H., David, S., Maillard, D., ... Bertrand, D. (2007). Country report: medical education in France. *Medical Education*, 41, 295–301.
- Sood, R. (2008). Medical education in India. Medical Teacher, 30, 585–591.
- Sterkenburg, A. J., Barach, P., Kalkman, C. J., Gielen, M., & Ten Cate, O. T. J. (2010). When do supervising physicians decide to entrust residents with unsupervised tasks? *Academic Medicine*, 85, 1408–1417.
- Sultana R. G. (2009). Competence and competence frameworks in career guidance: complex and contested concepts. *International Journal for Educational and Vocational Guidance*, 9, 15–30.
- Suzuki, Y., Gibbs, T., & Fujisaki, K. (2008). Medical education in Japan: a challenge to the healthcare system. *Medical Teacher*, 30, 846–850.

T

- Tavares, W., & Eva, K. W. (2012). Exploring the impact of mental workload on raterbased assessments. Advances in Health Sciences Education: Theory and Practice, DOI: 10.1007/s10459-012-9370-3.
- Teichler, U. (2003). Mutual recognition and credit transfer in Europe: experiences and problems. *Journal of Studies in International Education*, 7, 312.
- Ten Cate, O. (2007). Medical education in The Netherlands. *Medical Teacher, 28*, 752–757.
- Ten Cate O. (2006). Trust, competence and the supervisor's role in postgraduate training. *British Medical Journal*, 333, 748–751.
- Ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education*, 39, 1176–1177.
- Ten Cate, O., & Scheele, F. (2007). Competence-based postgraduate training: Can we bridge the gap between educational theory and clinical practice? *Academic Medicine*, 82, 542–547.
- Ten Cate, O., Snell, L., & Carraccio, C. (2010). Medical competence: The interplay between individual ability and the health care environment. *Medical Teacher*, 32, 669–675.
- Ten Cate, O., Snell, L., Mann, K., & Vermunt, J. (2004). Orienting teaching toward the learning process. *Academic Medicine*, 79, 219–228.
- Turnbull, J., & Van Barneveld, C. (2002). Assessment of clinical performance: intraining evaluation. In G. R. Norman, C. P. M. Van der Vleuten, D. I. Newble (Eds.), *International handbook of research in medical education* (pp.793–810). Dordrecht, The Netherlands: Kluwer Academic.
- Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review*, 3, 130–154.

V

- Van der Schaaf, M. F., & Stokking, K. M. (2011). Construct validation of content standards for teaching. Scandinavian Journal of Educational Research, 55, 273–289.
- Van der Velden, L. F. J., & Hingstman, L. (2003). *The medical training continuüm: What makes it so long.* Utrecht, The Netherlands: Nivel [in Dutch].
- Van der Vleuten, C. P. M. (1996). The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education*, 1, 41–67.
- Van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 703–719.
- Van Hell, E. A., Kuks, J. B. M., Schönrock-Adema, J., Van Lohuizen, M. T., & Cohen-Schotanus, J. (2008). Transition to clinical training: influence of pre-clinical knowledge and skills, and consequence for clinical performance. *Medical Education*, 42, 830–837.
- Van Merriënboer, J. J. G., Van der Klink, M. R., & Hendriks, M. (2002). Competencies: From complications toward agreement. Utrecht, The Netherlands: Educational Council of the Netherlands [in Dutch].

- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9, 257–280.
- Vidic, B., Weilauf, H. M. (2002). Horizontal and vertical integration of academic disciplines in the medical school curriculum. *Clinical Anatomy*, 15, 233–235.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: MIT Press.

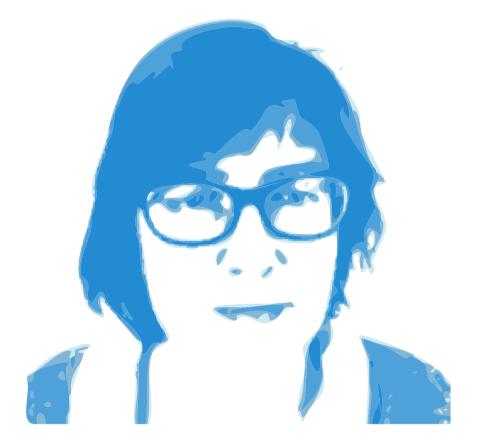
W

- Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: current state of the art. *Advances in Health Sciences Education*, 7, 63–80.
- Wass, V., & Archer, J. (2011). Assessing learners. In T. Dornan, K. Mann, A. Scherpbier,
 & Spencer J. (Eds.), *Medical Education: theory and practice* (pp.229–255). Toronto,
 Canada: Churchill Livingstone Elsevier.
- Wass, V., Van der Vleuten, C., Shatzer, J., & Jones, R. (2001). Assessment of Clinical Competence. *Lancet*, 357, 945–949.
- Watmough, S., Garden, A., & Taylor, D (2006). Pre-registration house officers' views on studying under a reformed medical curriculum in the UK. *Medical Education*, 40, 893–899.
- Wetzel, A.P. (2012). Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. *Academic Medicine*, 87, 1060–1069.
- Whitehouse, C. R., O'Neil, P., & Dornan, T. (2002). Building confidence for work as house officers: student experience in the final year of a new problem-based curriculum. *Medical Education*, 36, 718–727.
- Wijnen-Meijer, M., Ten Cate, T. J., Rademakers, J. J. D. J., Van der Schaaf, M., & Borleffs, J. C. C. (2009). The influence of a vertically integrated curriculum on the transition to postgraduate training. *Medical Teacher*, 31, 528–532.
- Wijnen-Meijer, M., Ten Cate, O. T. J., Van der Schaaf, M., & Borleffs, J. C. C. (2010). Vertical integration in medical school: effect on the transition to postgraduate training. *Medical Education*, 44, 272–279.
- Wijnen-Meijer, M., Ten Cate, O., Van der Schaaf, M., & Harendza, S. Graduates from vertically integrated curricula. *The Clinical Teacher*. Accepted for publication.
- Wijnen-Meijer, M., Van der Schaaf, M., Booij, E., Harendza, S., Boscardin, C., Van Wijngaarden, J., & Ten Cate, Th. J. An argument-based approach to the validation of UHTRUST: Can we measure how recent graduates can be trusted with unfamiliar tasks? Submitted.
- Wijnen-Meijer, M., Van der Schaaf, M., Nillesen, K., Harendza, S., & Ten Cate, O. Essential facets of competence that enable trust in graduates: A Delphi study among physician educators in the Netherlands. *Journal of Graduate Medical Education*. Accepted for publication.
- Williams, G. (1980). Western Reserve's experiment in medical education and its outcome. New York, NY: Oxford University Press.
- Wittert, G. A., & Nelson, A. J. (2009). Medical Education: revolution, devolution and evolution in curriculum philosophy and design. *Medical Journal of Australia, 191*, 35–37.

Wojtczak, A. (2002). Medical education terminology. Medical Teacher, 24, 357.

Y

- Yeates, P., O'Neill, P., Mann, K., & Eva, K. (2012). Seeing the same thing differently. Mechanisms that contribute to assessor differences in directly-observed performance assessments. Advances in Health Sciences Education: Theory and Practice, DOI: 10.1007/S10459-012-9372-1.
- Young, J. Q., Ranji, S. R., Wachter, R. M., Lee, C. M., Niehaus, B., & Auerbach, A.D. (2011). "July effect": impact of the academic year-end changeover on patient outcomes. *Annals of Internal Medicine*, 155, 309–315.



Dankwoord/Acknowledgements

Toen ik aan dit promotietraject begon, verwachtte ik wel dat ik anderen hierbij nodig zou hebben. Maar dat er zovel mensen by betrokken zouden raken had ik niet kunnen denken. Ik wil op deze plek graag iedereen bedanken die op de een of andere manier heeft geholpen om dit tot een goed einde te brengen. Allereekst mijn promotoken professor Olle ten Cate en professor jan Borleffs. Olle, hartelijk dank dat je op het idee kwam dat dit onderzoek misschien wel wat voor my zou zijn 1k weet nog dat mijn eerste reactie iets was als: "hoe kon je daar nou by, ik ben helemaal geen onderzoeker". Later ben ik gelukkig van mening ve-Randerd. Ik vond onze samenwerking erg prettig. Vooral het feit dat ik in de periode van de organisatie van UHTRUST altigd wel even tussendoor big je binnen kon lopen heb ik erg gewaardeerd. Jan. dank voor het in mij gestelde vertrouwen Door je vertrek naar Groningen, heb je letterlijk en figuurlisk een wat meer afstandeliske rol gekregen. Fisn dat je in de laatste fase wel weer meer betrokken was.

Jany Rademakers was min eerste copromotor. Het feit dat jij mij zou begeleiden, heeft er mede voor gezorgd dat ik het aandurfde om ermee te beginnen. Jammer dat je door je nieuwe baan deze taak niet lang kon uitvoeren, maar ik ben bly met de start die we samen hebben gemaakt. Ma-Rieke van der Schaaf was een goede opvolger. Marieke, ik ben heel bly dat je positief reageerde op myn vraag om min copromotor te worden jouw enorme kennis en ervaring op het gebied van onderwüskundig onderzoek, hebben er zeker voor gezorgd dat het onderzoek kwalitatief beter is geworden Na onze afspraken, waar de thee altyd klaar stond, was ik altijd week erg gemotiveerd om door te gaan. Ook heel leuk dat je bereid was om te helpen met praktische zaken, zoals boodschappen doen voor de UATRUSTdag. Ik hoop dat onze samenwerking op de een of andere manier kan worden voortgezet.

Sue Kilminster, Trudie Roberts, Naomi Quinton, Shelley Fielden and Miriam Zukas: Thank you for the opportunity to stay for a period at the University of Leeds and work together. Despite the problems with the ethical approval, ive had a great time. Naomi and Shelley, thanks to your help with finding a better apartment and your lists with suggestions for restaurants and shops, I really felt at home in Leeds Bill Burdick, it was nice working with you Our annual meetings at the AMEE-conference were always very efficient. Our common project took much longer than I expected, but Im proud of the result.

By het UHTRUST-onderzoek had ik helemaal veel mensen nodig. Anne Oosterbaan is het langst by dit project betrokken geweest, vanaf de voorbereidingen tot en met de analyses. Zonder jou was het echt niet mogelijk geweest om de organisatie rond te krigen. Je inzet was fantastisch, je dacht altijd kritisch mee en ik kan me meerdere vrydagavonden herinneren waarop wy samen tot laat bezig zijn geweest met planningen, instructies en talloze lijst jes. Je was ook een enorme steun tijdens de behoorlijk stressvolle pilot in Hamburg. Toch hebben we daar ook gelachen en ik bewaar goede herinneringen aan de ontspannende wandeling Rondom het meer de dag erna Chantalle Burgers heeft ook een belangrijke bijdrage geleverd. Je had volgens mij geen idee waar je aan begon toen ik je vroeg, maar ik ben heel blij dat je de sprong in het diepe hebt gewaagd Ook is hebt wat avonden en weekenden moeten opofferen om de deadlines te halen, maar het is allemaal gelukt. De casussen (zelfs de extra casus die op het laatste moment moest worden gemaakt vanwege de EHEC-epidemie in Hamburg) waren helemaal in orde voor het hoofdonderzoek en op de dag zelf was je een geweldig "hoofd lab en kadiologie".

Eveline Boois, Kirstin Nillesen, Margreet Manrique en Stephan Contant: als onderwisskunde stagiaires hebben jullie een belangriske bijdrage geleverd aan de pilotfase. Leuk dat jullie ook bij het hoofdonderzoek weer mee hebben geholpen.

- lk heb ook dankbaar gebruik gemaakt van de expertise van een aantal junior docenten. Vooral Evelien de Moel
- heeft veel gedaan, van het zoeken van de juiste CT-
- scans tot aan het plakken van (heel veel) stickers op
- telefoons en USB-sticks Hartelijk dank daarvoor! Daarna
- hebben ook jorike Huiskes en Lonneke Alofs meegeholpen
- met aanvullende onderzoekjes Yolanda Hoogenberk en
- Linda Molman dank voor jullie hulp bij de catering en het
- bestellen van allerlei artikelen. Verder waren natuurlijk de
- artsen die mee hebben geholpen met de casussen, andere

assessoren, verpleegkundigen, kandidaten en medewerkers onnisbaar. Ik wil met name Annet van Royen hier noemen, die vanaf het allereerste begin betrokken is geweest bij het UATRUST-project

Sigrid Harendza, thanks for your collaboration in the UHTRUST-project Despite some difficulties we had fun together. I really appreciate your honesty and openness. I also want to thank you for the German translation of the summary

Medewerkers van het team kwaliteitszorg, Andrea Fast, Petra Simon, Lenneke Ainsworth, Margreet Manrique en Madelon Lohuis: hartelijk dank voor jullie begrip dat ik soms wat (te) weinig tijd voor jullie had. En erg leuk dat een aantal van jullie mee hebben geholpen tijdens de UHTRUST-dagen Dit laatste geldt ook voor de studentassistenten Chris de Witte en Sannemarije van Waes Chris, daarnaast ook bedankt voor je enorme hulp met Refworks.

Sjoukje van den Broek, dank voor je bijdrage aan het ontwikkelen van de casussen en vooral voor je hulp bij het hoofdonderzoek in Hamburg Ik vind het erg leuk dat we blijven samenwerken op onderzoeksgebied.

Feikje van Stiphout, Inge Pool, Anouk van der Gijp en Cécile Ravesloot: Ik vond onze lunches en andere afspraken altijd gezellig. Het was prettig om onderzoekservaringen te kunnen uitwisselen. Ik blijf graag op de hoogte van de voortgang van jullie onderzoeken.

Hanneke Mulder, dank voor je begrip als ik in sommige periodes meer tijd voor het onderzoek nodig had en voor het bewaken van mijn jaarwrenkaart; dat laatste met wisselend succes, maar dat lag niet aan jou

Willie Hols, je bent de beste kamergenoot die ik me kan wensen. Dank voor de gezelligheid en het fungeren als klankbord in de afgelopen jaren.

Ook naast het werk zijn er veel mensen belangrijk geweest in de afgelopen jaren Eline, je bent al heel lang een erg goede vriendin en jij was gedurende lange tijd de enige in män vielendenkring die gepromoveered was. Jä was daarmee een soort voorbeeld en ik vond het prettig dat je altäd goed een voorstelling kon maken van waar ik mee bezig was Erg bedankt voor je steun op belangräke momenten. Gelukkig hebben we deze keer geen "bankjesmoment" meegemaakt.

Liem Che heb ik leren kennen bis de Onderwissinspectie. We zijn daar ongeveer gelijk begonnen en het duurde niet lang voor we bis het kopieerapparaat aan elkaar bekenden dat dat niet bepaald onze droombaan was We zijn dan ook niet lang collega's geweest, maar we zijn wel vriendinnen geworden. Je hebt heel lang in Limburg gewerkt en gewoond, maar Cyriel heeft ervoor gezorgd dat je naar de randstad bent verhuisd het is helemaal leuk dat je nu in Gouda werkt. Onze lunches tijdens mijn schrijfdagen gaven me altijd weer nieuwe energie voor de rest van dag. Ik vind het leuk dat je mee hebt geholpen tijdens de UATRUST-dag en ik ben blij dat je mijn paranimf wilt zijn

Ook andere vrienden wil ik bedanken voor de gezellige en ontspannende momenten in de afgelopen jaren Michèle en Annemiek, leuk dat julie mee hebben geholpen tüdens de grote UHTRUST-dag Michèle, Als ik nog eens een "chef catering" nodig heb, dan weet ik je te vinden Sigrid, dank voor je hulp bü de ondertiteling van het Duitse filmpje en de samenvatting. José, bedankt voor je professionele kledingadvies.

Ouders en schoonouders; hartelijk dank voor de steun en het voortdurend tonen van belangstelling voor mijn onderzoek

Min allerliefste Maarten, de belangrijkste persoon in min leven het waren pittige jaren, met jouw project en min proefschrift. We hebben het allebei volbracht en ik heb veel zin in de volgende, hopelijk iets relaxtere, periode samen Fantastisch dat jij de vormgeving van dit boekje hebt gedaan, waardoor het een mooie co-productie is geworden Ik ben heel gelukkig dat jij aan mij zijde wilt staan als paranimf, maar vooral in de rest van mijn leven.

Marjo Winen-Meijer

Gouda, november 2012

Curriculum vitæ

Marjo Wijnen-Meijer was born on November 10th, 1971 in Leersum, the Netherlands. In 1990 she completed her pre-university education at the Lingecollege in Tiel. She subsequently started studying Educational Sciences at Utrecht University, specializing in 'curriculum development' and 'educational consultancy'. She obtained her MSc degree in 1996.

After graduation, she worked at educationalist at Comenius College in Capelle aan den IJssel (school for secondary education) and at Rotterdam University of Applied Sciences. Since 2005 she works at the Center for Research and Development of Education at UMC Utrecht. Among other things, she was involved in the modernizing of postgraduate medical curricula and she has been project manager of a project for continuing education of nurses. Since 2008 she is head of the Quality Assurance team. In November 2006 she started her PhD project.

Marjo Wijnen-Meijer is geboren op 10 november 1971 in Leersum. Ze behaalde in 1990 haar Gymnasium-diploma aan het Lingecollege in Tiel. In 1991 behaalde ze het propedeutisch examen Pyschologie en Pedagogiek aan de Universiteit Utrecht en in datzelfde jaar startte ze met de doctoraalstudie onderwijskunde. Zij specialiseerde zich op het gebied van curriculumontwikkeling en onderwijskundig advies.

Na haar afstuderen in 1996 heeft ze diverse functies als onderwijskundige gehad, onder meer bij het Comenius College in Capelle aan den IJssel en de Hogeschool Rotterdam. Vanaf 2005 is ze werkzaam bij het Expertisecentrum voor Onderwijs & Opleiding van het UMC Utrecht. Ze is onder andere betrokken geweest bij de modernisering van de medische vervolgopleidingen en heeft een project geleid op het gebied van nascholing van verpleegkundigen. Sinds 2008 is ze hoofd van het team Kwaliteitszorg. In november 2006 is ze gestart met haar promotieonderzoek.



Index (chapters)



Assessment

Frame of reference training 7, 8, 9, 10, 11 Instruments 1,9, 10 Authentic performance assessment 9, 10, 11 Authentic simulation procedure 9, 10, 11 Direct Observation of Procedural Skills—DOPS 9 Entrustment decisions 1, 3, 7, 8, 9 Mini-clinical evaluation—Mini-CEX 1, 9 Objective Structured Clinical Examination—OSCE 9 Utrecht Hamburg Trainee Responsibility for Unfamiliar Situations Test-UHTRUST 9, 10, 11, A, B Levels of assessment 9 Miller's pyramid 9 Reliability 9 Scoring forms 9, 10 CARE 9, 10, B Entrustable Professional Activities 9, 10, B Facets of competence 9, 10, BPost Patient Encouter Form—PPEF(9, 10, B) Self-assessment 3 Utility aspects 9 Acceptability 9 Costs 9 Feasibility 9 Validity 9, 10, 11 Argument-based approach 9 Threats to validity 9

Expertise development

Adaptive experts 1, 11 Coping with unfamiliar situations 9, 10, 11 Clinical reasoning 1 Competency 1, 11 Competency frameworks 1, 7 Facets of competence 7, 8, 9, 10, 11 Deliberate practice 1, 11 Entrustable Professional Activities—EPAS 1, 3, 9, 10, 11 Expertise 1, 11 Skill acquisition 1, 11 Stages of expertise development 1, 11 Zone of Proximal development 1, 3, 11

Medical school curricula

Case-based learning 1 Competency-based education 1, 7, 8, 11 Discipline based 1 H-shape curriculum 1, 4, 5, 11 Horizontal integration 1 Longitudinal integrated clerkships 1, 11 Models of curriculum development 1 PRISMS framework 1 RIFLE framework 1 SPICES model 1, 6 Problem-based learning 1 Vertical integration 1, 4, 5, 6, 10, 11 Definition 1, 6, 11 Early clinical clerkships 1, 4, 5, 11 Extended clerkships 1, 4, 5, 11 Increasing responsibility 1, 3, 4, 5, 6, 11 Z-shape curriculum 1, 4, 5, 11

Methodology

Delphi study 7 Design based research 1, 7, 8, 9, 10 Document analyses 2 Literature review 7 Questionnaire study 1, 2, 3, 4, 5, 6 Qualitative questionnaire study 2 Ranking study 8

School effectiveness

Factors 1 Models 1

Structure of medical education

Degrees after medical school 2 Stages of medical education 2 Internship 2 Medical school 2 Postgraduate education 2 Residency 2 Social service 2 Undergraduate education 2 Unrestricted practice 2, 3, 11

Support of expertise development

Community of Practice—CoP 1, 11 Scaffolding 1 Supervison 3, 11 Degrees of supervision 3 Workplace learning 1, 11 Internalization 11 Socialization 1, 11

Transfer

High-road transfer 1, 11 Low-road transfer 1, 11

Transitions

Medical school to postgraduate training 1, 4, 5, 6, 10, 11 Admittance to residency 4, 5, 11 Career choice 4, 5, 11 Preparedness for work and postgraduate training 5, 6, 10, 11 Preclinical to clinical education 1, 4, 5, 11 Types of transition 1, 3, 11 Licensure transition 3, 11 Responsibility transition 3, 11