DEVELOPING CLINICAL COMPETENCE

PAUL F. WIMMERS

Developing Clinical Competence

Paul F. Wimmers

Cover design by Paul Wimmers
Video frames of supervised training by Tatum Korin

ISBN-10: 90-9021068-7
ISBN-13: 978-90-9021068-1

With summary in Dutch
Keywords: medical expertise / diagnostic process / cognitive psychology

Printed in the Netherlands by Printpartners Ipskamp B.V.

Developing Clinical Competence

Ontwikkeling van Klinische Competentie

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus,
Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 19 oktober 2006 om 11:00 uur

door

Paul Ferdinand Wimmers
geboren te Amsterdam

Promotiecommissie

Promotor:
Prof.dr. H.G. Schmidt

Overige leden:
Dr. R.M.J.P. Rikers
Prof.dr. L. Wilkerson
Prof.dr. D.J.G.M. Duncker

Table of Contents

CHAPTER ONE

Defining Clinical Competence: An Introduction

"You cannot care for a patient unless you have an idea of what is wrong; an [accurate] diagnostic decision has to be made before you treat." (Barrows & Pickell, 1991)

"A good doctor is one who is shrewd in diagnosis and wise in treatment; but, more than that, he is a person who never spares himself in the interest of his patients." (Sir Hugh Cairns, 1949)

A PROPER DEFINITION of clinical competence and its components is important to serve as a criterion for validating medical educational programs and to assure a minimum level of competency at the end of medical school and beyond during residency. Understanding clinical competence is crucial not only for medical education, assessment, and licensing examinations, but also for society and its responsibility for the quality of health care. Many clinical educators, committees, specialty boards, and expert panels within and outside the medical profession have noticed the importance of defining what is required of a health care professional and have spent time and effort in attempting to describe clinical competence. Webster dictionary defines being competent as the quality of having sufficient knowledge, judgment, skill, or experience for some purpose. It could be reasoned that if one of the primary tasks of a doctor is diagnosing illnesses and providing interventions to improve the condition of patients, then a clinically competent doctor would be someone who has the knowledge, judgment, skill, and experience to diagnose correctly and, in addition, who is capable of providing appropriate treatment interventions (Burg, Lloyd, & Templeton, 1982). A simple definition like this can, however, not function as a thorough framework for clinical competence. A general definition of clinical competence turns out to be not as easy as a dictionary suggests, as the medical profession is already concerned with this issue for many decades (Neufeld & Norman, 1985; Stern, 2006).

This chapter will introduce a working definition of clinical competence. This working definition is not intended to replace any other definitions, but to provide a framework for the subsequent chapters. This thesis consists of a collection of six studies that each focus on a different aspect of clinical competence. In the studies discussed, clinical competence is mostly related to performance or the diagnostic problem-solving process. The studies appear here in three parts: (1) Determinants of clinical competence development, (2) The nature of clinical competence, and (3) The development of clinical competence.

Historical perspectives on clinical competence

White (1959) suggested that competence be conceptualized as effective interaction with the environment. This means that clinical competence manifests itself only in observed behaviors or practice. A description of competence according to the tasks of the clinical encounter, such as history taking, physical examination, use of laboratory tests, patient management, record keeping, etc. was helpful to get insights on the purpose and consequence of physicians' behaviors. Hubbard and colleagues, for example, described a method developed by John Flanagan and the staff of the American Institute for Research in the early 1960s to realistically define the characteristics of clinical competence (Hubbard, Levit, Schumacher, & Schnabel, 1965). Senior physicians and residents who had direct responsibility for the supervision of students were asked to record good and bad medical practices in clinical situations. It is interesting to mention that the most frequently reported (good) practices in this so called "critical-incident technique" (see for the original description of the critical-incident technique, Flanagan, 1954) were all based on the diagnostic process, such as: taking a history thoroughly and performing a physical examination in an orderly manner; accurately recognizing the patient's condition from observation of clinical signs; including further information in diagnosis; correctly suspecting obscure diagnosis despite the obvious symptoms and signs of another diagnosis; and taking appropriate emergency action when indicated (Hubbard et al., 1965).

Definitions of clinical competence were initially focused on diagnostic problem solving (a major responsibility of a doctor). In time, definitions became more detailed and in line with the fast growing demands and expectations of society on health care delivery. A good doctor was more than a diagnostic problem solver. For example, initially the American Board of Internal Medicine distinguished between four different dimensions of clinical competence, whereby problem solving was the core aspect: (1) abilities (i.e., knowledge, technical skills, and interpersonal skills), (2) problem solving skills (i.e., data-gathering and diagnoses), (3) the nature of the medical illness (the problems encountered by the physician), and (4) the social and psychological aspects of the patient problem, especially those which relate to diagnosis and management (ABIM, 1979). In a later report of the American Board of Internal Medicine, more elements were added: communication skills, professionalism (e.g., ethical practice, understanding diversity, responsible attitude), and system-based practice (i.e., understanding of the health care system to improve and optimize health care), (ABIM, 2002). The Institute of Medicine of the National Academies formulated five core competencies that were also much more than just problem solving: providing patient-centered care, employing evidence-based medicine, applying quality improvement, utilizing informatics, and working in interdisciplinary teams (Greiner & Knebel, 2003). The Canadian Medical Education Directions for Specialists (CanMeds) formulated seven key competencies. (Their mission was to describe essential roles of physicians in the context of global trends towards greater demands for public accountability, rising patient consumerism, rapidly evolving medical science and technology, and fiscal restraints on health care spending.) Only the first competency is directly related to diagnostic problem solving. In addition, the physician must be able to be a clinical decision-maker, communicator, collaborator, manager, health advocate, scholar, and a professional (CanMEDS, 1996; Frank & Langer, 2003). These later descriptions emphasize that a clinically competent physician is indeed a highly qualified and specialized professional able to function in a society that puts high demands on the professional.

Determinants of clinical competence development

A clear and concise definition of clinical competence seems most pressing for assessment purposes. George E. Miller (1990) distinguished several hierarchical layers of competence to function as a framework for within which assessment might occur (see Figure 1). The different layers in Miller's model represent a developmental sequence of stages, in other words, a horizontally layered hierarchical categorization of clinical competence. All levels are needed and have their own important impact on clinical competence. In his framework for clinical competence a distinction is made between knows (knowledge), knows how (competence), shows how (performance) and does (action). Knowledge is at the base of this triangle shaped framework and action is at the top. A student, resident, or physician needs the knowledge that is required to carry out professional functions effectively, a prerequisite for being clinical competent. The next two layers, competence and performance, which follow upon knowledge, are often used interchangeably; however, competence means that a physician can apply his/her knowledge in concrete situations, while performance is the ability to use this knowledge to perform concrete actions. The final top layer represents what a physician actually does during day-to-day practice.

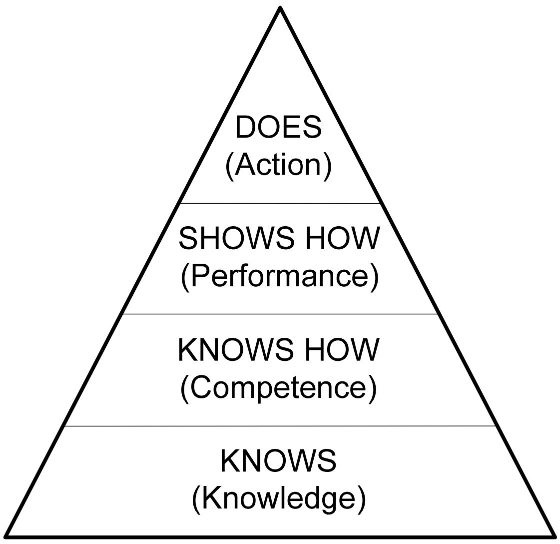


Figure 1. *Framework for clinical assessment by Miller (1990)*

A distinction between competence and performance is often made in the literature. Senior (1976), for example, defined competence as what a physician is capable of doing and performance as what a physician actually does. The former would, in this case, be related to the first three layers of the pyramid of Miller (1990) and the latter to the top layer (See also, Rethans, Van Leeuwen, Drop, Van der Vleuten, & Sturmans, 1990; or, Van der Vleuten et al., 2000). The implication is that (student) performance must be measured in order to assess competence, and many different tests are probably needed. Assessment of medical students has focused mostly on "knows" and "knows how," the base of the pyramid: recall of factual knowledge and the application of this knowledge in problem solving. However, such examinations may fail to document what students will do when faced with a real patient. To determine someone's clinical competence, observing behaviors in action is needed, and this is represented by the top layer of the pyramid in Miller's model (Miller, 1990; Wass, Van der Vleuten, Shartzer, & Jones, 2001).

The symbiosis between assessing and defining becomes clear in this distinction, simply because a proper definition of clinical competence is needed for clinical assessment. Despite the fact that Miller's pyramid is primarily intended to serve as a framework to define and categorize different assessment tools, his model gives a good idea about which characteristics are influencing the development of clinical competence. The layers represent how students build their knowledge during the preclinical years and how competence and performance are shaped in action during the latter years of clerkships and residency, wherein the clinical encounter is crucial.

Chapter two ("The Predictability of Performance in Medical School: A Comparison of Grade Subgroups") focuses on the predictive value of level of pre-university achievement on performance during medical school for an entire cohort versus subgroups. Specifically, it explores its influence on the two distinctive phases of medical education: the preclinical and the clinical phase. Indicators for level of pre-university and preclinical achievement consisted of written knowledge-based assessments (knows and knows how). In the latter clinical years direct observations of professional performance and oral end-of-clerkship examinations are used to assess students' level of competence. Chapter three ("Influence of Clerkship Experiences on Clinical Competence") focuses on the relationship between the nature and volume of patient encounters and the learning outcomes during clerkships. After determining the variation of students' clinical experiences within and across sites, this study attempted to explain the causes of this inter-site variation in clinical experiences and, in addition, to investigate the consequence of this variation in clinical experiences and quality of supervision on clinical competence. Clinical competence is indicated by direct observation of professional performance, a practical end-of-clerkship examination, and a theoretical end-of-clerkship examination.

The nature of clinical competence

Before the 1970s, research on clinical competence was for the most part focused on general observable abilities. For example, Elstein, Shulman and Sprafka (1978) initially discovered that the medical problem-solving process was characterized by generating multiple hypotheses early in the patient encounter. The number of hypotheses generated and the amount of information collected by doctors were essentially the same for all specialties, which suggested the existence of a general competency that can be applied to different clinical contents (see also, Barrows, Norman, Neufeld, & Feightner, 1982; Barrows & Pickell, 1991; Gale, 1982). However, based on further work from Elstein, it turned out that someone who was able to diagnose one patient's problem was not necessarily able to diagnose a different patient's problem. Therefore, it was concluded that clinical competence is highly dependent on the particular content of the situation. This phenomenon, named "content specificity (Elstein et al., 1978)," emerged in many studies using different kind of assessment methods, such as patient management problems (Neufeld & Norman, 1985), written tests (De Graaf, Post, & Drop, 1987), oral tests (Swanson, 1987), performance-based tests (Van der Vleuten & Swanson, 1990), standardized-patient tests, (De Champlain, Macmillan, King, Klass, & Margolis, 1999), and computer-based clinical performance assessments (Fitzgerald et al., 1995). The ability to solve clinical problems did not seem to be a general, content-independent characteristic of doctors. The reasoning that clinical competence is more than solving clinical problems was replaced by a perspective that knowledge is an essential factor in all competencies (see for example, Van der Vleuten et al., 2000).

Chapter four ("Clinical competence: General Ability or Case-specific?") focuses on whether clinical competence can be considered a general, content-independent ability or whether competence in this area is dependent on case-specific knowledge (Barrows et al., 1982; Elstein, 1972; Elstein et al., 1978; McGuire, 1976). For this purpose, individual oral end-of-clerkship examination scores of students on 10 different clerkships are analyzed in two separate modeling steps using structural equation modeling (SEM) techniques. Chapter five ("Clinical Competence through the Eyes of an Educator: Differences in Perceived Importance of Student Performance on the Wards and on Clerkship Examinations") uses a survey that lists 21 individual competencies. This survey was administrated among clinical educators and physicians of different hospitals and disciplines involved with student learning on clerkship rotations. Our mean point of interest was to determine whether there is a difference between what is important for normal daily performance on the wards versus performance on clerkship examinations.

The development of clinical competence

Studies into the nature of this knowledge, its application and development started to dominate the field of medical education research in the 1980s and 1990s (Boshuizen & Schmidt, 1992; Chi, Glaser, & Rees, 1982; Norman, Tugwell, Feighter, Muzzin, & Jacoby, 1985; Patel, Evans, & Groen, 1989; Patel & Groen, 1986a, 1986b; Schmidt & Boshuizen, 1992, 1993a). Much of this research was focused on the transition from theory to practice and the consequences for knowledge development and its structure. A well-known and useful example of a theory that used the development of medical knowledge as the cornerstone of clinical expertise is the knowledge encapsulation theory proposed by Schmidt and Boshuizen (1992). They reasoned that knowledge acquired during the first years of medical school integrates with clinically relevant knowledge in the latter practical years of medical school. Basic biomedical science knowledge of the first years becomes encapsulated into clinical concepts by repeated exposure to real clinical problems. Theories based on the integration of biomedical and clinical knowledge were not only successful in explaining such well documented phenomena as the above mentioned content specificity of problem solving (Patel et al., 1989; Schmidt, Norman, & Boshuizen, 1990), but also in explaining the often found "intermediate effect" in clinical case recall. This robust phenomenon often documented in the medical expertise literature consists of the finding that medical students of intermediate levels of expertise outperform both experts and novices in clinical case recall after diagnosing cases (e.g., Boshuizen & Schmidt, 1992; Patel et al., 1989; Schmidt & Boshuizen, 1992, 1993b; Schmidt et al., 1990).

Chapter six ("Inducing Expertise Effects in Clinical Case Recall through the Manipulation of Processing") is directed at the distinction between the "intermediate effect" and expertise effect in clinical case recall. Despite the consistent finding of intermediate effects in clinical case recall, in some instances, expertise effects are found (Norman, Brooks, & Allen, 1989). The current study manipulates case processing to explore under what conditions a shift occurs from an intermediate effect in recall towards an expertise effect. Chapter seven ("Effects of Level of Expertise on Data-gathering Behavior during different Stages of the Diagnostic Process") relates to previous research in medical problem solving that has been plagued by inconsistent findings about data-gathering behavior. There is uncertainty about the amount of information needed to mentally represent and solve a diagnostic problem and whether this is related to level of expertise. Some studies have proven that experts are more efficient in data-gathering (Rimoldi, 1955, 1961), while others relate expertise to spending more time and selecting more information (e.g., Chi, Feltovich, & Glaser, 1981; Van Gog, Paas, & Van Merriënboer, 2005). This study attempts to explain these findings and consideres how the amount of patient-data gathered differs during subsequent stages of the diagnostic problem-solving process and how it is related to the level of expertise.

Proposed categorization of clinical competence for this thesis

A categorization of clinical competence will be proposed here to function as a framework. As said, this framework is not intended to be a replacement or improvement of any other existing model, but is intended to serve as a model for the understanding of the subsequent chapters. The format of the categorization is derived from an often used division of competence into knowledge, skills, and attitudes (e.g., Rice & Sinclair, 1995) and further specified according to a categorization of clinical competence given by Norman (1985).

"Diagnostic Problem Solving and Clinical Judgment," a component of clinical competence (see Table 1), is subdivided into "the stages of the diagnostic process." These stages are derived from the study of Hubbard et al., (1965) and modified according to studies of Brug and Lloyd (1983) and Epstein and Hundert (2002).

Table 1. *Categorization of clinical competence and its application*

Cognitive Abilities	
	Biomedical and clinical knowledge and the ability to apply it to concrete situations
Diagnostic Problem Solving and Clinical Judgment	
a.	Obtaining sufficient information from clinical history and patient notes
b.	Performing a focused physical examination
c.	Utilizing and applying laboratory tests methods correctly
d.	Utilizing and applying medical procedures correctly
e.	Arriving via a reasonable differential diagnoses at a final diagnosis
Interpersonal Skills	
	Effective communication with patients and colleagues
Professional Qualities	
	Respectful and professional relationships with patients and in the provision of health care

References

ABIM (1979). American Board of Internal Medicine: clinical competence in internal medicine. *Annals of Internal Medicine, 90*, 402-411.

ABIM (2002). *Residents: Evaluating your clinical competence. New competencies for internal medicine*. Philadelphia, Pennsylvania: American Board of Internal Medicine Clinical Competence Program.

Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning process of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine, 5,* 49-56.

Barrows, H. S., & Pickell, G. C. (1991). *Developing clinical problem-solving skills: a guide to more effective diagnosis and treatment* (First ed.). New York, London: Norton Medical Books: W. W. Norton & Company, Inc.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16,* 153-184.

Burg, F. D., & Lloyd, J. S. (1983). Definitions of competence: A conceptual framework. In J. S. Lloyd & D. G. Langsley (Eds.), *Evaluating the skills of medical specialists* (pp. 1-71). Chicago: American Board of medical Specialties.

Burg, F. D., Lloyd, J. S., & Templeton, B. (1982). Competence in Medicine. *Medical Teacher, 4*(2), 60-64.

CanMEDS (1996). *Skills for the new millennium: report of the societal needs working group.* Ottawa, Canada: CanMEDS 2000 Project Societal Needs Working Group Report.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-76). Hillsdale, NJ: Erlbaum.

De Champlain, A. F., Macmillan, M. K., King, A. N., Klass, D. J., & Margolis, M. J. (1999). Assessing the impacts of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administrated standardized patient examination. *Academic Medicine, 74*(10), S52-S54.

De Graaf, E., Post, G. J., & Drop, M. J. (1987). Validation of a new measure of clinical problem-solving. *Medical Education(21),* 213-218.

Elstein, A. S. (1972). *Methods and theory in the study of medical inquiry.* Cambridge, MA: Harvard University Press.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning.* London, Cambridge, Mass: Harvard University Press.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA, 287*(2), 226-235.

Fitzgerald, J. T., Wolf, F. M., Davis, W. K., Barclay, M. L., Bozynski, M. E., Chamberlain, K. R., et al. (1995). A preliminary study of the impact of case specificity on computer-based assessment of medical student clinical performance. *Evaluation and the Health Professions, 17*(3), 307-321.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Frank, J. R., & Langer, B. (2003). Collaboration, communication, management, and advocacy: Teaching surgeons new skills through the CanMEDS Project. *World Journal of Surgery, 27*(8), 972-978.

Gale, J. (1982). Some cognitive components of the diagnostic thinking process. *British Journal of Educational Psychology, 52*, 64-76.

Greiner, A. C., & Knebel, E. (Eds.). (2003). *Health profession education: A bridge to quality.* Washington. D. C.: The National Academies Press.

Hubbard, J. P., Levit, E. J., Schumacher, C. F., & Schnabel, T. G. (1965). An objective evaluation of clinical competence: new technics used by the National Board of Medical examiners. *New England Journal of Medicine, 272*, 1321-1328.

McGuire, C. H. (1976). Simulation technique in the teaching and testing of problem solving skills. *Journal of Research in Science Teaching, 13*(2), 89-100.

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

Neufeld, V. R., & Norman, G. R. (1985). *Assessing Clinical Competence*. New York: Springer-Verlag.

Norman, G. R. (1985). Defining competence: a methodological review. In V. R. Neufeld & G. R. Norman (Eds.), *Assessing Clinical Competence.* New York, NY: Springer-Verlag.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by experts medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology; Learning, Memory and Cognition, 15,* 1166-1174.

Norman, G. R., Tugwell, J. P., Feighter, J. W., Muzzin, L. J., & Jacoby, L. L. (1985). Knowledge and clinical problem-solving. *Medical Education, 19,* 344-356.

Patel, V. L., Evans, D. A., & Groen, G. J. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 53-112). Cambridge: The MIT Press.

Patel, V. L., & Groen, G. J. (1986a). Knowledge-based solution strategies and the use of biomedical knowledge by medical students. *Medical Education, 24,* 129-136.

Patel, V. L., & Groen, G. J. (1986b). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*(1), 91-116.

Rethans, J. J., Van Leeuwen, Y., Drop, M., Van der Vleuten, C. P. M., & Sturmans, F. (1990). Competence and performance: two different constructs in the assessment of quality of medical care. *Family Practice, 7,* 168-174.

Rice, C. A., & Sinclair, M. (1995). Competency-based objectives for clinical training. *Canadian journal of medical technology, 57*(3), 136-140.

Rimoldi, H. J. A. (1955). A technique for the study of problem solving. *Educational and Psychological Measurement, 15*, 450-461.

Rimoldi, H. J. A. (1961). The test of diagnostic skills. *Journal of Medical Education., 36*(1), 73-79.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York, NY: Springer-Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On Acquiring Expertise in Medicine. *Educational Psychological Review, 5*(3), 205-221.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*(3), 338-351.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611-621.

Senior, J. R. (1976). *Toward the measurement of competence in medicine.* Philadelphia: National Board of Medical Examiners.

Stern, D. T. (Ed.). (2006). *Measuring Medical Professionalism.* New York: Oxford University Press.

Swanson, D. B. (1987). A measurement framework for performance-based tests. In I. R. Hart & R. M. Harden (Eds.), *Further developments in Assessing clinical competence* (pp. 13-45). Montreal: Can-Heal Publications.

Van der Vleuten, C. P. M., Scherpbier, A. J. J. A., Dolmans, D. H. J. M., Schuwirth, L. W. T., Verwijnen, G. M., & Wolfhagen, H. A. P. (2000). Clerkship assessment assessed. *Medical Teacher, 22*(6), 592-600.

Van der Vleuten, C. P. M., & Swanson, D. B. (1990). Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine, 2*, 58-76.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Wass, V., Van der Vleuten, C., Shartzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet, 357*, 945-949.

White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297-333.

CHAPTER TWO

The Predictability of Performance in Medical School: A Comparison of Grade Subgroups

This study is conducted at and funded by Erasmus MC—University Medical Center Rotterdam (EUR).
An earlier version of this paper was presented at the 86th Annual Meeting of the American Educational Research Association (AERA) in Montreal, Canada, April 11-15, 2005, with the title: Admission Methods and their Predictive Value for Success in Medical School across Institutions.

Reference
Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (in preparation, 2006). The Predictability of Performance in Medical School: A Comparison of Grade Subgroups.

Abstract

BACKGROUND Indicators based on cognitive performance are often considered the best predictors for academic performance in medical school. Most studies that advocate this viewpoint analyzed a whole cohort and were only successful for prediction of preclinical performance. Far less is known about the relationship between level of preadmission GPA and its effect on curriculum transitions from preclinical to clinical.

PURPOSE The study investigated the predictive value of preadmission GPA for preclinical and clinical performance and the applicability of predictive values for an entire cohort and for subgroups of a cohort based on the level of preadmission GPA (low, intermediate, and high achievers). In addition, the study investigated whether students' consistency of performance is related to the subgroup they belong to, independent of phase transitions of the curriculum.

METHODS The predictability and consistency of performance across transitions in the curriculum for all subgroups were explored in order to get insight into the interaction between the performance of the subgroups and the phases of the curriculum.

RESULTS Performance of the low-achiever group is inconsistent and not predictable; performance of the intermediate group is predictable for the preclinical phase only, while the performance of the high-achiever subgroup is best predicted by preadmission GPA and is less influenced by the effects of curriculum transitions.

CONCLUSIONS Level of preadmission GPA can serve not only as a predictor of performance, but also as an indicator for *consistency* of performance during medical school. High achieving students seem less affected by curriculum transitions and keep performing at a consistent and highly predictable level.

PERFORMANCE OF STUDENTS at the university in general and medical school in particular involves a very complex interaction between student-related factors, such as intelligence and motivation (Perrot, Deloney, Hastings, Savell, & Savidge, 2001), and external factors, such as teaching and social environments (Woloschuk, Harasym, & Temple, 2004).

Both the student-related factors and the external factors are very diverse and subject to change over time (Rothman, 1973). One of the most consistent student-related factors at various levels of education is cognitive performance. Indicators based on past cognitive performance, such as grade point average (GPA) and MCAT scores, are considered the best and most widely used predictor and criterion for future academic performance in medical school. Correlations found between preadmission GPA and preclinical performance are moderate to high (range from 0.25 to 0.60), (Ferguson, James, & Madeley, 2002; Gottheil & Miller-Micheal, 1957; Julian, 2005; Mitchell, 1990; Peat, Woodburry, & Donner, 1982; Salvatori, 2001; Wiley & Koenig, 1996). Apparently, cognitive measures of previous performance are the most consistent predictors for future academic performance because such measures are representing a similar underlying ability (Ferguson et al., 2002; Rheault, 1988).

However, medical education consists of two distinctive phases: a preclinical and a clinical phase. The first phase emphasizes primarily the acquisition of basic science knowledge, while in the latter phase, the emphasis shifts towards practice and the application of knowledge in a clinical setting. This major transition from theory to practice is often experienced as difficult and stressful for students (Prince, van de Wiel, Scherpbier, van der Vleuten, & Boshuizen, 2000). It is well studied that practical experiences during the clerkships have a restructuring impact on the knowledge base of the students (Schmidt & Boshuizen, 1992, 1993a). Boshuizen (2005) describes this transition as a discontinuity in the development of expertise. The predictability of clinical performance during clerkships or the later practice of medicine from preadmission GPA becomes indeed more complicated (Gottheil & Miller-Micheal, 1957; Korman, Stubblefield, & Martin, 1968; Murden, Galloway, Reid, & Colwill, 1978; Rheault, 1988; Ronai, Golman, & Shanks, 1984; Salvatori, 2001). The study of Peat and colleagues (1982) for example, could explain only nine percent of the variance in clinical performance by preadmission performance. Such data underline the decrease due to a distinctive change of learning environment, such as the clinical phase. In addition, it seems logical in view of the wide diversity of student related factors that predictive factors for study performance may vary for subgroups of students. In fact, it has been shown that the predictive value of preadmission GPA is different for subgroups of students, such as students of racial/ethnic minority groups (Koenig, Sireci, & Wiley 1998; Lynch & Schneider, 2000) and students who used English as a second language (Chan-Ob & Boonyanaruthee, 1999).

In the present study, we investigated the predictive value of preadmission GPA for performance during medical school by students who performed at a low, intermediate, and high level before admission to medical school. The goal of this study was to investigate the predictive value of GPA's for performance in different phases of medical education (i.e., preclinical and clinical phase) and the predictive value of GPA's for an entire cohort versus subgroups. Furthermore, the study investigated whether students are progressing at a consistent level throughout medical school, independent of the phase transitions of the curriculum, and whether this consistency is related to the grade-subgroup they belong.

Methods

Background

Defining subgroups based on grades is a commonly used procedure by admission committees at Dutch universities. The admission procedure used to select students has been accomplished by a weighted lottery according to such subgroups since the 1970s. It is assumed that students who pass the Dutch National License Examination of public preparatory school (in Dutch: "Voortgezet Wetenschappelijk Onderwijs") are prepared to enroll in universities and eligible to apply for the lottery admission process. This unique system gives the opportunity to study a student population based on a more complete range of grades, since the lottery brings in students with relatively low high school GPA's as well as high ones. The eligible candidates are divided into categories based on their high school grades. The weight of the lottery draw is based on the grade-subgroup of the candidate.

Participants

Medical students (N = 329) who completed medical school successfully in the year 2000/2001 and 2003/2004 were considered in this study (146 male, 183 female). The medical program takes 6 years and is divided in two distinctive phases. The preclinical phase occupies the first 4 years and focuses primarily on theory and basic science knowledge, and the clinical phase that occupies the final 2 years is devoted to clinical practice.

Procedure

The students were classified into three groups based on their preadmission GPA (GPA of public preparatory school): low, intermediate, and high achievers. Low achievers were students with a mean preadmission GPA between 5.5-7 (n = 184), intermediate achievers were students with a mean preadmission GPA between 7-8 (n = 107) and high achievers (honor students) were students with a mean preadmission GPA above 8 (n = 38). Preadmission GPA's were used as predictor for success in medical school for the whole cohort and the different grade-subgroups, respectively. Preclinical and clinical performance were determined by students' mean grades of all course examinations within that phase measured on a 10-point scale, with a passing grade being an average of 5.5 or higher. Where a student repeated a course, only the result of the first attempt was included in this study. This was done to avoid strongly skewed to the right distributions (correlation between first-attempt and final-attempt grades was .91).

The preclinical phase consisted of 31 courses. "Preclinical performance" is the mean of the final examinations for these courses. The reliability of this criterion is .94 (Cronbach's alpha). In the clinical phase, students passed through 10 different clerkships in a fixed sequence, starting with internal medicine and ending with family medicine (internal medicine, pediatrics, neurology, psychiatry, surgery, gynecology, dermatology, otorhinolaryngology, opthalmology, and family medicine, respectively). All clerkships ended with an independent final examination taken orally. Most examinations consisted of the completion of a patient file consisting of patient history, physical examination, and diagnosis. It is suggested that oral examinations cover three components of competence: knowledge, problem-solving ability, and personal characteristics (Muzzin & Hart, 1985). The clinical performance variable was the mean of these 10 clerkship finals. The reliability coefficient of this criterion was .82 (Cronbach's alpha).

In addition, to investigate whether students perform at a consistent, repeatable level through medical school, and to be able to compare the subgroups with each other, internal consistency reliability (Cronbach's alpha) across all curriculum phases was calculated for each grade-subgroup.

In this study, we used the data of medical students who completed medical school successfully. However, on average, 15% of the admitted students dropped out, mainly during the first preclinical years. The dropouts were divided, 80%, 20%, and 1% over the low, intermediate, and high-achievers subgroups, respectively.

Analysis

SPSS 13.01 is used to calculate descriptive statistics, correlations, and explained variances of performance in all the phases. Skewness and kurtosis statistics are used to explore whether normality assumptions are met for all used variables. Levene's test for equality of variances is used to test if the predictor and criterion variables have equal variances for all subgroups. Subgroups were analyzed by using one-way ANOVAs and ANOVAs for repeated measures in order to differentiate between group effect, curriculum effect, and its interaction. Fisher's least significant difference (LSD) was used to make post hoc comparisons between the different subgroups. Furthermore, standardized scores were used to make performance during the different school phases comparable (i.e., public preparatory school, preclinical, and clinical education). Cronbach's alphas were used as indicator for the consistency of performance within grade-subgroups. An alpha level of .05 was used for all statistical analyses.

Results

Descriptive statistics of all variables for the whole cohort as per grade-subgroup are listed in Table 1.

Skewness and kurtosis statistics were within range of normality for all variables: preadmission GPA, preclinical GPA, and clinical GPA. Levene's test for equality of variances indicated that homogeneity of variances between grade-subgroups has been met for all variables. The Levene statistic is for preadmission GPA, F (2, 326) = .005, $p > .05$, preclinical phase, F (2, 326) = 1.668, $p > .05$, and clinical phase, F (2, 326) = .205, $p > .05$. The assumption of homoscedasticity was met, indicating that the variability in scores for one variable is roughly the same at all values of the other variable. The correlations between preadmission GPA and performance across medical school, and the correlations between the subsequent phases within medical school are listed in Table 2.

For the complete cohort (N = 329), the correlation between preadmission GPA and preclinical performance was high (r = .54, $p < .001$), the explained variance was 28.6%. The correlation between preclinical performance and clinical performance was .43, $p <$ .001, explained variance is 18.5%.

The correlation between preadmission GPA and clinical performance, on the other hand, is .25, $p < .001$ with an explained variance of 6.1%. In line with former findings, preadmission GPA obviously is a better predictor for preclinical performance than for clinical performance.

Table 1. *Descriptive statistics for all grade-subgroups*

			Grade-subgroups					
	Overall (N=329)		Low (n=184)		Intermediate (n=107)		High (n=38)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Preadmission GPA	6.94	.73	6.40	.30	7.35	.29	8.34	.33
Preclinical phase	6.61	.76	6.31	.61	6.84	.70	7.40	.75
Clinical phase	7.78	.39	7.70	.37	7.85	.36	7.97	.43

Table 2. *The predictability of performance for whole cohort and grade-subgroups*

		Grade-subgroups		
	Overall (N=329)	Low (n=184)	Intermediate (n=107)	High (n=38)
GPA → Preclinical phase (4 years)	0.54^{**}	0.10	0.36^{**}	0.43^{**}
GPA → Clinical phase	0.25^{**}	0.03	0.03	0.37^{*}
Preclinical phase → Clinical phase	0.43^{**}	0.32^{**}	0.38^{**}	0.52^{**}

* $p < .05$, ** $p < .01$, two-tailed

Focusing on the predictive value of preadmission GPA for performances in medical school for grade-subgroups of students, it appeared that for the low-achiever subgroup (n = 184) the predictive value of preadmission GPA for preclinical performance and clinical performance was non-significant. Thus, performance during medical school *could not* be predicted with preadmission GPA in the low-achiever subgroup. The predictive value of preadmission GPA for preclinical performance of the intermediate-achiever subgroup (n = 107) was .36 ($p < .001$), with explained variance of 12.8%, and lost its predictability during the clinical years. The predictive value of preadmission GPA for preclinical performance of the high-achiever subgroup (n = 33) was .43 ($p < .001$), with explained variance of 18.4%. And the predictability of preadmission GPA for clinical performance was .37 ($p < .001$), with explained variance of 13.7%. For the preadmission GPA subgroups, low, intermediate, and high achievers, correlations between preclinical performance and clinical performance were .32, .38, and .52, respectively, ($p < 0.01$) with explained variances of 10.4%, 14.5%, and 27.0%, respectively. Taken together, these findings imply that the predictability of performances during medical school is dependent on the grade-subgroup.

Consistency of performance for all grade-subgroup was compared by using Cronbach's alpha coefficient of internal consistency. This internal consistency reliability coefficient was for the low, intermediate, and high grade-subgroup, .362, .488, and .649, respectively. The higher grade-subgroups performed most consistent.

Repeated measures analysis of variance showed a main effect of curriculum, F (2, 652) = 22.07, $p < .001$, indicating that the different subgroups are performing differently per curriculum phase. Furthermore, a significant subgroup by curriculum interaction was found, F (4, 652) = 39.21, $p < .001$. This demonstrated that there were differences within and across grade-subgroups and the slope within each grade-subgroup was dependent on the curriculum phase. Figure 1 shows grade as a function of curriculum phase for all grade-subgroups using standardized grades.

One-way ANOVA's for all subsequent phases in the curriculum revealed this decreasing effect. F-ratio's were for preadmission performance, preclinical performance, and clinical performance, 814.52, 52.77, and 9.95, respectively, (*df* = 2, 326), $p < .001$. Employing pairwise comparisons between the means of the grade-subgroups in all subsequent phases of the curriculum revealed that all subgroups are significantly different from each other, except between the means of the intermediate and high achievers subgroup during clinical performance.

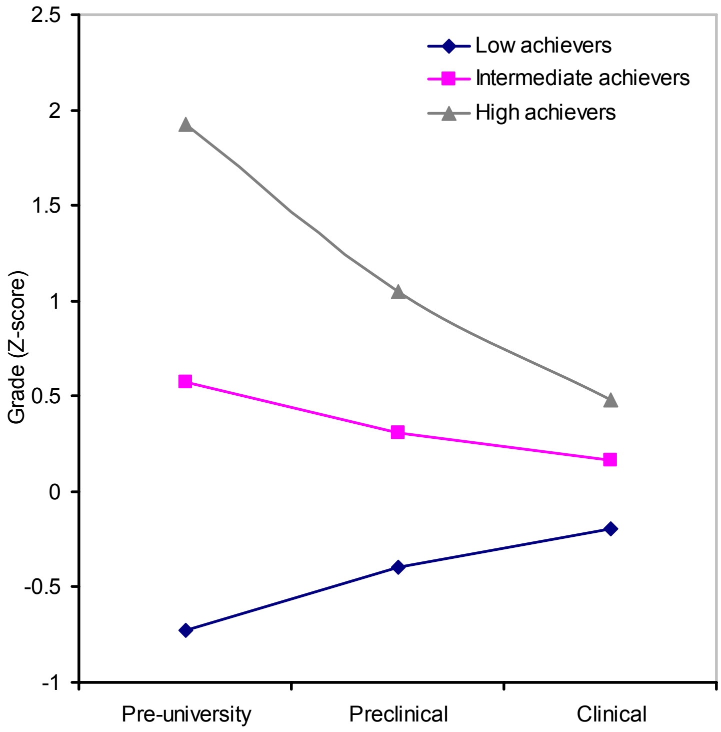


Figure 1. *Level of grades (in Z-scores) as a function of curriculum phases for all subgroups*

Discussion

Preadmission GPA is one of the most widely used indicators for academic performance. Most studies using GPA as a predictor analyzed a whole cohort and were mostly successful for predicting preclinical performance (Ferguson et al., 2002; Gottheil & Miller-Micheal, 1957; Julian, 2005; Mitchell, 1990; Peat et al., 1982; Ronai et al., 1984; Salvatori, 2001; Wiley & Koenig, 1996). In this study, we analyzed the relationship between student preadmission performance and performance during different phases of medical school from a grade-subgroup perspective. In order to investigate the influence of curriculum transitions on subgroup performance, consistency of performance was compared between subgroups.

The findings indicated an expected good predictive value of preadmission GPA for preclinical performance if we consider performance of the entire cohort. Preadmission GPA as predictor of clinical performance was much lower and in line with previous studies: six percent of the variance could be explained (see, for example, the study of Peat et al., 1982). Preadmission performance and preclinical performance are most likely bound by common abilities, as Norman (2004) pointed out: they are probably an index of both intelligence and motivation, as well as an indication of the ability to master topics essential for medicine. Clinical clerkships, on the other hand, are mainly practical learning environments that differ greatly from the preclinical learning environment, in which other characteristics are probably accountable for successful performance, such as interpersonal skills, effective working relationship with colleagues and supervisors, ability to work with patients, clinical problem-solving skills, or appropriate professional behavior (Mann, Ruedy, Millar, & Andreou, 2005; Murden et al., 1978; Stern, Frohna, & Gruppen, 2005). The transition from preclinical to clinical education is often described as difficult and stressful for students. The loss of predictive power of preadmission GPA for this phase in the medical curriculum can be attributed to the change of environment. There is research evidence that students often encounter difficulties bridging the gap from theory to practice (Boshuizen, 2005; Prince et al., 2000). In addition, it is well described in the literature that clinical assessment is different from classroom learning and tailored to the clinical setting (Van der Vleuten et al., 2000).

Are the above listed relationships between student preadmission performance and performance during different phases of medical school also applicable for the different subgroups as identified by their level of preadmission GPA? Are the correlations calculated for the entire cohort perhaps covering up deviant behavior of distinctive subgroups? A closer look at performance patterns within subgroups revealed different insights and can be summarized as follows: Performance of the low-achiever group is not predictable and inconsistent; performance of the intermediate group is predictable for the preclinical phase, while the performance of the high-achiever subgroup is best predicted by preadmission GPA and is less influenced by the effects of curriculum transitions. In comparing the Cronbach's alpha coefficients among the subgroups then there could be concluded that level of consistency of performance is related to level of preadmission GPA.

Performance of the low-achiever subgroup is not predictable. Preadmission GPA is not a good predictor for more than half of the student population. This could indicate that other, perhaps non-cognitive, variables are better predictors for performance in this subgroup (Albanese, Snow, Skochelak, Hugett, & Farell, 2003; Moulaert, Verwijnen, Rikers, & Scherpbier, 2004; Webb et al., 1997). Performance of the intermediate grade subgroup seems predictable only during preclinical years, not for the clinical years. On the other hand, it seems, based on the data of the present study, that academic performance for both phases within medical school is predictable by preadmission GPA of the high achievers. In addition, the high-achievers are consistent performers during medical school. The question of interest here is why the high achiever group is able to maintain this high consistency. If indeed different characteristics are important during different curriculum phases, then does this mean that this group has it all? Or does it mean that the emphasis of assessment during clinical training is also based on cognitive related components of competence?

By combining our findings with the findings of above mentioned studies of Boshuizen (2005) and Prince et al., (2000), which describe the transition from theoretical to clinical training as problematic, it could be concluded that the transition difficulties are related to the level of preadmission GPA and probably related in some way to cognitive abilities of the individual student. Students reported negative experiences in the clinical phase to be related with professional socialization and difficulties in the use of knowledge and skills. Furthermore, it is suggested that students adapt their learning methods to the assessment program of the curriculum phase (Prince et al., 2000). High-achieving students seem better equipped for those changes. Low-achieving students seem most vulnerable and highly influenced by curriculum transitions. The dropout ratios are highest for this grade-subgroup (80% of all dropouts belong to this grade-subgroup). The intermediate group seems to perform consistently till they reach the clinical phase; the predictive value of preadmission GPA is lost after this transition. The high achieving students drop out hardly at all and stay the better performers during medical school, but their performance seems to drop more as expected by a ceiling-effect. The, for this group, perhaps less stimulating curriculum could be partly responsible for this declining effect. In addition, the consistency of performance for this latter group could indicate that cognitive abilities are still highly valued and rewarded in the clinical phase.

In comparing performance levels among the subgroups between the different phases, then the group differences are getting smaller. For example, clinical performance is not significantly different anymore between the intermediate- and high-achieving subgroup. The effect of the curriculum seems to make the group differences smaller.

Limitations of this study that should be considered are related to the reliability of the predictor variable. Despite the widely use of grades as predictor and criterion, the consistency and meaningfulness of the use of preadmission GPA could be questioned, partly because grades are often dependent on the specific school or program. In our case, the effect of school is low because preadmission GPA is determined by a national examination in the final year. The grade average, however, is a composite of six grades from several sources: four required classes (mathematics, physics, chemistry, and biology) and two classes that could vary per individual student (e.g., philosophy, arts, history).

In conclusion, learning is a complex interaction between the student and its environment. Studies that search for reliable cognitive predictors for student performance during medical school consider often the whole cohort. Individual students behave and react differently to the changing learning environment, and the results of this study indicated that different performance characteristics are applicable for different students. The level of performance is not only an indicator of students' cognitive abilities but also representative of level of predictability and consistency of performance during medical school.

While this study design, at first sight, may not appear to be of value to anyone outside of the Netherlands, studying the Dutch system gave the opportunity to focus on performance of students with relatively low grades. Identifying subgroup-dependent characteristics will open the door to a more specific perspective on student performance and its predictability. In addition, it could lead to a better understanding of the impact of curriculum transitions on student learning. The learning environment is often designed based on the 'average' students and is too uniform for a diverse cohort. Identifying these specific, student-related characteristics early in their educational career could be helpful, for example, in providing a more effective schooling for students at risk; not only for students with relatively low grades, as identified in this study, but also for students recruited to represent ethnic minorities. Many medical schools enroll students from less fortunate backgrounds to promote cultural and ethnic diversity as represented by a modern society. Further research should therefore be directed at investigating which combinations of factors could be used to define groups of students with specific behaviors and study patterns.

References

Albanese, M. A., Snow, M. H., Skochelak, S. E., Hugett, K. N., & Farell, P. M. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine, 78*(3), 313-321.

Boshuizen, H. P. A. (2005). Does practice make perfect? A slow and discontinues process. In H. P. A. Boshuizen, R. Bromme & H. Gruber (Eds.), *Professional Learning: Gaps and Transistions on the way from Novice to Expert (Vol. 2)*. Dordrecht / Boston / London: Kluwer Academic Publishers.

Chan-Ob, T., & Boonyanaruthee, V. (1999). Medical students selection: which matriculation scores and personality factors are important? *Journal of Medical Association of Thailand, 82*, 604-610.

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

Gottheil, E., & Miller-Micheal, C. (1957). Predictor variables employed in research on the slection of medical students. *Journal of Medical Education, 32*(2), 131-147.

Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine, 80*, 910-917.

Koenig, J. A., Sireci, S. G., & Wiley , A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine, 73*, 1095-1106.

Korman, M., Stubblefield, R. L., & Martin, L. W. (1968). Patterns of success in medical school and their correlates. *Journal of Medical Education, 43*, 405-411.

Lynch, T. G., & Schneider, P. D. (2000). Availability of content information does not improve performance on computerized case simulations. *The American Journal of Surgery, 179*(6), 472.

Mann, K. V., Ruedy, J., Millar, N., & Andreou, P. (2005). Achievement of non-cognitive goals of undergraduate medical education: perceptions of medical students, residents, faculty and other health professionals. *Medical Education, 39*(1), 40-48.

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

Moulaert, V., Verwijnen, M. G. M., Rikers, R. M. J. P., & Scherpbier, A. J. J. A. (2004). The effects of deliberate practice in undergraduate medical education. *Medical Education, 18*, 1044-1052.

Murden, R., Galloway, G. M., Reid, J. C., & Colwill, J. M. (1978). Academic and personal predictors of clinical success in medical school. *Journal of Medical Education, 53*(711-719).

Muzzin, L. J., & Hart, L. (1985). Oral examinations. In V. R. Neufeld & G. R. Norman (Eds.), *Assessing clinical competence*. New York: Springer-Verlag.

Norman, G. R. (2004). Editorial - The morality of medical school admissions. *Advances in Health Sciences Education, 9*, 79-82.

Peat, M., Woodburry, M. G., & Donner, A. (1982). Admission average as predictor of undergraduate academic and clinical performance. *Physiotherapy Canada, 34*, 211-214.

Perrot, L. J., Deloney, L. A., Hastings, J. K., Savell, S., & Savidge, M. (2001). Measuring student motivation in health professions' colleges. *Advances in Health Sciences Education, 6*, 193-203.

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). Qualitative analysis of the transition from theory to practice in undergraduate training in a PBL-medical school. *Advances in Health Sciences Education, 5*(2), 105-116.

Rheault, W. (1988). Relationship between academic achievement and clinical performance in a physical therapy education program. *Physical Therapy, 68*(3), 378-380.

Ronai, A. K., Golman, M., & Shanks, C. A. (1984). Relationship between past academic performance and results of specialty in-training examinations. *Journal of Medical Education, 59*, 341-344.

Rothman, A. I. (1973). A comparison of persistent high and low achievers through four years of undergraduate medical training. J*ournal of Medical Education, 48*, 180-182.

Salvatori, P. (2001). Reliability and Validity of Admissions Tools Used to Select Students for the Health Professions. *Advances in Health Sciences Education, 6*, 159-175.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York, NY: Springer Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On Acquiring Expertise in Medicine. *Educational Psychological Review, 5*(3), 205-221.

Stern, D. T., Frohna, A. Z., & Gruppen, L. D. (2005). The prediction of professional behaviour. *Medical Education, 39*(1), 75-82.

Van der Vleuten, C. P. M., Scherpbier, A. J. J. A., Dolmans, D. H. J. M., Schuwirth, L. W. T., Verwijnen, G. M., & Wolfhagen, H. A. P. (2000). Clerkship assessment assessed. *Medical Teacher, 22*(6), 592-600.

Webb, C. T., Sedlacek, W., Cohen, D., Shields, P., Gracely, E., Hawkins, M., et al., (1997). The impact of nonacademic varables on performance at two medical schools. *Journal of the National Medical Association, 89*(3), 173-180.

Wiley, A., & Koenig, J. A. (1996). The validity of the Medical College Admission Test for predicting performance in the first two years of medical school. *Academic Medicine, 71*(10 Suppl.), S83-S85.

Woloschuk, W., Harasym, P. H., & Temple, W. (2004). Attitude change during medial school: a cohort study. *Medical Education, 38*(5), 522-534.

CHAPTER THREE

Influence of Clerkship Experiences on Clinical Competence

This study is conducted at and funded by Erasmus MC—University Medical Center Rotterdam (EUR).
An earlier version of this paper is presented at the Annual Meeting of the Association for Medical Education in Europe (AMEE), Bern, Switzerland, August 31 – September 3, 2003 with the title: Modeling Clinical Competence in a Medical Internship: The Impact of Variation in Real Clinical Experiences.

Reference
Wimmers, P. F., Schmidt, H. G., & Splinter, T. A. W. (2006). Influence of clerkship experiences on clinical competence. Reprinted from *Medical Education, 40*(5), 450-458, © 2006 with kind permission of Blackwell Publishing.

Abstract

BACKGROUND Clerkship experiences are considered crucial for the development of clinical competence. Yet whether there is a direct relationship between the nature and volume of patient encounters and the learning outcome is far from clear. Some evidence in the literature points towards the importance of clinical supervision on student learning, but the relationship between clinical supervision, patient encounters, and student competence remains unclear.

PURPOSE This study aimed firstly to determine the variation of students' clinical experiences within and across sites; secondly, to identify the causes of this variation, and thirdly, to investigate the consequences of this variation on students' competence.

METHODS Clerkship students at 12 hospital sites recorded their patient encounters in logbooks. Site characteristics that might influence the variation in patient encounters were collected. Student competence was determined by 3 independent indicators: a practical end-of-clerkship examination; a theoretical end-of-clerkship examination, and an evaluation of professional performance. A model was developed to test the available clerkship data using Structural Equation Modeling (SEM) software.

RESULTS Analysis of the logbooks revealed a large variation in the number of patients encountered by students. The average patient length of stay, number of patients admitted, and quality of supervision accounted partly for this variation. An increased number of patient encounters did not directly lead to improved competence. Quality of supervision turned out to be crucially important because it directly impacted student learning and also positively influenced the number of patient encounters.

CONCLUSION Monitoring the effectiveness of clerkships by merely asking students to keep a tally of the problems and diseases they encountered, without paying attention to the quality of supervision, does not contribute towards improving student learning.

THE MEDICAL CURRICULUM is often designed according Flexner's two-stage model with a sharp division between basic medical science and clinical practice (Flexner, 1910). Medical education in the preclinical years is predominantly theoretical and students learn basic medical science, such as biochemistry, anatomy, physiology, and pathology. The following years are centered on clinical experience. The patient encounter is thought to be crucial during that period and is considered to be the essence of clinical education (Neufeld & Norman, 1985; Snell, Battles, Bedford, & Washington, 1998; Witzke, Koff, McGeagh, & Skinner, 1990). This assumption is widely accepted and seems logical: to become a baker you have to practice baking bread.

The structured preclinical years of medical education contrast with the less structured clinical years. Throughout preclinical training, all students at a school complete the same program, study the same material, read the same books and take the same examinations. The clinical years, on the other hand, are less uniform; clerkship rotations are carried out at different hospital sites, with different supervisors and with variations in patient mix. This variation in clinical education between hospital sites and between students remains an eyesore of medical education. It highlights the need for reform and the importance of uniform standards in the current structure of the clerkship model (Nutter & Whitcomb, 2001).

One of the first attempts to examine and structure the content of clinical clerkships involved the introduction of patient logbooks. These logbooks have proven their value in comparing and identifying intersite and interstudent differences in patient encounters and have shed light on the range of signs, symptoms and diseases seen during rotations (Dolmans, Schmidt, Van der Beek, Beintema, & Gerver, 1999; Ferrell, 1991; Gruppen, Wisdom, Anderson, & Woolliscroft, 1993; Hobbs, Mongam, & Miller, 1987; McGraw & Lord, 1997; McLeod & Snell, 1991; Patricoski, Shannon, & Doyle, 1998; Raghoebar-Krieger & Bender, 1997). In addition, it has been demonstrated that students' knowledge increases significantly during clerkships (Butterfield & Libertin, 1993). Schwartz, et al., (1994) confirmed that students who started with approximately the same knowledge levels on the pretest ended with very different knowledge levels on the post-test. This indicates that learning environment has consequences for learning outcomes. Nevertheless, finding a direct relationship between the volume of patient encounter and the learning outcomes of students is more difficult. Studies that have attempted to find correlations between the number or variation of patients seen during clerkships and performance in the end-of-clerkship examinations have been unable to do so (McManus, Richards, Winder, & Sproston, 1998; Van Leeuwen et al., 1997). The observed lack of correlations was attributed to the nature of the examinations given: the written knowledge-based examinations may have failed to assess skills and knowledge gained from clinical experience. Gruppen et al., (1993) assessed students' clinical knowledge at the end of a clerkship with what was probably a more appropriate test. The investigators used a diagnostic recognition test of common clinical problems and related the scores to the clinical experiences of the students. Although the increase in learning outcomes was significant, again no correlation with clinical experience was found.

Despite the results of the aforementioned studies, students consider that exposure to real patients is important. Dolmans et al., (2002), for example, evaluated medical students' perceptions of the effectiveness of clerkships. The degree of perceived effectiveness depended on the number and variation of patients, and the quality of supervision. It is interesting to note that the role of the clinical supervisor was considered even more important when the number of patients and variation of diseases was low. Châtenay et al., (1996) failed to find any differences between students with little clinical experience and students with much clinical experience on end-of-clerkship examinations, but found examination performance to be influenced by a combination of much clinical experience in emergency admissions and feedback given by the supervisor. A positive effect of supervision is also supported by Griffith et al., (1997, 1998) who found that teaching quality (an aspect of supervision) and supportive housestaff improved student performance during clerkships.

In summery, the variation in students' clinical experiences during clerkships is a frequently identified problem, but the site characteristics responsible for this variation are largely unknown. Previous researchers who have attempted to find direct relationships between patient encounters and student competence were unsuccessful. Some evidence in the literature points towards the importance of clinical supervision on student learning, but the relationships between clinical supervision, clinical encounters, and student competence remains unclear. In the present study we sought to explain the causes of these intersite variations in patient encounters. Moreover, we investigated the influence of this variation in patient encounters and overall quality of supervision on students' clinical competence. As indicators of clinical competence, we used the outcomes of the practical end-of-clerkship examination, the theoretical end-of-clerkship examination, and a rating of the professional performance of the student by the supervisor on site. In a stepwise procedure, two models were devised that explained the causes and consequences of the variation in clinical experiences. These two models were brought together in a final "clerkship competency model." All models were developed and tested with structural equation modeling (SEM) techniques (Bentler, 2003).

Method

Participants

A total of 227 students rotated through the internal medicine clerkship at Erasmus MC—University Medical Center Rotterdam, the Netherlands during the academic year 1999/2000. This 12-week clerkship takes place after 4 years of preclinical instruction and is the first of 10 different clerkships spread over a 2-year period. For this clerkship, the students were divided over 14 different hospital sites, consisting of three sites at the academic hospital and 11-affiliated hospital sites. Institutional approval has been was received for this study.

Materials and procedure

Students were instructed to document clinical patient encounters in a logbook. The pocket logbook provided the instructions for how to complete the patient encounter data. Patient encounters were documented by date, date of birth, gender and diagnoses. The supervisor of the site had to sign every logged encounter on a daily basis.

All students had to complete an evaluation form at the end of the clerkship consisting of five questions about the quality of supervision. The first two questions were scored on a four-point Likert scale and the next three were open questions: (1) How was the supervision of the attending physician (resident)? (2) How was the supervision of the clinical supervisor? (3) How often did an educational discussion take place with the clinical supervisor? (4) How often did you have a performance interview? (5) Are the implications of the patient history, physical examination and clinical encounter reports discussed? In order to obtain a global score of level of quality of supervision the mean of all answers was used.

To explain the causes of inter-hospital variation in patient encounters the following site information was collected: number of beds, number of beds used for educational purposes, number of staff, average length of patient stay, number of patient admissions, site occupancy, and number of peer clerks.

Assessment of applied knowledge

The end-of-clerkship examination comprised two independent parts. The practical part consisted of the completion of a patient file, including history taken, physical examination and diagnosis of a patient. This examination was taken at the hospital site in week 11 and judged by the clinical supervisor. The theoretical part of the examination covered at least five different topics related to internal medicine. An authorized teacher administrated this examination orally at the home university and provided the student with a final grade based upon overall performance during this examination.

Assessment of professional performance

The clinical supervisors rated the clinical performance of the student every three weeks during the clerkship. This rating scale consisted of seven items, scored on a four-point Likert scale: (1) the student's general attitude towards clinical practice; (2) the quality of the student's interaction with patients; (3) the quality of the student's interaction with staff; (4) the amount of student's available knowledge; (5) the student's competence in taking a case history; (6) the student's competence in presenting the patient history; and (7) the student's general clinical competence. All items together give an indication of students' clinical performance; therefore the mean of all answers was used. Reliability analysis of the rating scale based on standardized items = .85 (Cronbach's alpha).

Students' clinical competence was determined by the 3 indicators mentioned previously (Figure1): the practical end-of-clerkship examination, the theoretical end-of-clerkship examination, and professional performance.

Analysis of the logbook data

The logbooks were analyzed and the number of patients, the number of diseases, and the variety of diseases were derived from the logbook data. The number of diseases refers to the incidence of diseases of the whole period of 12 weeks, regardless of repetition. Variety of diseases, on the other hand, refers to the number of unique diseases encountered by the student. The diseases were compared with a short list of most common diseases in internal medicine. This short list of 52 diseases is based on the appendix of the Dutch national objectives for medical education in 2001. This appendix lists almost 250 different diseases considered to be important because of common occurrence or degree to which they are life-threatening (Metz, Verbeek-Weel, & Huisjes, 2001). This list is designed to represent the exemplary diseases that medical students must have seen and/or be familiar with by the time they graduate. Our list of 52 diseases was coded in order to statistically analyze the patient encounters as documented in the logbooks.

Statistical Analysis

The analysis was carried out in two successive steps: The first step involved defining the variation in terms of number of patients, quality of supervision, and end-of-clerkship examinations. The analysis was conducted by using one-way ANOVAs. Fisher's least significant difference (LSD) was used to make post hoc comparisons between the different groups. An alpha level of .05 was used for all statistical analyses.

In the second step, a model was developed to test the available clerkship data of the students using a Structural Equation Modeling (SEM) software package EQS version 6.1. (Bentler, 2003). The advantage of this statistical methodology is that models can be built that reflect complex relationships because any variable, whether observed or not observed, can be used to predict any other variable. In a complex learning environment, like a clerkship, it is almost impossible to directly measure complex constructs such as "clinical competence." A factor stands for such a construct and is measured by other indicator variables indirectly. The relationships between the observed variables and the unobserved variable or factor are defined in terms of regression weights, here called path coefficients.

The SEM software package (Bentler, 2003) could produce several fit indices indicating how well the tested model reflects the observed covariance structure. Fit indices are arranged according class: absolute, parsimonious, and incremental (Hu & Bentler, 1999). Every class of fit indices reflects somewhat different facets of the model fit; therefore fit indices from different classes are used to support model's acceptability. The fit indices used in this study are the model chi-square statistic (χ^2), whereby a nonsignificant value indicates that the observed and model-implied covariance matrices are not significantly different. The Bentler Comparative Fit Index (CFI), the Bentler-Bonett Normed Fit Index (NFI), and the Jöreskog-Sörbom Goodness-of-Fit Index (GFI), must all be above .90 to indicate good fit. Finally, Steiger and Lind Root Mean Square Error of Approximation (RMSEA) should have a value below .06 to indicate proper fit (Kline, 1998).

Results

Logbook analysis

Altogether, 179 of the 227 students completed their logbooks, giving a compliance rate of 78.9%. A total of 24 of the 179 returned logbooks were excluded from the study, for the reason that these logbooks were incomplete or had not been used by the students for the entire clerkship period. The remaining 152 students encountered an average of 43.3 patients (SD = 21.5) and diagnosed an average of 32.4 diseases (SD = 17.8). The average number of different diseases encountered was 16.3 (SD = 5.9). Note the large *SD*'s. Descriptive statistics per site are presented in Appendix A.

Inter-hospital variation

The number of patients, the number of diseases, and variation of diseases encountered, differed all significantly between hospital sites (F-values are 16.924, 13.377, and 12.593, respectively, df 13, 138, $p < .01$). Post hoc multiple comparisons between the sites revealed that for the number of patients, each site significantly differed, on the average, from 8 other sites (range 4 till 13); for number of diseases, this was 7.5 (range from 5 till 13); and for the variety of diseases encountered, this was 7.6 (range from 4 till 13).

Quality of supervision was also significantly different between hospital sites, F (13, 138) = 1.847, $p < .05$. Multiple comparisons revealed that each site differed significantly, on average, from 1 other site (range 0 till 5).

The assessment measures of clinical competence indicated that only the practical end-of-clerkship examination was significantly different between hospital sites, F (13, 138) = 2.327, $p < .05$. Multiple comparisons revealed that each site differed significantly, on the average, from 2.6 other sites (range 0 till 10). No significant differences were found between hospitals for the other assessment measures of clinical competence (i.e., the theoretical end-of-clerkship examination and professional performance).

The Clerkship Competency Model

A stepwise procedure, as recommended by Jöreskog and Sörbom (1999), was used to come to a final hypothesized model. Two different models are tested that would be nested in a final hypothesized model as displayed in Figure 1. The first model focused on the causes of the variation in the number of patients encountered; the second model focused on the consequences of the variation in the number of patients encountered for clinical competence.

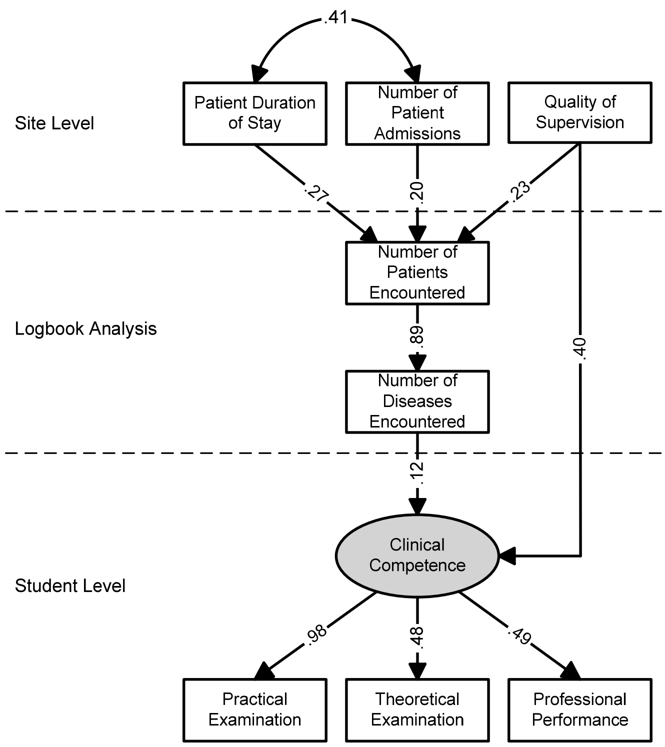


Figure 1. *Clerkship Competency Model*

Focusing on explaining the intersite variation in actual number of patients seen, we initially brought all possible hospital site information into a model as cause indicators of number of patients encountered. This model did not fit the data well and the Wald test for dropping parameters indicated that the variables, "total number of beds," "number of staff," "number of peer clerks," and "site occupancy" did not have a significant influence on the variable "number of patients encountered," and should be removed from the model to increase fit. The Wald statistic estimates the amount the model's overall χ^2 would increase if a particular free parameter is eliminated or dropped from the model (Kline, 1998). In addition, after rerunning this model, the Lagrange Multiplier test for adding parameters specified to add the parameter between "number of patient admissions" and "average patient length of stay." Lagrange Multiplier approximates the amount by which the model's overall χ^2 would decrease if a particular parameter were freely estimated. The overall fit of the model would increase if the parameter were added to the model (Kline, 1998). The resulting model with three cause indicators loading on the variable "number of patients encountered" showed proper fit (see "site level" in Figure 1). The χ^2 statistic (2, $N = 152$) = 2.03, $p = .36$, is indicating that the proposed model is not significantly different from a perfect fit. The other fit statistics are the Comparative Fit Index (CFI), the Normed Fit Index (NFI), and the Goodness-of-Fit Index (GFI). The values of CFI, NFI, and GFI in the hypothetical model are 1.00, 1.00 and .99, respectively (values above .90 indicate a good fitting model). The other fit statistic, which focuses on the Root Mean Square Error of Approximation (RMSEA), has a value of .011, also indicate good fit also (the value must be below .06).

The second model is developed to test whether the number of patients and the quality of supervision could have an impact on the clinical competence of students (see "student level" in Figure 1). This model is based around a latent variable, labeled "clinical competence," with three indicator variables: "practical end-of-clerkship examination," "theoretical end-of-clerkship examination," and "professional performance." The reliability coefficient of this measure of clinical competence, as if they were 3 items on a single scale, is .67 (Cronbach's alpha based on standardized items). The variable "number of patients encountered" is a cause indicator of "number of diseases encountered," (see "logbook analysis" in Figure 1). These two variables are closely related; most patients' log entries are marked with a single disease. Sometimes the student filled in more diseases and in some cases the student left the disease entry of the logbook blank. The path coefficient between these two variables is .89, $p < .05$. The variables "number of diseases encountered" and "quality of supervision" are cause indicators of the latent variable "clinical competence." This model's fit indices indicate a borderline fit. The χ^2 statistics (8, $N = 152$) = 16.29, $p = .04$. The values of CFI, NFI, and GFI are .96, .98 and .97, respectively. The RMSEA has a value of .083. The Lagrange Multiplier test for adding parameters specified that we should add the parameter between "quality of supervision" and "number of patients encountered" to improve significantly the χ^2 statistics with 11.31. The readjusted model has good fit indices. The χ^2 statistic (7, N = 152) = 4.54, p = .72. The values of CFI, NFI, and GFI are respectively 1.00, .99, and .99. The RMSEA has a value of .000.

The two models discussed earlier are nested within the final internal medicine clerkship competency model. The generated correlation matrix, means and standard deviations of all observed variables used in the final model are presented in Appendix B. The fit statistics indicate that our final hypothesized model represents the causal processes between the variables well. The χ^2 statistic (17, N = 152) = 22.74, p = .16. The values of CFI, NFI, and GFI are .99, .95, and .97, respectively. The RMSEA has a value of .047.

Let us discuss the model of Figure 1 in more detail. The variables used in this model could be categorized in three different levels, labeled as site level, logbook analysis level, and student level. Variables from the first level are: ”patient length of stay,” “number of patient admissions,” and “quality of supervision;” they have factor loadings of .27, .20, and .23, $p < .05$, respectively, with the variable “number of patients encountered.” The percentage explained variance by these predictors is 21.7%. This first level is based on our first model, discussed above. The variables of the following levels of the model in Figure 1, logbook analysis level and student level, are based on the logbook data and their consequence on student’s competence. The variables are: “number of patient encountered,” “number of diseases encountered,” “practical end-of-clerkship examination,” “theoretical end-of-clerkship examination,” and “professional performance.” The variables, “practical end-of-clerkship examination,” “theoretical end-of-clerkship examination,” and “professional performance” are the indicator variables of the unobserved construct or factor, named “clinical competence,” with factor loadings of respectively, .98, .48, and .49, $p < .05$. In addition, the measured variables, “quality of supervision,” with a factor loading of .40, $p < .05$, and the “number of diseases encountered,” with a factor loading of .12, $p > .05$, are both cause indicators of clinical competence. However, only the influence of quality of supervision was significant. Therefore, variation in the latent variable “clinical competence” is caused (partly) by variation in the observed variable “quality of supervision.” The explained variance of the latent variable “clinical competence” is 19.3%.

Discussion

As it is generally assumed that a sufficient number of patient encounters is essential for the development of clinical competence (Bentler, 2003; Neufeld & Norman, 1985; Snell et al., 1998; Witzke et al., 1990) large variations in patient encounters (Châtenay et al., 1996; Ferrell, 1991; Gruppen et al., 1993) must have far-reaching consequences for student learning. Studies examining the relative growth in knowledge gained during clerkships confirmed significant increases (Butterfield & Libertin, 1993; Schwartz et al., 1994). Students who started with approximately the same knowledge levels ended with very different knowledge levels at the end of the clerkship. This indicates that the learning environment does have consequences for the learning outcome. However, no previous studies have found a direct relationship between the number of clinical encounters during rotations and clinical competence (Châtenay et al., 1996; Gruppen et al., 1993; McManus et al., 1998; Van Leeuwen et al., 1997).

The purposes of the present study were, after determining the degree of variation of students' clinical experiences within and across sites, to explain the causes of this intersite variation in clinical experiences and, in addition, to investigate the consequence of such variation in clinical experiences and quality of supervision on clinical competence.

The present study confirmed that differences between hospital sites are bigger than expected given the differences within sites. Hospital sites account for a large amount of the variation in the number of patients encountered by students during their clerkships. The search for possible causes of this intersite variation within the hospital itself revealed that, of all variables measured, only length of patient stay, number of patient admissions, and quality of supervision significantly explained the variation in patients (explained variance in this study: 21.7%). Despite this variation in patient encounters, the consequence for student competence was limited. This casts doubt on the overall assumption that exposure to a sufficient number of patient encounters is essential for the development of clinical competence. Instead, our clerkship competency model (Figure 1) showed that quality of supervision has a significant and direct effect on students' clinical competence. Griffith and colleagues investigated the direct impact of the quality of teaching on student performance during clerkships and found that the difference in the scores on pre- and postclerkship examinations of the students of the "best teachers"was significantly higher than the equivalent differences for other students. Even the supportive guidance of other housestaff lead to better performance (Griffith et al., 1997, 1998). However, according to student perceptions, the effectiveness of clinical rotations was not caused by a single indicator but was dependent on a combination of patient mix and supervision. The degree of perceived effectiveness depended on the number and variation in patients, and the quality of supervision. The role of the clinical supervisor was considered more important when the number of patients and variability of diseases was low (Dolmans et al., 2002). Our results support the evidence for supervision's crucial role during clerkships. It appears that the clinical supervisor can have a stimulating effect on student learning and student learning environment, resulting in more patient encounters.

An interaction between encountered patients and supervision was also found in the study of Châtenay et al., (1996). Better performances at examinations related to high-volume experience in combination with feedback. Such data are in agreement with Ericsson's deliberate practice model of expertise. Deliberate practice is characterized by training that is highly structured and closely related to the level of acquired performance. Practice alone is not enough: hours spent on practice in general had no correlation with level of expertise. Repetitive experiences in combination with supervised training are key elements of expertise. It takes a long period for medical students to acquire the knowledge and skills needed to become a physician. Equally, a long period of supervised training is needed to facilitate the gradual acquisition of the skills required to treat patients (Ericsson, 2004; Ericsson & Charness, 1994).

The limitations of this study concern the reliability of the logbook data, the assessment of clinical competence, and the assessment of the quality of supervision. Previous research has shown that students seem to have an overall tendency to under-report the number of patients and diseases seen (see for details Raghoebar-Krieger, Sleijfer, Bender, Stewart, & Popping, 2001). The significant intersite variation in patient encounters, however, is not dependent on students' willingness to fill in logbooks and even doubling the number of patients for each student will not correlate with a significant increase of clinical competence.

The individual indicators used for the assessment of clinical competence, which were based on an oral examination, may be questioned for their reliability. However, similar measures are used for the remaining 9 clerkships and the degree of internal consistency among these 10 clerkship finals, as if they were items on a single scale, would be .82 (Cronbach's alpha), indicating fairly strong internal consistency despite different clerkships and many different examiners (observation of 227 students, Wimmers, Splinter, Hancock, & Schmidt, in press, 2006), (See also, Daelmans, Scherpbier, Van der Vleuten, & Donker, 2001). With respect to the latent construct of clinical competence, we used the covariation among 3 independent indicators. The reliability coefficient for these 3 indicators, as if they were items on a single scale, would be .67, indicating an acceptable reliability. Although our findings can be supported with other current research, the low correlations between number of patient encounters and clinical competence may be caused by the unreliability of our measures.

Measurement of "quality of supervision" is based on student's ratings of 5 different items. All answers were averaged to obtain an indicator of the overall quality of supervision. A more precise instrument is needed to identify what aspects of supervision in particular affect student learning.

Conclusions

Although, the learning environments offered by most hospital sites are not specially designed for educational purposes we are convinced that clinical clerkships are an essential part of the learning experience medical students require if they are to become competent physicians. However, if the learning effect of merely seeing patients is questionable and the supervised training provided in most hospital sites is limited, then Flexner's legacy for medical education (Flexner, 1910) may also be questioned. This evokes some interesting questions for further research. What and how do students learn from patient encounters and, more specifically, what other aspects of clinical supervision are needed for an optimal learning effect? Which other components of the hospital environment contribute to the development of students' clinical competence? In order to turn hospital sites into effective learning environments, answers to these questions should be put into a new model based on "deliberate practice," whereby clinical training is highly structured and closely related to the level of acquired performance: effective repetitive experiences in combination with high-quality supervised training.

References

Bentler, P. M. (2003). EQS (Version 6.1 Beta) [Computer Software]. Encino, CA: Multivariate Software, Inc.

Butterfield, P. S., & Libertin, A. G. (1993). Learning outcomes of an ambulatory care rotation in internal medicine for junior medical students. *Journal of General Internal Medicine, 8*(4), 189-192.

Châtenay, M., Maguire, T., Skakun, E., Chang, G., Cook, D., & Warnock, G. L. (1996). Does the volume of clinical experiences affect performance of clinical clerks on surgery exit examinations? *The American Journal of Surgery, 172*(4), 366-372.

Daelmans, H. E. M., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., & Donker, A. J. M. (2001). Reliability of clinical oral examinations re-examined. *Medical Teacher, 23*(4), 422-424.

Dolmans, D. H. J. M., Schmidt, A., Van der Beek, J., Beintema, M., & Gerver, W. J. (1999). Does a student log provide a means to better structure clinical education? *Medical Education, 33,* 89-94.

Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., Essed, G. G. M., Scherpbier, A. J. J. A., & Van der Vleuten, C. P. M. (2002). The impacts of supervision, patient mix and numbers of students on the effectiveness of clinical rotations. *Academic Medicine, 77*(4), 332-335.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10 Suppl), S70-S81.

Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychology, 49*(8), 725-747.

Ferrell, B. G. (1991). Demonstrating the efficacy of the patient logbook as a program evaluation tool. *Academic Medicine, 66*(9 Suppl), S49-S51.

Flexner, A. (1910). *On Medical Education in the United States and Canada.* New York: Carnegie Foundation for the Advancement of Teaching.

Griffith, C. H., Wilson, J. F., Haist, S. A., & Ramsbottom-Lucier, M. (1997). Relationships of how well attending physicians teach to their students' performances and residency choices. *Academic Medicine, 72*(10 Suppl), S118-S120.

Griffith, C. H., Wilson, J. F., Haist, S. A., & Ramsbottom-Lucier, M. (1998). Do students who work with better housestaff in their medicine clerkships learn more? *Academic Medicine 73*(10 Suppl), S57-S59.

Gruppen, L. D., Wisdom, K., Anderson, D. S., & Woolliscroft, J. O. (1993). Assessing the consistency and educational benefits of students' clinical experiences during an ambulatory care internal medicine rotation. *Academic Medicine, 68*(9), 674-680.

Hobbs, J., Mongam, P., & Miller, M. D. (1987). A system for assessing clerkship experience using a logbook and microcomputers. *Family Medicine, 19*(4), 287-290.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-31.

Jöreskog, K. G., & Sörbom, D. (1999). *LISREL 8 : Structural equation modeling with the SIMPLIS command language* (4 ed.). Hillsdale, NJ: Scientific Software International, Erlbaum.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling.* New York, NY: The Guilford Press.

McGraw, R., & Lord, J. A. (1997). Clinical activities during a clerkship rotation in emergency medicine. *Journal of Emergency Medicine, 15*(4), 557-562.

McLeod, P. J., & Snell, L. (1991). Casemix in internal medicine clerkship: Educational value of the clinical problems seen. *Journal of General Internal Medicine, 6*, 455-459.

McManus, I. C., Richards, P., Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ, 316,* 345-350.

Metz, J. C. M., Verbeek-Weel, A. M. M., & Huisjes, H. J. (2001). *Blueprint 2001: training of doctors in the Netherlands. Appendix: list of skills, list of clinical pictures.* Nijmegen: Mediagroep Nijmegen.

Neufeld, V. R., & Norman, G. R. (1985). *Assessing Clinical Competence.* New York: Springer-Verlag.

Nutter, D., & Whitcomb, M. (2001). *The AAMC Project on the Clinical Education of Medical Students.* Washington, D.C: Association of American Colleges.

Patricoski, C. T., Shannon, K., & Doyle, G. A. (1998). The accuracy of patient encounter logbooks used by family medicine clerkship students. *Family Medicine, 30*(7), 487-489.

Raghoebar-Krieger, H. M. J., & Bender, W. (1997). A comparison of the Dutch Blueprint standards (theory) with the experiences of students in clerkships in Groningen (practice). *Journal of Cancer Education, 12,* 85-88.

Raghoebar-Krieger, H. M. J., Sleijfer, D. T., Bender, W., Stewart, R. E., & Popping, R. (2001). The reliability of logbook data of medical students: an estimation of interobserver agreement, sensitivity and specificity. *Medical Education, 35*(7), 624-631.

Schwartz, R. W., Donnelly, M. B., Sloan, D. A., & Young, B. (1994). Knowledge gain in a problem-based surgery clerkship. *Academic Medicine, 69,* 148-151.

Snell, L. M., Battles, J. B., Bedford, J. A., & Washington, E. T. (1998). Verifying the curriculum of a family medicine clerkship. *Medical Education, 32,* 370-375.

Van Leeuwen, Y. D., Düsman, H., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R. P. T. M., et al. (1997). Factors Influencing the Growth in Knowledge of Trainees in General Practice. *Advances in Health Sciences Education, 2*(1), 61-70.

Wimmers, P. F., Splinter, T. A. W., Hancock, G. R., & Schmidt, H. G. (in press, 2006). Clinical competence: General ability or case-specific? *Advances in Health Sciences Education.*

Witzke, D. B., Koff, N. A., McGeagh, A. M., & Skinner, P. D. (1990). Developing a computer-based system to document and report students' clinical encounters. *Academic Medicine, 65*(7), 440-441.

Appendix A

Number of students (total n = 152), means, and standard deviations per hospital site

		Hospital sites													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Students *n*		8	7	10	8	11	14	15	8	6	16	16	13	6	14
Quality Supervision	*M*	6.1	6.8	3.9	6.3	6.6	6.3	6.9	7.1	6.6	6.3	5.8	6.3	5.4	6.7
	SD	1.3	1.0	2.0	1.8	1.8	1.7	2.4	1.1	2.2	2.2	1.7	1.4	1.7	2.2
Patients encountered	*M*	82.4	66.1	13.7	41.0	45.2	44.6	36.9	43.3	46.2	69.5	32.6	31.9	41.0	28.6
	SD	23.6	15.6	5.4	18.1	13.6	8.6	11.7	9.2	18.9	20.8	15.0	8.1	11.7	7.0
Diseases encountered	*M*	69.1	46	9.6	33.3	35.9	32.5	31.6	30.4	36.0	47.1	25.4	26.5	31.7	14.6
	SD	22.9	14.4	4.1	16.8	13.2	8.7	11.2	6.5	18.4	16.5	11.3	7.6	5.9	7.7
Different diseases	*M*	25.5	20.6	7.4	17.5	18.1	17.3	16.3	18.1	19.5	21.1	14.1	14.3	13.8	9.9
	SD	3.7	3.8	2.4	6.2	4.5	4.0	4.2	2.2	3.8	5.4	3.8	3.4	2.9	4.7
Practical Examination	*M*	7.6	7.6	7.2	7.5	8.4	8.0	8.1	7.5	7.7	7.8	7.4	7.4	7.2	7.5
	SD	.7	.5	1.1	.9	.5	.7	.6	.5	.8	.8	.7	.5	1.0	.7
Theoretical Examination	*M*	7.6	7.6	7.2	7.4	7.6	7.6	7.7	7.6	7.2	7.5	7.7	7.6	7.0	7.1
	SD	.7	1.0	.6	1.0	.9	1.0	.9	.5	.8	.7	.6	.5	.6	.8
Professional performance	*M*	7.7	7.9	7.7	7.7	8.0	7.9	7.9	7.9	7.7	7.7	7.8	7.9	7.9	7.8
	SD	.5	.2	.3	.3	.1	.2	.2	.1	.2	.4	.4	.1	.2	.1

Appendix B

Means, standard deviations, and intercorrelations of all observed variables

Observed variables	*Mean*	*SD*	1	2	3	4	5	6	7	8
1. Patient length of stay	10.6	1.2	--	.41**	.07	.37**	.29**	.16*	-.04	-.06
2. Number of patients admissions	1850.5	679.1		--	.11	.34**	.23**	.15	.14	-.01
3. Quality of supervision	6.2	1.9			--	.27**	.25**	.42*	.15	.28**
4. Number of patients encountered	43.3	21.5				--	.89**	.24*	.10	.05
5. Number of diseases encountered	32.4	17.8					--	.21*	.11	.05
6. Practical examination	7.7	.8						--	.48**	.48**
7. Theoretical examination	7.5	.8							--	.24**
8. Professional performance	7.8	.3								--

* $p < .05$, ** $p < .01$, two-tailed

CHAPTER FOUR

Clinical Competence: General Ability or Case-specific?

This study is conducted at and funded by Erasmus MC—University Medical Center Rotterdam (EUR).
An earlier version of this paper is presented at the 85th Annual Meeting of the American Educational Research Association (AERA) in San Diego, California, USA, April 12-16, 2004, with the title: How content specific is clinical competence?
It received the Best Paper by an Established Investigator Award from AERA's Division I Awards Committee in 2005.

Reference
Wimmers, P. F., Splinter, T. A. W., Hancock, G. R., & Schmidt, H. G. (in press, 2006). Clinical competence: General ability or case-specific? Reprinted from *Advances in Health Sciences Education* © 2006 with kind permission of Springer Science and Business Media.

Abstract

BACKGROUND Before the 1970s, research into the development of clinical competence was mainly focused on general problem-solving abilities. The scope of research changed when Elstein and colleagues discovered that individual ability to solve clinical problems varies considerably across cases. It was concluded that problem-solving abilities are highly dependent on domain-specific knowledge rather than on general problem-solving skills. Elstein called this phenomenon "case specificity."

PURPOSE The finding of content specificity will be contrasted with the existence of a general clinical problem-solving ability, and the relationship between preclinical knowledge and a problem-solving ability will be investigated.

METHODS A correlation matrix was calculated with clerkship final scores from 10 disciplines to examine the magnitude of the interrelations. A confirmatory factor analysis was applied to the corresponding covariances using structural equation modeling to investigate whether scores on finals shared any common variance across clerkships. Finally, two additional models were tested to examine the nature of the relationship between preclinical knowledge and problem solving.

RESULTS Low to moderate correlations across clerkship disciplines were found, supporting the original findings of content specificity. Further investigation showed that in addition to specific knowledge, a general, content-independent ability is needed to perform on these examinations.

CONCLUSIONS Clinical competence, as measured in this study, is based on a combination of specific preclinical knowledge and a problem-solving ability. Case specificity fits perfectly well in this interactional perspective on clinical problem-solving but does not explain it. The phenomenon "case specificity" is therefore not solely a result of content knowledge, but of level of experience and level of case difficulty.

MOST DAILY ACTIVITIES of physicians are concerned with solving clinical problems and managing the consequences of disease. Therefore, a reasonable educational approach would be to emphasize teaching the fundamental problem-solving skills necessary to solve these problems. Based on this assumption, medical education in the sixties began to emphasize the clinical problem-solving process, sometimes at the expense of teaching medical knowledge.

This shift from teaching knowledge to teaching general problem-solving skills was elicited in part by research in psychology and artificial intelligence. Newell and Simon, for instance, attempted to design computer programs that could solve a wide variety of problems using general domain-independent problem-solving rules (Newell & Simon, 1956, 1972). Pioneers of research into medical problem solving, following the lead of the influential Newell and Simon work, looked for general medical problem-solving skills by identifying the number and sequence of questions while solving a medical problem (Rimoldi, 1955, 1961), similarities in data-gathering behavior (Donnelly, Gallagher, Hess, & Hogan, 1974; Juul, Noe, & Nerenberg, 1979), or the structure of the pathways throughout the diagnostic reasoning process (McGuire, 1976; McGuire & Babbott, 1967).

Elstein, Shulman and Sprafka (1978) initially discovered that the medical problem-solving process was characterized by generating multiple hypotheses early in the patient encounter. The number of hypotheses generated and the amount of information collected by doctors were essentially the same for all specialties, which suggested the existence of a general problem-solving ability that can be applied to different clinical contents (see also, Barrows, Norman, Neufeld, & Feightner, 1982; Barrows & Pickell, 1991; or, Gale, 1982). However, further research revealed that generating hypotheses early in the diagnostic process turned out to be an inaccurate indicator of level of expertise, because the process was shown to be the same for novices and experts. The differences between novices and experts seemed to be, rather, in the quality of the generated hypotheses and in the practical experience of an expert (Elstein, 1972; Elstein & Schwartz, 2002; Neufeld, Norman, Barrows, & Feightner, 1981). Elstein and colleagues also discovered that the performance of physicians was not consistent across problems. Diagnostic performance on one clinical case did not predict the performance on another case as indicated by low correlations between performance on different cases. This phenomenon was called *case-specificity* (Elstein et al., 1978) or *content-specificity* (McGuire, 1976) of medical expertise, and it emerged in many studies using different kind of assessment methods, such as patient management problems (Neufeld & Norman, 1985), written tests (De Graaf, Post, & Drop, 1987), oral tests (Swanson, 1987), performance-based tests (Van der Vleuten & Swanson, 1990), standardized-patient tests (De Champlain, Macmillan, King, Klass, & Margolis, 1999), and computer-based clinical performance assessments (Fitzgerald et al., 1995).

Obviously, the ability to solve clinical problems did not seem to be a general, across the board, characteristic. Problem-solving skills could not be separated from the contents of the knowledge needed to diagnose a case, and the solving of medical problems required sufficient understanding of specific diseases and how they manifest themselves in the human body. In fact, most current theories consider expertise as a process of extension of causal knowledge about a domain and emphasize the importance of knowledge structures and its organization in memory (e.g., Chi & Glaser, 1985; Chi, Glaser, & Rees, 1982; Patel & Groen, 1986; Schmidt & Boshuizen, 1993; Schmidt, Norman, & Boshuizen, 1990). Twenty-five years later, the content-specificity perspective almost entirely dominates research on expertise and assessment methodology. Some researchers in medical education even assume that the search for general problem-solving skills is (or at least should be) completely abandoned (Eva, 2003; Norman, 2005). However, controversy about the relationship between knowledge and general problem-solving skills still remains (Eva, Neville, & Norman, 1998; McAuliffe, 1978; Norman, Tugwell, Feighter, Muzzin, & Jacoby, 1985; Peverly, 1991).

The study presented here is an attempt to reopen the discussion about this old research dilemma. Since the discovery of content specificity in the seventies, new statistical methods have become available, enabling us to provide the general problem-solving ability/content specificity divide with a new perspective.

In the present study, scores on clerkship finals of 10 different medical disciplines were used to examine the relative contributions of content specificity and general clinical problem-solving ability. Specifically, three models relating scores on clerkship finals were considered and compared: (1) a completely restricted general-factor model, assuming that there is only a general ability underlying performance on clerkship finals; (2) an independence model, assuming that, in fact, no relationships exists among the specific clerkship examinations; and (3) a combined general/specific factor model, assuming that both a general underlying factor and content-specific factors influence performance.

These three models, as displayed in Figure 1, were analyzed and compared using structural equation modeling (SEM) techniques. Once the best model for the data was determined, a secondary analysis examined the relationship of medical knowledge, acquired during the first years of medical training, with the clerkship finals in the context of the model selected.

Materials and Methods

The clinical data used in this study were collected from fifth-year and sixth-year medical students' clerkship rotation finals during the years 1999-2000 and 2000-2001 at Erasmus MC, University Medical Center Rotterdam. Dutch medical curricula consist of a six-year program. The first four years are mainly theoretical in nature and focus on the acquisition of biomedical knowledge. The last two years are reserved for clinical instruction. The cohort studied consisted of 227 students who successfully completed their first four preclinical years at medical school and rotated through all clerkships. The students passed through 10 clerkships in a fixed sequence, starting with internal medicine and ending with family medicine (internal medicine, pediatrics, neurology, psychiatry, surgery, gynecology, dermatology, otorhinolaryngology [ENT], ophthalmology, and family medicine, having a duration of 12, 6, 5, 6, 12, 6, 3, 3, 3, and 4 weeks, respectively). Each clerkship ended with an independent final examination administrated orally. Most of these examinations consisted of the completion of single patient file (i.e., history, physical examination, and diagnosis). The clinical examiner provides the student with an overall grade expressed on a 10-point scale (the grades as used in this study). Despite the often disputed reliability of an oral examination, it is suggested that the strength of oral examinations as a measure of clinical competence is related to its face validity as a measurement technique. It covers three components of competence: knowledge, problem-solving ability, and personal characteristics (Muzzin & Hart, 1985), (note that the estimated reliability across 10 orals in this study was fairly strong, alpha is .82). Students' level of preclinical medical knowledge was measured by the average of all first attempt course grades of the first four preclinical years at medical school (scores for 31 courses expressed on a 10-point scale).

Analysis

Three models relating scores on clerkship finals, shown in Figure 1, were analyzed and compared using structural equation modeling (SEM) techniques: (1) a completely restricted general factor model indicating that a general (problem-solving) aptitude must be underlying the data; (2) an independence model, hypothesizing that the scores on the various clerkship finals had nothing to do with each other and were each produced by a (knowledge) factor specific to the domain-at-hand; and (3) a combined general/specific factor model. If a general clinical problem-solving ability completely explains all variance of, and covariance among, scores, then a completely restricted general-factor model will be appropriate. At the other extreme, if clinical performance across clerkship finals is specific and entirely depends on the content of that particular clerkship, then zero correlations between the various clerkship examinations should exist in the population, and an independence model should be appropriate for these data. (This model cannot technically be analyzed as it creates a model-implied covariance matrix with perfect multicollinearity; however, an approximation to this model will be analyzed, as described below.) Finally, if relationships among variables are explained by a single underlying general factor, but variance in variables remains unexplained by that factor, then a hybrid model is warranted in which both a single general factor and multiple content-specific factors are hypothesized.

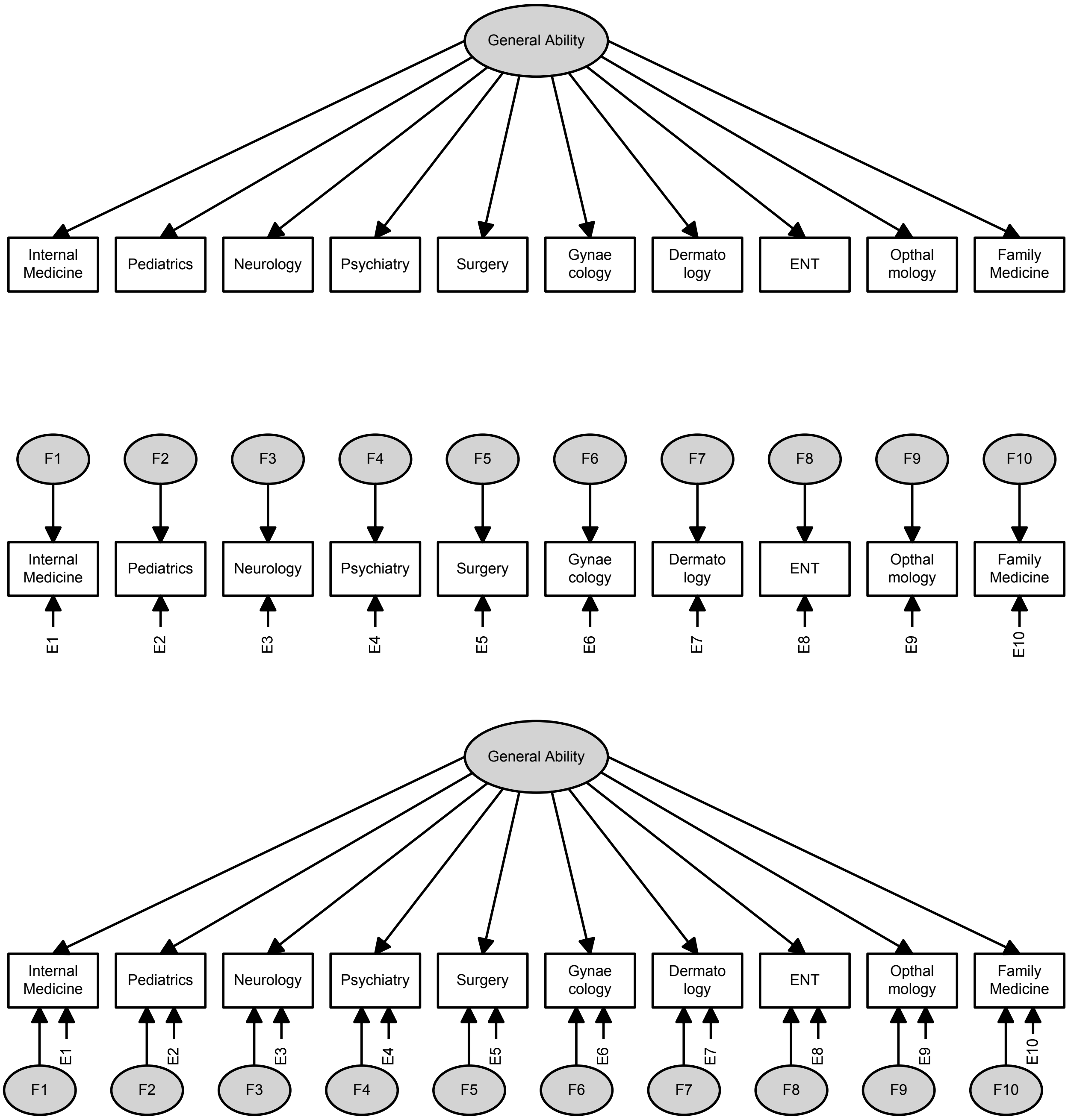


Figure 1 a/b/c. *Conceptualized models: (1) completely restricted general factor model, (2) independence model, and (3) combined general/specific factor model*

As a second modeling step in this study, we determined the influence of preclinical knowledge on clerkship final performance. This was done after having selected the best one of the three above-mentioned models for understanding clerkship final performance.

To conduct the SEM analyses the software package EQS 6.1 (Bentler, 2003) was used, through which maximum likelihood estimation was employed and the expectation maximization (EM) algorithm was used to address minor issues of missing data. When evaluating a hypothesized model, the SEM software package produces several data-model fit indices indicating the degree of consistency between the relationships in the observed data and the relationships implied by the given model. The indices selected for use in this study fall into three classes—absolute, parsimonious, and incremental— and follow the recommendations of Hu and Bentler (1999). A representative of the first class is the model χ^2 statistic, whereby a nonsignificant value indicates that the observed and model-implied covariance matrices are not statistically significantly different, indicating fit of the model to the data. Also absolute in nature is the Standardized Root Mean-Square Residual (SRMR), which represents the average standardized discrepancy between observed and model-implied relations; a value below .08 indicates reasonable fit. Parsimonious fit indices, which adjust for a model's complexity, include Steiger and Lind's Root Mean Square Error of Approximation (RMSEA), which should have a value below .06 to indicate reasonable fit. Another parsimonious index used is Akaike's Information Criterion (AIC), which addresses the issue of parsimony in the assessment of data-model fit across competing models. Although no overall numerical level of acceptability is assigned to this index, its comparative nature is such that smaller values represent better relative fit. Finally, an incremental fit index compares the fit of a given model to one positing complete independence of the variables (the independence model). Bentler's Comparative Fit Index (CFI) was used for this purpose, where a value above .95 indicates acceptable data-model fit. As each of the three classes of fit index reflects a somewhat different aspect of data-model fit, fit indices from each class were used to assess each model's overall acceptability. Models with a hierarchical relationship, that is, where one model is a special (constrained) case of another, were also compared statistically using a χ^2-difference test (Byrne, 2001; Hancock & Mueller, 2006; Kline, 1998).

Results

Descriptives

The intercorrelations among all observed variables, including means and standard deviations, are shown in Appendix A. The correlations among the finals of the 10 clerkship rotations varied from .18 to .46 ($p < .01$), with a mean value of .31. If we were to assess the degree of internal consistency among these 10 clerkship finals, as if they were items on a single scale, the reliability coefficient (Cronbach's alpha) for such a scale would be .82, indicating fairly strong internal consistency.

In a review of clinical assessments, Norman et al., (1985) showed that a variety of studies found correlations with values in the range of .10 to .30 between different problems (see also, Van der Vleuten & Swanson, 1990). If our investigation would stop here (as most of the earlier studies of the content-specificity hypothesis do), we would conclude that the intercorrelations in the present study were a little higher than expected, but still relatively low and, as a result, offering support for the content-specificity assumption.

With regard to the preclinical knowledge variable, which itself had an alpha of .94 across its 31 examinations, notice that its correlations with the clerkship finals ranged from .07 to .33 (averaging .19). This indicates that the relationship between the preclinical knowledge variable and each of the clerkship finals is low to moderate.

The Underlying model

The first step in the modeling portion of this study was to investigate the relative influence of content-specific knowledge and general problem-solving ability on performance. We therefore compared the three models described previously: a completely restricted general factor model, an independence model, and a combined general/specific factor model.

The completely restricted general factor model (shown in Figure 1a) posited a single underlying factor as completely explaining the covariance among the variables as well as all of their variance. As mentioned previously, this model represents perfect multicollinearity and hence cannot be analyzed as described. Instead, a model was analyzed in which a single general factor was assumed to explain 90% of each variable's variance, and thus variables would intercorrelate approximately .90 as well. Data-model fit results indicated: χ^2 (45) = 8023.64, p < .001; SRMR = .24; RMSEA = .89 (90% confidence interval from .87 to .90); AIC = 7933.64; and CFI = .000. Thus, this model fits exceedingly poorly.

The independence model, as shown in Figure 1b, on the other hand, is the model in which variables are assumed to be uncorrelated, meaning that only content specificity, as indicated by the factors F1-F10, is needed to perform on that particular clerkship. This model fit poorly by all assessments: χ^2 (45) = 509.31, p < .001; SRMR = .29; RMSEA = .21 (with 90% confidence interval from .20 to .23); AIC = 419.31; and CFI = .000 (by definition, for the independence model). In short, there is clearly nonrandom dependence among the clerkship variables, to be investigated in the remaining models.

Taken together, the performances on clerkship finals are neither fully independent, as would be expected if performance on each final was purely dependent on specific knowledge of that clerkship, nor could a single underlying factor completely explain performances on all clerkship finals, as would be expected if a general ability was responsible for performances on clerkship finals. Therefore, a combined model was examined in which a general factor explained some of the clerkship finals' interrelations, and specific factors (representing content specificity) were allowed as well to explain unique variance in the finals. This model is depicted in Figure 1c. When analyzing such a model, the software cannot distinguish between the specific factors and remaining error variability; hence, as is customary, these two terms were collapsed into a single residual term. The results were as follows: χ^2 (35) = 35.98, p = .42; SRMR = .037; RMSEA = .011 (with 90% confidence interval from .000 to .049); AIC = -34.02; and CFI = 1.00. Thus, this combined model fits the data extremely well. As to be expected, the combined general/specific model has statistically significantly better data-model fit than the independence model: $\Delta\chi^2 = 473.33$, $\Delta df = 10$, $p < .001$. The same applies to the comparison with the restricted dependence model: $\Delta\chi^2 = 7987.66$, $\Delta df = 10$, $p < .001$. The fact that the combined model χ^2 value was not statistically significant implies that the model is statistically indistinguishable from a model that perfectly explains the data.

The model of choice therefore, is the model combining a general factor with content-specific factors. It meets all criteria for excellence of data-model fit, and is preferred over the alternatives on the basis of explanatory ability and parsimony. The standardized path coefficients in this model from the general factor to the clerkship finals ranged from .43 to .64 ($p < .05$), indicating that the percentage of variance in clerkship finals remaining to be explained by the specific factors and remaining measurement error ranges from 61% to 81%. The variance accounted for by the general factor is 38%. Based solely on the intercorrelations (Appendix A) one could conclude that the average explained variance between clerkship finals was les than 10%.

The role of preclinical subject-matter knowledge

From the observed correlations in Appendix A, it can be deduced that the preclinical knowledge variable (the average of all 31 first attempt course grades of the preclinical years of medical school) obviously has some type of relationship with performance on the clerkship finals. Thus, having selected the best model for understanding clerkship examination performance, the second modeling step in this study was to determine the role of preclinical knowledge within this model. We examined two hypothesized models, as shown in Figure 2. The first hypothesized that preclinical knowledge impacts on general problem-solving ability directly, and, hence, clerkship final performance only indirectly. In this view, the acquisition of knowledge contributes to the development of a general problem-solving skill but does not influence performance directly. The second model considered preclinical knowledge as independent of the general problem-solving skill and impacting on clerkship performance directly, which is analogous to impacting the specific factors. For each model, the preclinical knowledge variable was added to the general/specific factor model as described.

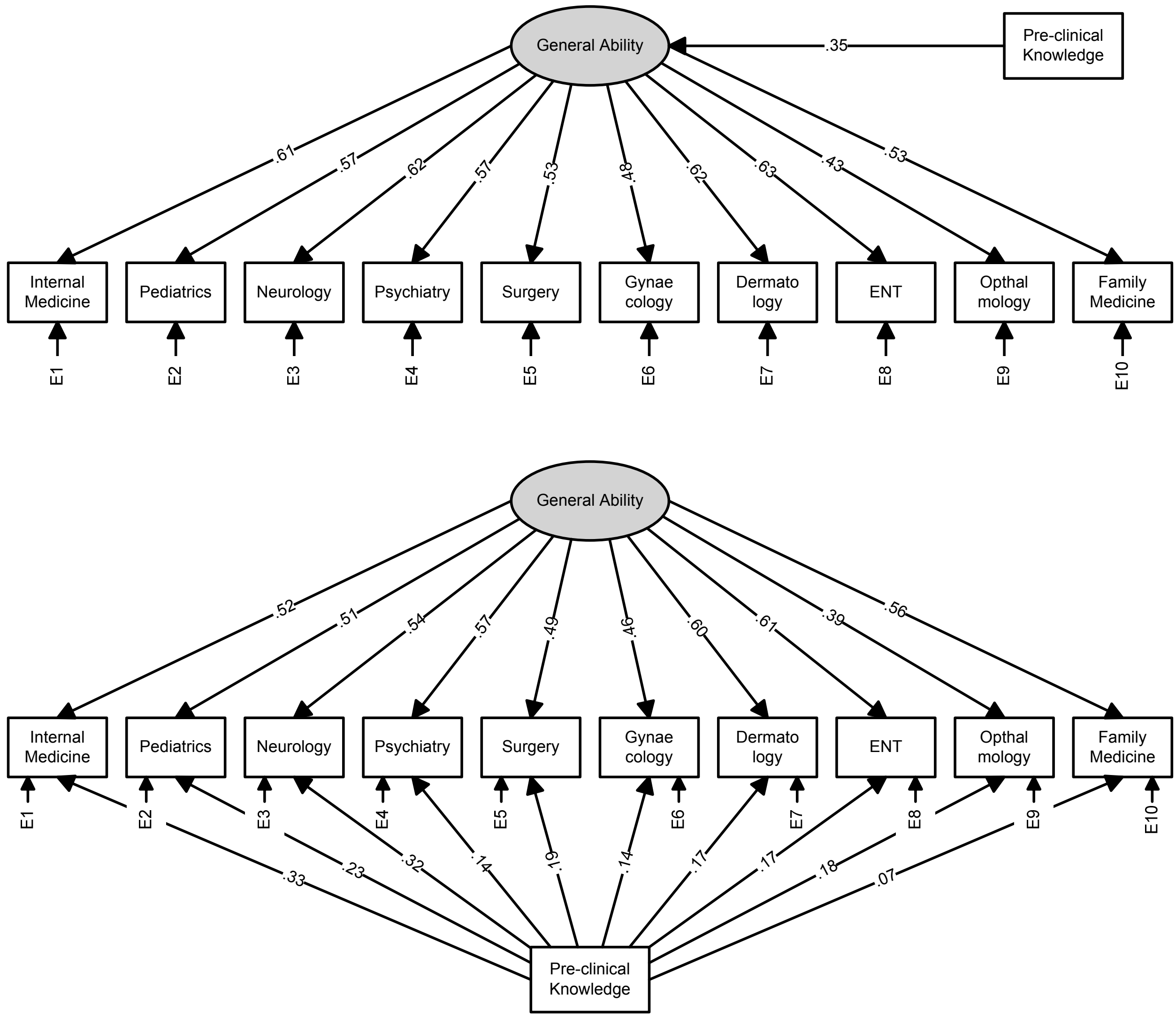


Figure 2 a/b. *One-factor model with preclinical knowledge influencing the factor and one-factor model with preclinical knowledge directly influencing clerkship final scores*

For the first model, in which preclinical knowledge influences general problem-solving ability directly and clerkship final performance only indirectly, the data-model fit was good: χ^2 (44) = 54.67, p = .11; SRMR = .043; RMSEA = .035 (with 90% confidence interval from .00 to .06); AIC = -33.33; and CFI = .98. For the second model, in which preclinical knowledge influences clerkship final performance directly, the data-model fit was also very good: χ^2 (35) = 35.78, p = .38; SRMR = .033; RMSEA = .015 (with 90% confidence interval from .00 to .05); AIC = -34.22; and CFI = 1.00. Because the two models do not have a hierarchical relationship, they cannot be compared using a χ^2-difference test. In such case, the smaller AIC is used to determine the superior model. This is the second model, in which preclinical knowledge influences clerkship final performance directly. It seems that general problem-solving ability and subject-matter knowledge contribute independently to clerkship finals performance. However, both seemed to be needed for clerkship performance. The path coefficients between preclinical knowledge and the clerkship finals are in the range of .07 to .33. Thus, preclinical knowledge plays an inconsistent role with regard to performance in each of the clerkship finals. In fact, the influence of preclinical knowledge seems to decrease over time, as the values for the later clerkships seem to be getting smaller. And this decreasing trend is not related to the length of the clerkships.

Discussion

Research into the development of expertise before the seventies was mainly focused on the search for general problem-solving abilities (Ernst & Newell, 1969; Newell & Simon, 1972). This focus of research changed when Elstein and colleagues discovered that diagnostic performance on one clinical case had no predictive value for the performance on another case. It was concluded that problem-solving abilities are highly dependent on specific content knowledge of the particular problem at hand rather than on a general skill (Elstein et al., 1978; Patel & Groen, 1986).

The present study revisited the old research dilemma of whether clinical competence could be considered a general, content-independent ability or whether competence in this area is dependent on case-specific knowledge. For this purpose, we analyzed individual examination scores of students on 10 different clerkships in two separate modeling steps using structural equation modeling (SEM) techniques.

In the first modeling step of this study, three models underlying the scores on the clerkship finals were considered. The completely restricted general factor model reflected that only a general ability in clinical problem-solving explains clinical competence. This first tested model was clearly not supported by our data. This finding seems at variance with the opinions of those who believe that clinical competence essentially is a domain-independent skill, but is in agreement with earlier failures to model problem-solving as a purely content-independent skill (Elstein et al., 1978; Newell & Simon, 1972). The second model tested was designed to reflect a pure content-specificity perspective. This model failed to find sufficient support in our data as well. Since neither a pure content-specificity assumption nor a pure general-ability assumption could explain the data sufficiently, a third model was tested that combined both assumptions. This model appeared to explain performance of clerkship finals quite well. The combined model shows us that a single underlying factor (with a total explained variance of 38%) and unexplained specific variances relates to clinical problem solving (i.e., clerkship examination performance), and in addition, it shows us that it is possible to separate a general process from its content aspects statistically.

The findings reported in this article are, to some extent, at variance, not only with the content-specificity perspective dominant in the medical education literature, but also with cognitive psychology literature documenting the pervasive role of knowledge in problem-solving (Chi, 1985; Chi & Glaser, 1985; Schmidt et al., 1990). This is not to say that over the past 30 years there were no dissenting voices (Eva et al., 1998; McAuliffe, 1978; Norman, Tugwell et al., 1985; Peverly, 1991). In the late seventies, for example, Berner et al., (1977) concluded that it is erroneous to assume that only knowledge accounts for differences in problem-solving performance. More recently, Eva and colleagues (1998) supported this point of view in a review of the literature about the etiology of content specificity. Peverly (1991) stated that much of the research supporting this content-specificity perspective is methodologically problematic. Depending on the nature of the problems used in this kind of research, findings may lead to underestimating or overestimating of the importance of knowledge in performance. If, for example, problems are too easy, then they may be little more than recall tasks for experts, and one would guess that problem-solving processes would not be generated (see for example, Schuwirth, Verheggen, van der Vleuten, Boshuizen, & Dinant, 2001). General processes may be more largely implicated when an individual is confronted with difficult problems as found in other fields than medicine (Glaser, 1984). There is evidence in medicine that case processing of experts changed only under the demands of the task, like an unusual case format or being forced to pay attention to the information units of the case (Norman, Brooks, & Allen, 1989; Wimmers, Schmidt, Verkoeijen, & van de Wiel, 2005). More evidence for general processes in problem-solving was provided by Berner, et al., (1977). They were able to distinguish three general elements of problems solving: (1) Creating initial problem lists; (2) ordering diagnostic procedures; and (3) the ability to arrive at a final diagnosis. In their study, they demonstrated more similarity in performance on the same elements of problem solving across five clinical cases than on different processes within a given case. Donnelly et al., (1974) were able to demonstrate consistency in information-gathering behavior (i.e., history, physical, laboratory, and imaging) across 10 patient management problems. Juul et al., (1979) found, across 24 patient management problems, in addition to a data-gathering factor (i.e., history, physical, and diagnostic procedures), a decision-making factor (i.e., diagnostic procedures, pathway, treatment, and diagnosis). A recent study of Wimmers and Kanter (2006) using computer-based clinical cases confirmed similar findings: high consistencies in data-gathering behavior among analogous stages of the diagnostic process across cases. See for comparable findings the pioneering studies of Rimoldi in the late fifties and early sixties where a card-drawing technique was applied during diagnosing a case in order to capture problem solving skills (Rimoldi, 1955, 1961). The solely knowledge-based perspective on problem solving as represented by the term “content-specificity“ seems therefore too much a swing in the opposite direction from a general problem-solving skill perspective (McAuliffe, 1978; Peverly, 1991).

In an attempt to understand our findings in more depth, a second modeling step was taken to determine the nature of the relationship between preclinical knowledge and general ability. Comparing two well-fitting models, it turned out that the model relating the preclinical knowledge indicator directly to each of the clerkships was superior. This particular model considered preclinical knowledge as independent of general problem-solving ability. And although both specific subject-matter preclinical knowledge and a general ability influenced performance independently, both seemed to be needed for clerkship performance. This indicates that a combination of specific preclinical knowledge and a general ability is required for clinical problem solving. Alexander and Judy (1988) put forward that competent learners weigh their content knowledge against the demands of the task and then bring in their appropriate form of strategic knowledge to perform on that task. It seems logical to conclude that strategies can be executed only in relation to domain-specific knowledge. However, as mentioned before, when domain-specific knowledge develops, domain-specific problem solving will become easier, and therefore less dependent on general strategies or abilities. The finding of case specificity fits perfectly well in this interactional perspective on clinical problem solving but does not explain it. Level of performance is not entirely dependent on content knowledge, but is highly dependent on level of experience and level of case difficulty. A general problem-solving ability will therefore also be dependent on level of experience and level of case difficulty.

Some limitations should be considered while interpreting the results of this study. The clinical reasoning data used were primarily based on oral end-of-clerkship examinations. We assumed that the general ability is related to student performance on clerkship finals. However, as mentioned before, it is suggested that oral examinations measure a combination of knowledge, problem-solving ability, and personal characteristics. In the latter models studied, we were able to statistically separate knowledge from a general ability, thereby ignoring the personal characteristics. Some part of the single factor found in our study could therefore be linked to personal characteristics of the examinee unrelated to problem-solving behavior but consistent across examinations. For instance, examinee's professional or personal appearance, motivation, communication skills, or showing of respect to the examiner may have been playing a role. The face-to-face interaction between examiner and examinee during orals requires also the examinee's emotional, personal, and social abilities. Consistent bias caused by the examiner is highly unlikely, since all clerkship sites and disciplines have different examiners (an excellent overview of oral examinations is given by Muzzin & Hart, 1985).

In comparing the models that include the preclinical knowledge variable (Figure 2), our choice, for reasons stated in the results section, is made for the later model, whereby preclinical knowledge influences clerkship performance directly rather than through the general factor. The fit indices were, however, very similar and under these conditions one could also choose the first model. In addition, the inconsistent role that preclinical knowledge seems to play with regard to performance in each of the clerkship finals seems to decrease over time as the values for the later clerkships seem to be getting smaller. The influence of the passage of time itself on level of preclinical knowledge might be considered. This observed trend could indicate that in time students rely more on clinical knowledge and less on basic science knowledge, a well studied phenomenon in the medical education literature (Schmidt & Boshuizen, 1993; Schmidt et al., 1990).

Conclusions

The data presented in this article at first sight seemed to support the content-specificity perspective: the relationship between performances on clerkship examinations tended to be relatively low, suggesting low covariability between cases. Nevertheless, further analyses of these data suggested that in addition to examination-specific knowledge or skill, a general, content-independent ability is needed to perform on these examinations. Domain-general abilities can be statistically distinguished from specific knowledge. In the light of our results, the words of Simon (1980) that "powerful general methods do exist and can be taught in such a way that they can be used in new domains" seem to get a different meaning again. However, it is clear that a general ability has a strong knowledge requirement: one cannot be developed without the other. The data discussed in this study seem to support the notion that specific knowledge and general abilities are both necessary conditions for clinical problem solving.

In order to get more insight into the quality of this general ability, it will be a challenge for contemporary research methodology and measurement techniques to *re*-search what aspects are analogous in problem solving. And maybe, for this purpose, we have to put aside the 'reliable' multiple short clinical cases and dig out again the old more complicated, realistic patient cases. The research can then be redirected towards the critical analysis of the interaction between domain-specific knowledge and domain-general strategies or abilities. The models tested in our study on a homogeneous group of participants should be applied to participants of different levels of expertise. In addition, it is important to provide participants with problems from different levels of difficulty, because it is reasonable to assume that if content knowledge is sufficient to solve a problem, the need for applying general strategies decreases.

References

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning process of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine, 5*, 49-56.

Barrows, H. S., & Pickell, G. C. (1991). *Developing clinical problem-solving skills: a guide to more effective diagnosis and treatment* (First ed.). New York, London: Norton Medical Books: W. W. Norton & Company, Inc.

Bentler, P. M. (2003). EQS (Version 6.1 Beta) [Computer Software]. Encino, CA: Multivariate Software, Inc.

Berner, E. S., Bligh, T. J., & Guerin, R. O. (1977). An indication for a process dimension in medical problem solving. *Medical Education, 11*, 324-328.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming.* MahWah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (1985). Changing conception of sources of memory development. *Human Development, 28*, 50-56.

Chi, M. T. H., & Glaser, R. (1985). Problem solving ability. In R. Sternberg (Ed.), *Human abilities: An information-processing approach* (pp. 227-257). San Francisco: W. H. Freeman & Co.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-76). Hillsdale, NJ: Erlbaum.

De Champlain, A. F., Macmillan, M. K., King, A. N., Klass, D. J., & Margolis, M. J. (1999). Assessing the impacts of intra-site and inter-site checklist recording discrepancies on the reliability of scores obtained in a nationally administrated standardized patient examination. *Academic Medicine, 74*(10 Suppl), S52-S54.

De Graaf, E., Post, G. J., & Drop, M. J. (1987). Validation of a new measure of clinical problem-solving. *Medical Education(21)*, 213-218.

Donnelly, M. B., Gallagher, R. E., Hess, J. W., & Hogan, M. J. (1974). *The dimensionality of measures derived from complex clinical simulations.* Paper presented at the Thirteenth Annual Conference on Research in Medical Education, Chicago, IL, US.

Elstein, A. S. (1972). *Methods and theory in the study of medical inquiry.* Cambridge, MA: Harvard University Press.

Elstein, A. S., & Schwartz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of cognitive literature. *British Medical Journal, 324*(729-732).

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning.* London, Cambridge, Mass: Harvard University Press.

Ernst, G., & Newell, A. (1969). *GPS: A case study in generality and problem solving.* New York: Academic Press.

Eva, K. W. (2003). On the generality of specificity. *Medical Education, 37*, 587-588.

Eva, K. W., Neville, A. J., & Norman, G. R. (1998). Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. *Academic Medicine, 73*(10 Suppl), S1-S5.

Fitzgerald, J. T., Wolf, F. M., Davis, W. K., Barclay, M. L., Bozynski, M. E., Chamberlain, K. R., et al., (1995). A preliminary study of the impact of case specificity on computer-based assessment of medical student clinical performance. *Evaluation and the Health Professions, 17*(3), 307-321.

Gale, J. (1982). Some cognitive components of the diagnostic thinking process. *British Journal of Educational Psychology, 52,* 64-76.

Glaser, R. (1984). Education and Thinking: The role of knowledge. *American Psychologist, 39*, 93-104.

Hancock, G. R., & Mueller, R. O. (2006). *Structural Equation Modeling: A Second Course (Quantitative Methods in Education and the Behavioral Sciences)*. Greenwich, Connecticut: IAP Information Age Publishing.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-31.

Juul, D. H., Noe, M. J., & Nerenberg, R. L. (1979). A factor analytic study of branching patient management problems. *Medical Education, 13*(3), 199-205.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling.* New York, NY: The Guilford Press.

McAuliffe, W. E. (1978). Studies of process-outcome correlations in medical care evaluations. *Medical Care, 16*(11), 907-930.

McGuire, C. H. (1976). Simulation technique in the teaching and testing of problem solving skills. *Journal of Research in Science Teaching, 13*(2), 89-100.

McGuire, C. H., & Babbott, D. (1967). Simulation technique in the measurement of problem-solving skills. *Journal of Educational Measurement, 4,* 1-10.

Muzzin, L. J., & Hart, L. (1985). Oral examinations. In V. R. Neufeld & G. R. Norman (Eds.), *Assessing clinical competence*. New York: Springer-Verlag.

Neufeld, V. R., & Norman, G. R. (1985). *Assessing Clinical Competence.* New York: Springer-Verlag.

Neufeld, V. R., Norman, G. R., Barrows, H. S., & Feightner, J. W. (1981). Clinical problem-solving by medical students: A longitudinal and cross-sectional analysis. *Medical Education, 15*, 315-322.

Newell, A., & Simon, H. A. (1956). The logic theory machine: A complex information processing system. *Transactions on Information Theory (Institute of Radio Engineers), IT-2*(3), 61-79.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norman, G. R. (2005). Editorial -- checklists vs. ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. *Advances in Health Sciences Education, 10*, 1-3.

Norman, G. R., Bordage, G., & Curry, L. (1985). A review of recent innovations in assessment. In R. E. Wakeford (Ed.), *Directions in Clinical Assessment* (pp. 9-27.). Cambridge: Cambridge University School of Clinical Medicine.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by experts medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology; Learning, Memory and Cognition, 15*, 1166-1174.

Norman, G. R., Tugwell, P., Feighter, J. W., Muzzin, L. J., & Jacoby, L. L. (1985). Knowledge and clinical problem-solving. *Medical Education, 19*, 344-356.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*(1), 91-116.

Peverly, S. T. (1991). Problems with the knowledge-based explanation of memory and development. *Review of Educational Research, 61*(1), 71-93.

Rimoldi, H. J. A. (1955). A technique for the study of problem solving. *Educational and Psychological Measurement, 15*, 450-461.

Rimoldi, H. J. A. (1961). The test of diagnostic skills. *Journal of Medical Education., 36*(1), 73-79.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On Acquiring Expertise in Medicine. *Educational Psychological Review, 5*(3), 205-221.

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611-621.

Schuwirth, L. W. T., Verheggen, M. M., van der Vleuten, C. P. M., Boshuizen, H. P. A., & Dinant, G. J. (2001). Do short cases elicit different thinking processes than factual knowledge questions do? *Medical Education, 35*(4), 348-356.

Simon, H. A. (1980). Problem solving and education. In D. T. Tuma & R. Reif (Eds.), *Problem solving and education: Issues in teaching and research*. Hillsdale, NJ: Erlbaum.

Swanson, D. B. (1987). A measurement framework for performance-based tests. In I. R. Hart & R. M. Harden (Eds.), *Further developments in Assessing clinical competence* (pp. 13-45). Montreal: Can-Heal Publications.

Van der Vleuten, C. P. M., & Swanson, D. B. (1990). Assessment of clinical skills with standarized patients: State of the art. *Teaching and Learning in Medicine, 2*, 58-76.

Wimmers, P. F., & Kanter, S. L. (2006, March 3-4). *Expert-novice differences in data-gathering behavior during diagnoses.* Paper presented at the Northeast Group on Educational Affairs, Philadelphia, Pennsylvania

Wimmers, P. F., Schmidt, H. G., Verkoeijen, P. P. J. L., & van de Wiel, M. W. J. (2005). Inducing expertise effects in clinical case recall. *Medical Education, 39*(9), 949-957.

Appendix A

Means, standard deviations, and intercorrelations of all variables (Students N = 227)

Clerkship	Mean	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. Internal medicine	7.66	.81	--	.45**	.33**	.30**	.30**	.31**	.30**	.37**	.35**	.30**	.33**
2. Pediatrics	7.48	.76		--	.34**	.24**	.28**	.32**	.31**	.31**	.35**	.29**	.23**
3. Neurology	7.56	.96			--	.33**	.38**	.25**	.22**	.44**	.37**	.35**	.32**
4. Psychiatry	7.71	.82				--	.31**	.33**	.22**	.37**	.46**	.35**	.14*
5. Surgery	7.85	.79					--	.24**	.22**	.35**	.28**	.31**	.19**
6. Gynecology	7.86	.62						--	.18**	.33**	.27**	.20**	.14*
7. Ophthalmology	7.64	.69							--	.25**	.30**	.22**	.18**
8. ENT	7.79	.59								--	.37**	.36**	.17**
9. Dermatology	7.77	.65									--	.35**	.17*
10. Family medicine	7.78	.65										--	.07
11. Preclinical knowledge	6.52	.70											--

Note: * Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed). ENT stands for Otorhinolaryngology

CHAPTER FIVE

Clinical Competence through the Eyes of an Educator: Differences in Perceived Importance of Student Performance on the Wards and on Clerkship Examinations

This study is conducted at and funded by Erasmus MC—University Medical Center Rotterdam (EUR).
An earlier version of this paper is presented at the 45th Annual Conference on Research in Medical Education (RIME), Association of American Medical Colleges (AAMC), Seattle, Washington, USA, October 27 - November 1, 2006, with the title: Being a good doctor vs. being a good student: How educators rate the importance of 21 competencies for functioning on the wards vs. assigning a grade.

Reference

Wimmers, P. F., Schmidt, H. G., Kanter, S. L., & Splinter, T. A. W. (in preparation, 2006). Clinical competence through the eyes of an educator: differences in perceived importance of student performance on the wards and on clerkships examinations.

Abstract

BACKGROUND Clinical rotations play an important role in the medical curriculum and are considered crucial for student learning. Examinations at the end of the clerkship intend to test what has been learned. However, competencies that were intended to be learned can differ from those that are assessed.

PURPOSE The aim of this study was to assess which competencies are considered important while students work on the wards and to what extent teachers consider these competencies during examinations.

METHOD A survey that consisted of 21 different competencies was administered to clinical teachers from the various medical specialties. First, two independent factor analyses were conducted to uncover underlying latent relationships among the different competencies using structural equation modeling. Second, we explored if there were significant discrepancies between what is important for performance on the wards and what is considered important during clinical examinations.

RESULTS Different competencies seem to be important for students while being examined, than during daily performance on the wards as indicated by level of ranked importance of the individual components of competence and their underlying factor structure.

CONCLUSIONS In the perception of clinical educators, it seems that what is important for adequate performance is not necessarily in alignment with what is required for examination. Taking into account the effects of assessment on learning, this phenomenon could jeopardize the development of clinical competence.

MEDICAL EDUCATION traditionally makes a distinction between basic medical science education in the first years and clinical practice in the later years. While working in professional practice, students integrate their basic biomedical knowledge of the first few years of medical education with clinical experience. During the specialty-specific clerkship rotations students get the opportunity to take part in the care of patients and observe the physicians and supporting staff in action. In these environments primarily intended for patient care, students build their basic clinical competence needed for high quality patient care.

There is general agreement on the vital importance of practical learning during clinical rotations and the significance of seeing real patients. However, despite the growing body of research studies into clinical learning and assessment, empirical evidence about how and what students actually learn during clerkships is still limited (Jolly, 1994; Van der Hem-Stokroos, Scherpbier, Van der Vleuten, De Vries, & Haarman, 2001). For example, studies that attempted to find relationships between the number or variety of patients seen by students during clerkships and performance at end-of-clerkship examinations were not able to do so (McManus, Richards, Winder, & Sproston, 1998; Van Leeuwen et al., 1997; Wimmers, Splinter, & Schmidt, 2006). This is perhaps not strange, because learning during clerkships is dependent on multiple factors and clinical assessment takes place at different settings by different examiners on students who met different supervisors and experienced different kind of patients.

Describing what competencies should be learned on the wards is a prerequisite for assessment. Early definitions of competence were focused on the tasks of the clinical encounter. Hubbard and colleagues (1965), for example, described a method developed by John Flanagan and the staff of the American Institute for Research in the early 1960s whereby senior physicians and residents who had direct responsibility for the supervision of students were asked to record good and bad medical practice in clinical situations (so called critical incidents). It was interesting to note that all good incidents were related to the diagnostic process, such as: history taken, physical examination, use of laboratory test, patient management, etc. Definitions of clinical competence were more often based on the diagnostic process. The American Board of Internal Medicine defined four different dimensions of clinical competence whereby problem solving was the core aspect: (1) abilities (i.e., knowledge, technical skills, and interpersonal skills), (2) problem solving skills (i.e., data-gathering and diagnoses), (3) the nature of the medical illness (the problems encountered by the physician), and (4) the social and psychological aspects of the patient problem, especially those which relate to diagnosis and management (ABIM, 1979). In time, definitions became more extensive and in line with the fast growing demands of society on health care delivery. In a later report of the same American Board of Internal Medicine, for example, more competencies were added, such as: communication skills and professionalism, that were not directly important for the diagnostic process, (ABIM, 2002).

Many organizations related to the medical profession tried to specify the components of clinical competence. One of the most well known classical divisions is: knowledge, skills, and attitudes (see for a recent use of this model, Rice & Sinclair, 1995). Many later definitions built on this foundation and became more detailed, in line with the growing demands of society on health care delivery. For example, the Association of American Medical Colleges (Anderson et al., 1998) formulated four competencies, the institute of medicine of the national academies (Greiner & Knebel, 2003) formulated five competencies, CanMeds (1996) formulated seven, and the National Board of Medical Examiners (2002) eight.

In medical education, an all-inclusive definition of competencies seems most pressing for assessment purposes. The reliable measurement of all competencies formulated is not easy. George E. Miller (1990), for example, distinguished several hierarchical layers of competence to function as a framework for within which assessment might occur. In his model a distinction is made between knowledge and its application to problem solving, and how it is demonstrated in real situations. Most assessment instruments, even in clerkships are focused on the lower levels of Miller's competence model (knowledge and the application of knowledge), and if they are reflecting the higher levels (performance and action) then their quality is poor (Miller, 1990; Van der Vleuten et al., 2000). Another problem for the quality of current assessment methods is the above mentioned lack of a consistent and all-inclusive definition of clinical competence (Burg, Lloyd, & Templeton, 1982; Newble, 1978; Van der Vleuten et al., 2000).

Taken together, clinical competence seems to be a multifaceted concept and many different definitions are in circulation. Although most definitions are rooted in the diagnostic process, it is not a sufficient descriptor of clinical competence anymore. Nowadays descriptions emphasize that a clinically competent physician is more than a diagnostic problem solver, he/she is a highly qualified and specialized professional able to function in a high demanding society (Stern, 2006). An interesting inquiry is to investigate which components of clinical competence are considered most important for adequate performance and to what extent these competencies are applied to assessment and grading of medical students. Assessment of clinical competence is often based on a few observations or a single examination with (simulated) patient cases at the end of a clerkship. Clinical assessment must match the competencies being learned (Van der Vleuten et al., 2000; Wass, Van der Vleuten, Shartzer, & Jones, 2001).

In the present study, a survey that lists 21 competencies was developed and administrated to clinical educators and physicians of different hospitals and specialties involved with student learning and/or examination during clerkship rotations. The survey was developed based on several descriptions of components of competencies (ABIM, 1979, 2001, 2002; Anderson et al., 1998; CanMEDS, 1996; General-Medical-Council, 2001; Greiner & Knebel, 2003; NBME, 2002). The questions asked for each competency concerned the degree of its importance for student daily performance on the wards and the degree of usefulness of this same aspect of clinical competence for examinations. Methods of examination during clerkships consist mostly of checklists or rating scales to evaluate professional performance of the student as well as practical end-of-clerkship examinations, administered often orally by the students' supervising physician/clinical educator of the site, to examine knowledge and problem-solving ability. It has been stated previously that oral examinations cover three competencies: knowledge, problem-solving ability, and personal characteristics (See for a thorough overview about orals, Muzzin & Hart, 1985). Our main interest was (1) which aspects of clinical competence are considered important for adequate performance of students on the wards and to what degree are the same aspects of clinical competence employed for student examination (in terms of giving grades)? (2) Is there a significant discrepancy between what is important for performance on the wards in comparison with what is important for examinations? For this purpose, factor analyses were conducted, using structural equation modeling, to abstract underlying simple constructs (the components of competence) and to further specify relationships among the latent factors.

Method

Setting

This study was conducted at Erasmus MC—University Medical Center, Rotterdam, the Netherlands and its affiliated hospitals of the same region. Approximately 400 students enter this medical school each year. Medical schools in the Netherlands have a six-year program. The first four years are mainly focused on basic medical science and theory, and in the last two years, students rotate through several clerkships of the main 10 specialties at different hospital sites. The clerks are supervised, guided, and assessed by physicians and clinical educators of that particular site.

Participants

This study was administrated to clinical educators and physicians of different hospitals and specialties who work with medical students during their clerkships. The clinical educators and physicians were solicited by letter along with the developed survey. In order to reach as many clinical educators as possible, we included additional surveys and with the request to distribute copies among colleagues also involved in supervising clerks. Of the 120 letters sent, 218 surveys were returned from 17 different hospitals and 10 specialties over a period of six months.

Instrument

The draft survey was piloted and reviewed by 10 clinical educators. Specifically, we asked if items were clearly formulated, and if any items were missing. Their comments were implemented in the final version, which consisted of 21 items that could be rated on two similar response columns (a total of 21 items and 42 responses). The items consisted of short statements or phrases related to clinical competence (e.g., “skill in interview and physical examination”) and we labeled them “competencies.” In the first column, respondents could rate the level of importance of the respective item for good performance of students on the wards (further identified as “student performance on the wards”), and in the second column, respondents could rate the level of usefulness of the same item for student examination, in terms of giving grades (further identified as “student performance on examinations”). On the survey, the physicians could mark whether they were involved with student performance on the wards and/or student performance on examinations, and, in addition, the physicians could indicate their name and specialty. Answers on both columns were scored on 6-point Likert scales (1 = never, and 6 = very much). The complete survey of clinical competence is printed in Appendix A.

The advantage of this design of the survey was that each column gave unique information about the same item (competency) in a different context (i.e., performance vs. examination). Comparison of the ratings of the items on both columns gives information about the discrepancy between what is important for good performance on the wards and how useful the same item is for student performance at examinations.

Data analysis

First, we looked at differences among specialties using descriptives and ANOVAs. To test whether the items of the survey were internally consistent, we looked at the correlations between the respective item and the total sum score, the squared multiple correlations between the respective item and the other items, and the internal consistency if a respective item was deleted. The Likert response data were treated as continues variables. Then, exploratory factor analyses with rotation (oblique rotation) using SPSS 14.0 were conducted on the data of student performance on the wards and student performance at examinations separately in order to identify covarying variables and to reduce the observed variables into smaller number of factors or components.

To analyze further the underlying structure and to explore the structural relations among the identified factors we used Structural Equation Modeling (SEM) software package EQS 6.1. (Bentler, 2003). Several fit indices were used to determine if the models being tested reflected the observed covariance structure. Fit indices are arranged according to class: absolute, parsimonious, and incremental fit indices. All class of fit indices reflect somewhat different facets of the model fit, therefore, fit indices from each of the classes were used to support model's acceptability. Kline (1998) recommends using four different fit indices. The fit indices used in this study are: The normal chi-square, the chi-square fit index divided by degrees of freedom (χ^2/df), which indicates that the existing model's covariance structure is different from the observed covariance matrix. A value < 3 is recommended. Bentler's Comparative Fit Index (CFI) which compares the existing model fit with an independence model and should have a value of .90 or higher to accept fit of the model (a value of 1 would indicate a perfect fit). The Standardized Root Mean-square Residual (SRMR) represents the average standardized difference between the observed and the model-implied variances and covariances in the model. A value below .08 indicates reasonable fit. Finally, Steiger's Root Mean Square Error of Approximation (RMSEA) penalizes for lack of parsimony. For the latter two indices counts that the closer the value to 0, for a model being tested, the better the model fit. A value below .06 for the RMSEA is recommended for an adequate fit. See for a more specific discussion of cut-off criteria the papers of Hu and Bentler (1999), and March et al., (2004). Multivariate skew or kurtosis for the structural models was analyzed by using Mardia's normalized coefficient (values > 3 are indicative for violations of multivariate normality, Kline, 1998). The chi-square fit index is sensitive to violations of the assumption of multivariate normality. The Satorra-Bentler scaled chi-square which adjusts model chi-square for non-normality of the data can be used as an alternative. The Lagrange Multiplier test and the Wald test are used to assist in model modification. Modifications are made step by step. The Lagrange Multiplier test for adding parameters is designed to evaluate the statistical necessity of one or more restrictions on a model. The Wald Test, on the other hand, is a test on the free parameters. It evaluates whether a free parameter could possibly be zero in the population (Bentler, 2003; Kline, 1998).

Furthermore, it was investigated which components of competence (the factors) were rated most important for student performance on the wards and student performance at examinations, and whether there was a difference between the two ratings. For this purpose the average item score per factor was calculated and compared for each factor within and between both factor models using paired-samples *t*-tests and independent-samples *t*-tests, respectively. In order to make the comparison between factors of both models appropriate, the average of common items of each corresponding factor were used.

Results

Descriptive statistics

A total of 218 surveys were returned. A total of 23 surveys were excluded: three were doubles, three who marked on the survey that they were not involved with students and 14 who did not mark "involved in student performance on the wards" and "student performance on examinations" at all and were unidentifiable. The resulting surveys came from 10 specialties: 22 internal medicine, 29 pediatrics, 13 neurology, 31 psychiatry, 30 surgery, 29 gyneacology, 24 dermatology, 7 ENT, 10 ophthalmology, and 2 psychology. We did not include psychology in further analyses because these were not part of the main clerkship specialties. Useful for analyses were therefore 195 surveys. A total of 186 participants marked "involved with student performance on the wards;" 161 marked "involved with student performance at examinations;" and 156 marked both items. In order to make use of all remaining responders and to avoid list-wise deletion, SPSS's EM method was used to estimate missing data. An average of .36% and .68% were imputed per item for student performance on the wards and student performance at examinations, respectively.

Most items gave small negative values for skewness, what indicated data that are skewed left. The distribution of item "respect for the patient" of student performance on the wards deviated most from normality with values of -1.62 for skewness and 5.56 for kurtosis, meaning that this particular item was skewed left with a leptokurtic shape (an absolute skew index > 3.0 and an absolute kurtosis index > 10.0 are indicated as problematic, Kline, 1998). This particular item was rated as highly important for student performance on the wards by almost all responders. The deviation is not serious given the number of items and the item will not be excluded for further analyses.

We initially analyzed the within and between-specialty effects for all items of student performance on the wards and student performance at examinations, respectively, using repeated measures analysis of variance. The main effect of specialty was not significant for student performance on the wards, $F\ (8, 177) = .941$, $p = .48$, and student performance at examinations, $F\ (8, 152) = .763$, $p = .64$. The specialty by item interaction was significant for student performance on the wards, $F\ (160, 3540) = 1.667$, $p <.0005$, and for student performance at examinations $F\ (160, 3040) = 1.664$, $p <.0005$. The main effect of items was significant for student performance on the wards, $F\ (20, 3540) = 37.497$, $p < .0005$ and for student performance at examinations $F\ (20, 3040) = 54.834$, $p < .0005$. This indicated that there is no significant overall difference among specialties, but various items were scored differently among specialties and the scoring on the items differed among each other.

The uniformity among specialties is noteworthy; it seemed that most physicians have a similar mental model about what competencies were important for student performance on the wards. Almost all specialties gave "respect for the patient" the highest ranking and "dexterity" (followed by "assertiveness") the lowest ranking. It is interesting to note that what lies at the heart of clinical competence "biomedical knowledge" was among the lowest ranked items across specialties for student performance. Ranked importance for student performance at examinations showed different patterns. Among the highest ranked items belonged "knowledge of diseases" and "skill in history and PE" (psychiatry ranked "respect for the patient" highest) and "dexterity" and "assertiveness" received the lowest ranking.

The mean ratings and *SD*s for all competencies of student performance on the wards and student performance at examinations are listed in Appendix A.

Factor analysis

Since the rankings show similar patterns among specialties it was possible to define components of competence with the use of factor analysis using participants of all specialties. To test the fit of the factor structures against the covariance structure and to further specify relationships among the latent variables, structural equation modeling software was used. A stepwise procedure of model modifying is used to come to a final model.

Components of competence for student performance on the wards

The estimated reliability coefficient for all items of student performance on the wards was high: Cronbach's alpha = .90. The Cronbach's alpha did not improve if any of the items would be deleted. A single construct, related to clinical competence, is measured for items related to student ward performance. Factor analysis is used to define further the components of competence. Analysis of the eigenvalues and screeplots lead to the conclusion that a four-factor structure could explain the underlying structure of the data best. The four-factor solution accounted for 55.38% of the total variance. The first factor had an eigenvalue of 6.91 and accounted for 32.93% of the variance. Most factor loadings differentiated well between the factors and load strongly (values of .50 and higher) on the respective factor. Some items indicated low till moderate cross loadings on several factors.

The first structural model tested was an independent four-factor structure based on 18 items (the items, "assertiveness," "presentation skills," and "dexterity" were removed for reasons indicated above). Mardia's normalized estimate was 15.62 indicating that multivariate normality was violated and the Satorra-Bentler scaled χ^2 which adjust for non-normality will be used as an alternative for the normal-theory chi-square statistic. The fit indices of the independent four-factor model indicated poor fit: Satorra-Bentler (S-B) scaled χ^2 (135) = 380.88, $p < .001$, the $\chi^2/df = 2.82$, CFI = .72, SRMR = .242; RMSEA = .099 (90% confidence interval from .087 to .111). The second model tested had the same four-factor structure (i.e., the same items are loading on the same factors) but with no constrains among the factors. This model showed a much better fit: S-B χ^2 (129) = 227.05, $p < .001$, the $\chi^2/df = 1.76$, CFI = .89, SRMR = .066; RMSEA = .064 (90% confidence interval from .050 to .077). The Lagrange Multiplier test for adding parameters specified a further significant improvement (total $\Delta\chi^2 = 36.26$, $\Delta df = 2$, $p < .001$) of the model by adding parameters (cross loadings) between the first factor (named "cognitive abilities") and "conscientiousness," and between the second factor (named "interpersonal skills") and "respect of the patient." The already previous mentioned importance of "respect for the patient" seemed to be reflected in the double loadings on factors "interpersonal skills" and "professional abilities." The model is shown at the left side of Figure 1. The fit indices of this modified four-factor model were S-B χ^2 (127) = 198.248, $p < .001$, the $\chi^2/df = 1.54$, CFI = .92, SRMR = .058; RMSEA = .055 (90% confidence interval from .040 to .069) indicating greater fit. The individual competencies have an underlying simple structure and four factors could be defined. The factors correlated moderately to strongly with each other; the correlations ranged from .43 to .72. The factor "motivation" seemed to play a central role during performance of the students on the wards, since it is highly correlated to all the other factors (correlations ranged from .64 to .72). To be able to make realistic comparisons between the factors we choose to calculate the factor means by averaging the raw scores of the respective item indicators instead of using estimates based on a mean structure model. The average raw item score of factor "motivation" was highest (M = 4.89, SD = .63), followed by the "professional qualities" factor (M = 4.86, SD = .66). The factors "interpersonal skills" and "cognitive abilities" had an average item score of 4.81 (SD = .66) and 4.64 (SD = .65), respectively. Only the factor "cognitive abilities" was significantly different from all other factors (t-values (185) ranged from 3.206 to 5.474, $p < .005$). There were no significant factor differences across specialties.

Components of competence for student performance at examinations

The Cronbach's alpha for all items of student performance at examinations = .86. The reliability of the survey would not improve if any of the items were to be deleted. A single construct, related to clinical competence, underlies student performance at examinations. Factor analysis is used to define further the components of competence. Analysis of the eigenvalues and screeplots lead to the conclusion that for student performance at examinations a four-factor structure could explain the underlying structure of the data best. The four-factor solution accounted for 55.35% of the variance. The first factor with an eigenvalue of 6.10 accounted for 29.06% of the variance. Internal consistency did not improve if any of the items within a factor were deleted. Most factor loadings of these four factors differentiate well between the factors and load strongly (values of .50 and higher) on the respective factor. Two items had moderate cross loadings on other factors.

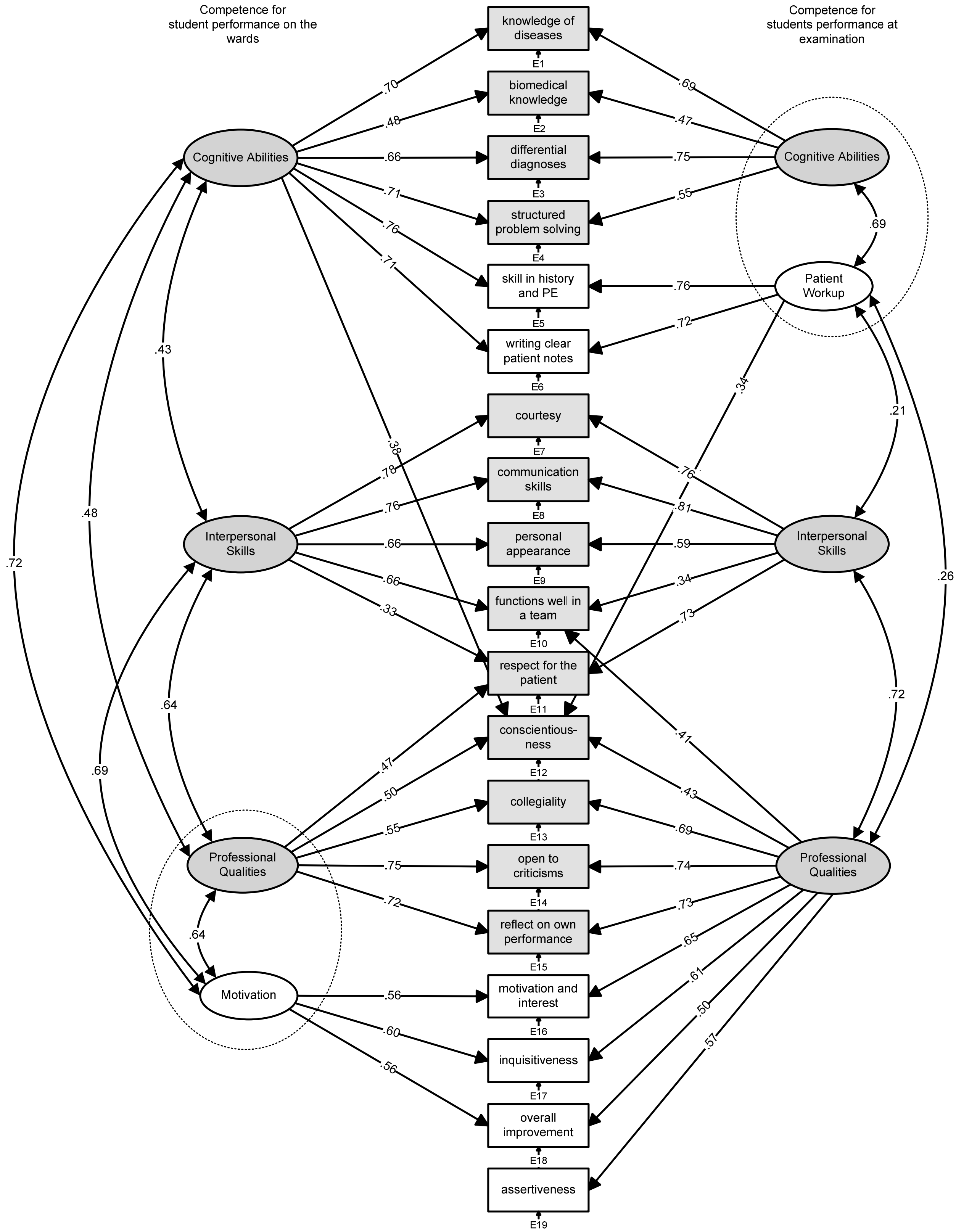


Figure 1. *Conceptualized model of the two four-factor models of the components of competence for student performance on the wards and for student performance at examinations. All results are significant at the .05 level. Note that the two models are analyzed independently and put together for comparison*

The first structural model tested had an independent four-factor structure based on 19 items (the items, “presentation skills,” and “dexterity” were removed on reasons indicated above). Mardia’s normalized estimate was 18.82 indicating that multivariate normality was violated and the Satorra-Bentler scaled χ^2 which adjust for non-normality will be used. The fit indices of the independent four-factor model indicated poor fit: Satorra-Bentler scaled χ^2 (152) = 354.89, $p < .001$, the $\chi^2/df = 2.33$, CFI = .73, SRMR = .199; RMSEA = .091 (90% confidence interval from .079 to .103). The second model tested had all factors correlated with each other. This model showed an increased fit of: χ^2 (146) = 241.16, $p < .001$, the $\chi^2/df = 1.65$, CFI = .88, SRMR = .066; RMSEA = .064 (90% confidence interval from .049 to .078). A borderline fitting model and further modification is needed.

The Wald test for dropping parameters specified to drop the nonsignificant factor correlations between the first factor (named “cognitive abilities”) and the third (named “interpersonal skills”), and between the first and the fourth (named “professional qualities”). The Lagrange Multiplier test for adding parameters specified a further significant improvement (total $\Delta\chi^2 = 17.71$, $\Delta df = 0$, $p < .001$) of the model by adding a parameter (cross loadings) between the second factor (named “patient workup”) and the item “conscientiousness,” and between the fourth factor and item “functions well in a team.” The model is shown at the right side of Figure 1. The fit indices of this modified four-factor model were χ^2 (146) = 223.45, $p < .001$, the $\chi^2/df = 1.53$, CFI = .90, SRMR = .076; RMSEA = .058 (90% confidence interval from .042 to .072) indicating better fit. Factor analyses indicated a four-factor structure as most appropriate. The interrelations between the factors are different in comparison with the interrelations between the factors of student performance on the wards and not all items load on the same factor. The factor “cognitive abilities” seemed to be independent from “interpersonal skills” and “professional qualities.”

The average item score of the factor “cognitive abilities” was highest ($M = 4.80$, $SD = .71$) followed by the more specific factor “patient workup” ($M = 4.61$, $SD = .78$). The factors “interpersonal skills” and “professional abilities” had average item values of 3.92 ($SD = .96$) and 3.97 ($SD = .77$), respectively. “Cognitive abilities” and the “patient workup” are obviously the most important components of competence, and “interpersonal skills” and “professional qualities” the least important for student performance at examinations (with consistent patterns across specialties). All factors are significantly different from each other; with the exception of the interrelationship between the factors “interpersonal skills” and “professional qualities.” The t-values (160) of the factor pairs that were significantly different ranged from 3.08 to 10.36, $p < .005$.

To investigate how the two factor models are related to each other we place the two models in one conceptualized figure as displayed in Figure 1. The left side shows the factor structure of competencies important for student performance on the wards and the right side shows the factor structure of competencies important for student examination. For both sides we could identify and label the commonly used three main dimensions of clinical competence: cognitive abilities, interpersonal skills, and professional qualities. The difference between both factor models is a more specific factor labeled "motivation" on the left side and a more specific factor on the right side labeled "patient workup." All items and factors common to both models are shaded. Despite the similarities between the two factor models, the structural models of the two factor models are too different to evaluate between-model differences in factor means using SEM. Re-specification of the structural models to achieve similarity will lead to poor fitting models (Bentler, 2005). In order to make global comparisons between the individual factors of both models possible, we consider the more specific factor "motivation" which is relatively highly correlated with the factor "professional qualities" (r = .64) and the more specific factor "patient workup" which is relatively highly correlated with the factor "cognitive abilities" (r = .69) both as one factor (as marked in Figure 1). This will lead to two similar and comparable three-factor models (i.e., cognitive abilities, interpersonal skills, and professional qualities). Subsequently, the average raw item score of similar items loading on the same factor are calculated to serve as indicator for the respective factor mean. For the average item score of factor "cognitive abilities" we found a significant difference between student performance on the wards and student performance at examinations. The mean difference was -.19, t (345) = -2.55, p = .011. The average item score difference of the factors "interpersonal skills" and "professional qualities" between student performance on the wards and student performance at examinations, were .89, t (345) = 9.93, p < .0005, and .66, t (345) = 8.78, p < .0005, respectively. This indicates that cognitive abilities are ranked higher for student performance at examinations, while interpersonal skills and professional qualities are

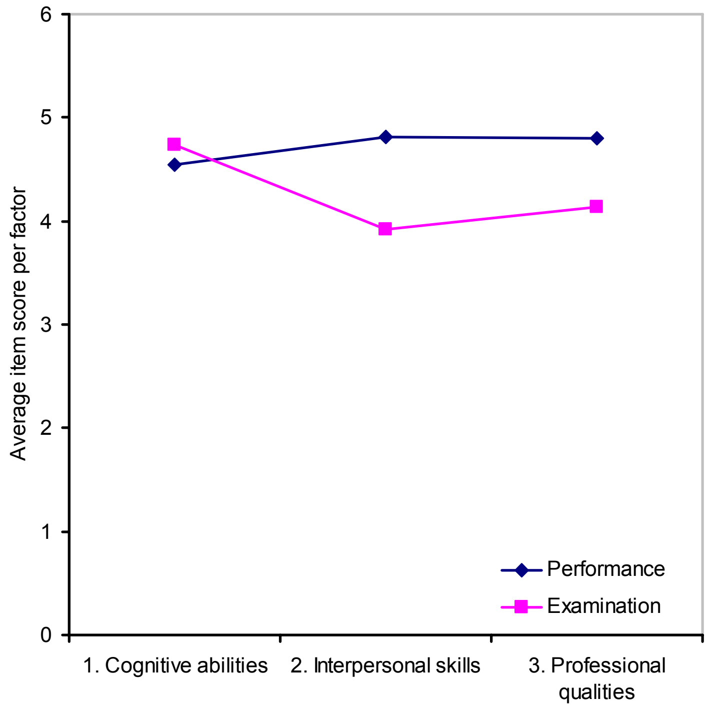


Figure 2. *Average item score per factor (cognitive abilities, interpersonal skills, and professional qualities) for student performance on the wards and student performance at examinations*

ranked higher for student performance on the wards. See figure 2 for the average item score per factor for student performance on the wards and student performance at examinations.

Discussion

Clinical rotations play an important role in the medical curriculum and are considered crucial for student learning. Clerkship assessment often relies on global evaluations of observed behaviors and examination of performance at the end of the clerkship. However, competencies that were intended to be learned can differ from what is assessed.

Discussions about defining components of competence are not new. Many organizations involved in medical education delivered as many different definitions (ABIM, 1979, 2001, 2002; Anderson et al., 1998; CanMEDS, 1996; General-Medical-Council, 2001; Greiner & Knebel, 2003; NBME, 2002). One of the main objectives for defining competencies was to serve as a framework for assessment purposes (Burg & Lloyd, 1983). We saw that the early definitions were all based on the diagnostic process (Hubbard et al., 1965). Later versions built on this foundation, but got a wider angle; further characteristics were related to, for example, providing good patient care, good communication skills, or professionalism. The continuum of the medical profession is wide and can go from dignity to respecting patients (See for a current overview about professionalism, Stern, 2006).

For the present study, a survey that listed 21 competencies was developed and administrated to clinical educators and physicians of different hospitals and specialties involved with student learning on clerkship rotations. The survey was intended to explore what competencies are considered important during adequate student performance on the wards and to what extent the same competencies were assessed at examinations.

The modeling step in this study was conducted to investigate the underlying simple factor structure. For the ratings on student performance on the wards and student performance at examinations a four-factor model explained the data best. In order to get a general idea about the similarities and differences between the factor structures, we combined both factor models in one conceptualized model, as depicted in Figure 1. The main findings can be summarized as follows: (1) Factor analysis revealed that three main factors with almost identical item loadings could be identified for both factor models. These three components of clinical competence followed the well known classic categorization in cognitive abilities, interpersonal skills, and professional qualities (Epstein & Hundert, 2002; Forsythe, McGaghie, & Friedman, 1986). (2) Looking at the model in more detail, then the factors are not considered equally important for student performance on the wards vs. student performance at examinations. Cognitive abilities are relatively more important for student performance at examinations and interpersonal skills and professional qualities are relatively more important for student performance on the wards. The difference could refer to a more commonly addressed problem that clinical educators very rarely directly observe students taking a history or doing a physical examination on the wards (Pulito, Donelly, Plymale, & Mentzer Jr, 2006). This aspect of the diagnostic process (in our model labeled: "patient workup") deserved, together with the factor "cognitive abilities," more attention during examinations. (3) Further differences between the two factor models seemed to be in a motivational factor during ward performance and a patient workup factor for student examination. Both minor factors were part of a main factor for the other factor model; professional qualities and cognitive abilities, respectively. (4) The relationships among the factors between the two factor models are remarkable different from structure. First, the cognitive abilities factor of the student performance on examinations model had no significant relationship with the factors: "interpersonal skills" and professional qualities" while all factors were moderately to strongly related with each other for student performance on the wards. Student examinations seemed to be centered on cognitive abilities. All competencies that load on this factor were rated higher for student examinations (i.e., knowledge of diseases, biomedical knowledge, generating differential diagnoses, and structured problem solving) and this factor was not correlated with the other main factors of the student performance at examinations factor model. The further branching of this cognitive abilities factor in patient workup (with loadings on skill in history and physical examination, writing clear patient notes, and conscientiousness) emphasized the importance of the diagnostic process during clerkship examinations.

An often addressed problem with regard to the definition of clinical competence is that there is little consensus about the components and that there is little agreement among specialties (Epstein & Hundert, 2002; Forsythe et al., 1986; Neufeld, 1985; Van der Vleuten et al., 2000). Our data, however, indicated the opposite. There is a high consensus among the different specialties on how important certain competencies are. For adequate performance on the wards, every specialty agreed that "respect for the patient" was the most important competency. For performance on examinations the emphasis fell on the diagnostic process ("knowledge of diseases" and "skill in history and PE"). The top ranked competencies were all related to the factors: cognitive abilities and patient workup. This was true across specialties. This does not mean that competencies are not dependent on a particular context or discipline (some differences among specialties were identified), but it means that, at least according the perspective of clinical educators, general characteristics can be identified which transcend a particular context.

The reader with a background in psychology would probably notice similarities with personality theories. The basic dimensions of personality according to the Big Five and their parallels with our model are: extraversion/emotionality ("interpersonal skills"), conscientiousness/agreeableness ("professional qualities"), and intellect ("cognitive abilities"). Extraversion/emotionality ("interpersonal skills") refers to an open expression of impulses, a sense of sociability. Sometimes it is a kind of dominance or confident assurance, whereby agreeableness refers to a sense of nurturance and emotional supportiveness. Conscientiousness/agreeableness (“professional qualities”) refers to characteristics such as planning, persistence, and purposeful striving towards goals. Intellect (“cognitive abilities”) refers mostly to thinking and knowledge. In this perspective it makes sense why extraversion/emotionality and conscientiousness/agreeableness are personality dimensions more valued during daily performance on the ward: a real life setting whereby the patient-doctor interaction is most important (McCrae & Costa, 1987).

Several limitations should be considered. This survey was administrated in a medical curriculum where oral end-of-clerkship examinations are still in use. Generalizability to other assessment tools could be limited (we could think about the modern deviations of oral examinations, such as the mini-clinical exercise and the Objective Structured Clinical Examination). Hence the debated reliability of oral examinations, it is suggested that these examinations measure a combination of knowledge, problem-solving ability, and personal characteristics. On the other hand, oral examinations are often valued because of their face validity. It can provide information about student’s clinical judgment capability and the ability to respond to changes in the environment (Muzzin & Hart, 1985).

The consensus found in the ranking of the competencies across specialties is partly related to the broad definition of the competencies used in this study. For example, “knowledge of diseases” was among the top rankings across specialties; however, it is logically to assume that the content of this knowledge is different for each individual specialty. It is possible to come to an agreement in the definition of clinical competence across specialties if competencies are broad defined. However, this conclusion can not be reversed, whereby in this particular example: “knowledge of diseases” is not a general trait that transcends among specialties.

Conclusions

Evaluating the degree of importance of different components of competence is important for assessment purposes. According to the perceptions of clinical educators, what is important for adequate performance on the wards is considered less important for examination. For daily performance of the student, characteristics like motivation, interpersonal skills, and other professional qualities (e.g., collegiality, working in teams, respect for the patient, communication skills) are more important than cognitive abilities (e.g., knowledge of diseases, problem solving skills, history taking and physical examination), while for clerkship examination the opposite is true. Assessment is considered one of the most difficult aspects of clinical education and faculty skills in assessment methods are often lacking (Mennin & Kalishman, 1998). Every attempt to make more reliable and valid measurement methods will have less chance of succeeding if clinical educators lack the necessary skills or think differently about student examination in contrast to what is required for adequate daily performance on the wards. If we take into account the effects of assessment on learning, then this observation could jeopardize the future development of clinical competence. Future research should focus on an explanation for this discrepancy in perception. A start can be made by using a more qualitative approach on the perceptions of clinical educators. They are the center of medical education and responsible for educating our new generation of doctors.

References

ABIM (1979). American Board of Internal Medicine: clinical competence in internal medicine. *Annals of Internal Medicine, 90*, 402-411.

ABIM (2001). *Evaluating your clinical competence: new competencies for internal medicine*. Philadelphia: American Board of Internal Medicine.

ABIM (2002). *Residents: Evaluating your clinical competence. New competencies for internal medicine*. Philadelphia, Pennsylvania: American Board of Internal Medicine Clinical Competence Program.

Anderson, M. B., Cohen, J. J., Hallock, J. E., Kassebaum, D. G., Turnbull, J., & Whitcomb, M. E. (1998). *Report 1 Learning Objectives for Medical Student Education-guidelines for Medical Schools.* Washington: AAMC.

Bentler, P. M. (2003). EQS (Version 6.1 Beta) [Computer Software]. Encino, CA: Multivariate Software, Inc.

Bentler, P. M. (2005). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software, Inc.

Burg, F. D., & Lloyd, J. S. (1983). Definitions of competence: A conceptual framework. In J. S. Lloyd & D. G. Langsley (Eds.), *Evaluating the skills of medical specialists* (pp. 1-71). Chicago: American Board of medical Specialties.

Burg, F. D., Lloyd, J. S., & Templeton, B. (1982). Competence in Medicine. *Medical Teacher, 4*(2), 60-64.

CanMEDS (1996). *Skills for the new millennium: report of the societal needs working group*. Ottawa, Canada: CanMEDS 2000 Project Societal Needs Working Group Report.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA, 287*(2), 226-235.

Forsythe, G. B., McGaghie, W. C., & Friedman, C. P. (1986). Construct validity of medical clinical competence measures: a Multitrait-Multimethod Matrix study using confirmatory factor analysis. *American Educational Research Journal 23*(2), 315-336.

General-Medical-Council. (2001). *Good Medical Practice.* London: General Medical Council.

Greiner, A. C., & Knebel, E. (Eds.). (2003). *Health profession education: a bridge to quality.* Washington. D. C.: The National Academies Press.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-31.

Hubbard, J. P., Levit, E. J., Schumacher, C. F., & Schnabel, T. G. (1965). An objective evaluation of clinical competence: new technics used by the National Board of Medical examiners. *New England Journal of Medicine, 272*, 1321-1328.

Jolly, B. C. (1994). *Bedside manner: teaching and learning in the hospital setting*. University of Maastricht, Maastricht.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling.* New York, NY: The Guilford Press.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing 438 approaches to setting cutoff values for fit indexes, dangers in

overgeneralizing Hu and Bentler's (1999) 439 findings. *Structural Equation Modeling, 11*, 320–341.
McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81-90.
McManus, I. C., Richards, P., Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ, 316*, 345-350.
Mennin, S. P., & Kalishman, S. (1998). Student assessment. *Academic Medicine, 73*(10 Suppl), S46-S54.
Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine, 65*(9 Suppl), S63-S67.
Muzzin, L. J., & Hart, L. (1985). Oral examinations. In V. R. Neufeld & G. R. Norman (Eds.), *Assessing clinical competence.* New York: Springer-Verlag.
NBME (2002). *Embedding professionalism in medical education: assessment as a tool for implementation*. Baltimore: National Board of Medical Examiners.
Neufeld, V. R. (1985). Historical perspectives on clinical competence. In V. R. Neufeld & G. R. Norman (Eds.), *Assessing Clinical Competence.* New York: Springer-Verlag.
Newble, D. I. (1978). How do we measure clinical competence? *The Medical Journal of Australia, 1*, 637-639.
Pulito, A. R., Donelly, M. B., Plymale, M., & Mentzer Jr, R. M. (2006). What do faculty observe of medical students' clinical performance. *Teaching and Learning in Medicine, 18*(2), 99-104.
Rice, C. A., & Sinclair, M. (1995). Competency-based objectives for clinical training. *Canadian journal of medical technology, 57*(3), 136-140.
Stern, D. T. (Ed.). (2006). *Measuring Medical Professionalism.* New York: Oxford University Press.
Van der Hem-Stokroos, H. H., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., De Vries, H., & Haarman, H. J. T. M. (2001). How effective is a clerkship as a learning environment? *Medical Teacher, 23*(6), 599-604.
Van der Vleuten, C. P. M., Scherpbier, A. J. J. A., Dolmans, D. H. J. M., Schuwirth, L. W. T., Verwijnen, G. M., & Wolfhagen, H. A. P. (2000). Clerkship assessment assessed. *Medical Teacher, 22*(6), 592-600.
Van Leeuwen, Y. D., Düsman, H., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R. P. T. M., et al. (1997). Factors Influencing the Growth in Knowledge of Trainees in General Practice. *Advances in Health Sciences Education, 2*(1), 61-70.
Wass, V., Van der Vleuten, C., Shartzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet, 357*, 945-949.
Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (2006). Influence of clerkship experiences on clinical competence. *Medical Education, 40*(5), 450-458.

Appendix A

Average response for all 21 competencies for student performance on the wards and student performance at examinations, and their mean difference across specialties (6-point Likert scale scores)

Competencies	Performance on the wards			Performance at examinations			Difference	
	N	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*
1. biomedical knowledge	186	4.12	1.19	161	4.56	1.11	156	-0.46
2. collegiality	186	4.76	0.98	161	3.68	1.30	156	1.09
3. good presentation skills	186	4.57	0.84	158	4.44	0.96	156	0.14
4. open to criticisms	186	4.67	0.90	161	3.69	1.14	156	0.94
5. conscientiousness	185	4.95	0.79	159	4.53	1.01	155	0.40
6. overall improvement	184	4.61	0.92	159	4.40	1.11	154	0.18
7. generating differential diagnoses	186	4.52	0.93	160	4.84	0.94	156	-0.38
8. assertiveness	186	3.94	0.93	161	3.17	1.03	156	0.75
9. motivation and interest	186	5.20	0.81	160	4.64	1.07	156	0.53
10. reflect on own performance	184	4.55	0.98	158	3.66	1.21	154	0.87
11. knowledge of diseases	186	4.57	0.92	160	5.00	0.92	156	-0.43
12. inquisitiveness	185	4.87	0.81	161	4.39	1.09	155	0.44
13. problem solving	184	4.82	0.78	160	4.78	0.93	154	0.08
14. dexterity	184	3.63	1.24	158	3.00	1.28	154	0.67
15. courtesy	185	4.64	0.95	161	3.82	1.28	155	0.86
16. good communication skills	186	5.04	0.75	159	4.33	1.11	156	0.71
17. respect for the patient	185	5.36	0.75	160	4.46	1.26	155	0.89
18. personal appearance	186	4.38	0.99	160	3.38	1.30	156	1.01
19. skill in history and PE	186	4.88	0.83	160	4.78	0.97	156	0.08
20. writing clear patient notes	185	4.65	0.93	161	4.51	1.05	155	0.10
21. functions well as part of a team	185	4.63	0.92	160	3.60	1.22	155	1.07

Note: PE stands for physical examination

Appendix B

CLINICAL COMPETENCE SURVEY

Name: …………………….... ………………….... Specialty: …………………….... …………………....

To what degree are these aspects of competence important during good performance of students on the wards, and to what degree are these aspects employed for examination of students on the wards?

I am involved with: (mark what is applicable)

- ○ student performance on the wards
- ○ examination of students on the wards
- ○ none of those

	Degree of importance for student ward **performance**						Degree of importance for student **examination**					
	1 = not at all					**6** = very much	**1** = not at all					**6** = very much
	1	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
1. biomedical knowledge	○	○	○	○	○	○	○	○	○	○	○	○
2. collegiality	○	○	○	○	○	○	○	○	○	○	○	○
3. good presentation skills	○	○	○	○	○	○	○	○	○	○	○	○
4. open to criticisms	○	○	○	○	○	○	○	○	○	○	○	○
5. conscientiousness	○	○	○	○	○	○	○	○	○	○	○	○
6. overall improvement during clerkship	○	○	○	○	○	○	○	○	○	○	○	○
7. generating differential diagnoses	○	○	○	○	○	○	○	○	○	○	○	○
8. assertiveness	○	○	○	○	○	○	○	○	○	○	○	○
9. motivation and interest	○	○	○	○	○	○	○	○	○	○	○	○
10 reflect on own performance	○	○	○	○	○	○	○	○	○	○	○	○
11. knowledge of diseases	○	○	○	○	○	○	○	○	○	○	○	○
12. inquisitiveness	○	○	○	○	○	○	○	○	○	○	○	○
13. structured problem solving	○	○	○	○	○	○	○	○	○	○	○	○
14. dexterity	○	○	○	○	○	○	○	○	○	○	○	○
15. courtesy	○	○	○	○	○	○	○	○	○	○	○	○
16. good communication skills	○	○	○	○	○	○	○	○	○	○	○	○
17. respect for the patient	○	○	○	○	○	○	○	○	○	○	○	○
18. personal appearance	○	○	○	○	○	○	○	○	○	○	○	○
19. skill in history and physical examination	○	○	○	○	○	○	○	○	○	○	○	○
20. writing clear patient notes	○	○	○	○	○	○	○	○	○	○	○	○
21. functions well as part of a team	○	○	○	○	○	○	○	○	○	○	○	○

What aspects of clinical competence are missing?

1. …………………………. 2. …………………………. 3. ………………………….

○ Yes, keep me informed about the research results

Thank you for your cooperation

CHAPTER SIX

Inducing Expertise Effects in Clinical Case Recall through the Manipulation of Processing

This study was conducted at the department of Psychology, Erasmus University Rotterdam (EUR), The Netherlands. The department of Psychology of the EUR funded this study. An earlier version of this paper was presented at the 84th Annual Meeting of the American Educational Research Association (AERA) at Chicago, Illinois, April 21-25, 2003. Trustfonds of the EUR partly supported participation of the presenting author.

Reference
Wimmers, P. F., Schmidt, H. G., Verkoeijen, P. P. J. L., & van de Wiel, M. W. J. (2005). Inducing expertise effects in clinical case recall. Reprinted from *Medical Education, 39*(9), 949-957, © 2005 with kind permission of Blackwell Publishing.

Abstract

BACKGROUND This study was directed at illuminating a well known phenomenon in the medical expertise literature, the “intermediate effect” in clinical case recall. This robust phenomenon consists of the finding that medical students of intermediate levels of expertise outperform both experts and novices in clinical case recall after diagnosing cases. It deals in particular with the findings of some researchers who have reported a monotonically increasing recall with level of expertise.

PURPOSE To address possible causes for this anomaly in medical expertise and to experimentally demonstrate how data elaboration can cause expertise effects in clinical case recall.

METHOD Expert nephrologists, intermediate level students and novices were presented with 6 medical cases under 3 different conditions: laboratory data cases without special instructions, laboratory data cases with instructions to elaborate, and cases with laboratory data and a relevant clinical context.

RESULTS Only when participants were required to elaborate on each of the information units presented to them did case recall show an expertise effect. If laboratory data are framed within the context of a patient’s history and physical examination data, the “intermediate effect” appears.

CONCLUSIONS The instructions used in the laboratory-data-only-plus-elaboration condition seem to have induced a deeper, more detailed, analysis of the patient case. It is therefore interesting to note that these instructions only affected the recall of the experts and had no effect on the novices’ or intermediates’ recall. We might conclude from this that expertise effects in clinical case recall are only produced when the normal processing of patient information is disrupted.

IT IS GENERALLY agreed upon that superior problem-solving performance by experts is caused by extensive knowledge in their area of specialization (Chi, Glaser, & Farr, 1988). These expert-unique knowledge structures facilitate the solving of domain-specific problems and support memory. Therefore, if experts are asked to recall information relevant to a problem processed, they tend to outperform novices or advanced students. The classic studies on chess expertise for instance, conducted by the Dutch psychologist Adriaan de Groot (1946) demonstrated a monotonically increasing function of expertise level and memory for mid-game chess positions. This increasing recall with level of expertise is also found in other domains, such as memory for bridge hands (Charness, 1979), baseball games (Spillich, Vesonder, Chiesi, & Voss, 1979), and computer programming lines (McKeithern, Reitman, Rueter, & Hirtle, 1981).

However, there are noticeable exceptions. If *medical* experts engage in diagnosing a case and subsequently are asked to write down the details of this case, they are often outperformed by advanced medical students who tend to produce more elaborate recall than either experts or novices. This peculiar phenomenon, known as the "intermediate effect in clinical case recall," is a robust phenomenon (Baramee, 2003; Boshuizen & Schmidt, 1992; Claessen & Boshuizen, 1985; Gilhooly et al., 1997; Muzzin, Norman, Feightner, & Tugwell, 1983; Patel & Groen, 1991; Rikers, Schmidt, & Boshuizen, 2000; Schmidt & Boshuizen, 1992, 1993a, 1993b; Schmidt, Boshuizen, & Hobus, 1988).

To explain these somewhat counterintuitive findings, some authors suggest that experts may learn as many important features from a case as students but that they only report the important data and leave out unimportant details. However, recently Eva, Norman, Neville, Wood, and Brooks (2002) have demonstrated that experts' recognition memory of case information presented to them is also poorer. As an alternative framework to explain the intermediate effect, the knowledge encapsulation theory was proposed (Schmidt & Boshuizen, 1992). This theory assumes that medical experts, while processing a case, chunk relevant information into higher level summarizing concepts (or "encapsulations"). When asked to recall a case, they tend to recall these concepts rather then the "raw" information. Hence the less extensive recall protocols of medical experts.

The intermediate effect in clinical case recall and its interpretation have been subject of considerable debate, not only because this finding is at variance with results from other domains of expertise, but also because some studies in medicine failed to display the effect. One of these exceptions is a study of Norman, Brooks, and Allen (1989) in which expert nephrologists and advanced medical students were presented with six cases consisting solely of laboratory data. In one of their experimental conditions, medical experts and advanced students diagnosed cases and were subsequently asked for recall. In addition, they were instructed to undertake, while processing the case, what the investigators called: "formulation of the patient's problem;" that is, they had to relate aloud the laboratory data to what they saw as the underlying pathology. When the participants were subsequently asked to recall the laboratory values, the experts outperformed the advanced students. So, unlike most other studies in medicine, Norman and colleagues were able to demonstrate an expertise effect in recall (Norman et al., 1989). The authors of this study suggest that experts' better and more detailed recall was the result of processing more information by experts. Of course, the question then is: why do experts process more information under this condition?

In an attempt to explain the findings of Norman et al., (1989), Verkoeijen, Rikers, Schmidt, Van de Wiel, and Kooman (2004) assumed that the atypical results of the Norman study were caused by an atypical task; doctors almost never process cases of laboratory data in isolation (i.e., without a relevant clinical context). Processing laboratory data in isolation might cause experts to pay more attention to each of these data separately, looking for the types of clues that are otherwise be provided by the complaints with which the patient presents to the doctor and the context in which these complaints have arisen. Therefore, these investigators replicated the Norman et al., (1989) study, and added a control condition in which participants had to process the same laboratory data, but this time framed within the context of a short case history in which patient's complaint, physical examination and other findings were briefly described. Their results indicate that, although processing laboratory-data-only took more time and led to higher recall than processing laboratory data in an appropriate clinical context, no effects of expertise on case recall were found. In fact, under both conditions, experts recalled the same number of laboratory data as advanced students. Thus, they were not able to replicate Norman's findings in terms of establishing the positive relationship between level of expertise and laboratory recall. The authors suggested (but did not actually test) that differences in the nature of the experts employed in both studies might be responsible for their failure to replicate. The experts in the Norman et al., study (1989) were predominantly nephrologists, whereas the expert groups in the Verkoeijen et al., (2004) study consisted mostly of internists. Because the laboratory data cases used by Norman et al., (1989) were primarily designed for nephrologists, this may account for the differences.

There may, however, be another explanation for the failure to replicate the findings of Norman et al., (1989)required participants to think aloud while explicitly relating the laboratory data to what they saw as the patient's problem. The reason for this instruction was to establish an expectation among participants that the investigators were interested in diagnosis rather than recall. This instruction may have induced the participants to *semantically elaborate* upon the laboratory data information. It is well known from the literature that semantic elaboration of stimuli leads to better memory for those stimuli (Craik & Lockhart, 1972). In addition, the procedure employed by Norman and colleagues resembles procedures used in the literature on self-explanation (Chi, 2000; Norman et al., 1989). When students are required to explain what they have read of a text while processing it, their understanding of the text increases (Chi, De Leeuw, Chiu, & LaVancher, 1994).

Students who solved problems while giving elaborative explanations about what they were doing and why they were doing it performed better than those who did engaged in self-explanation poorly (Chi, Lewis, Reimann, & Glaser, 1989).

In summary, we suggest that the particular elaboration procedure that Norman and colleagues (1989) required may have led their participants to more elaborate processing and, hence, better memory. The implication is that if they are not asked to elaborate, the processing by experts may be more shallow or more functional to the task at hand, and therefore the expertise effect in recall may not emerge. This may have been the case in the Verkoeijen et al., (2004) study, because it did not require participants to elaborate. To test this hypothesis, novices, intermediate medical students, and nephrologists were asked to diagnose the six cases used by both Norman and Verkoeijen. In the laboratory-data-only condition, participants simply read the cases, provided a diagnosis, and recalled the laboratory data. This was the condition applied in the Verkoeijen et al., (2004) study. In the second condition, described as the laboratory-data-only-plus-elaboration condition, participants related laboratory data to the problem while thinking aloud and then provided diagnosis and recall. This was the procedure used by Norman et al., (1989). A third condition, the laboratory-data-plus-context condition, required participants to process the same laboratory data in the context of a patient's complaint and other findings. This condition was similar to clinical context condition in Verkoeijen et al., (2004). This experiment enabled us to compare the various interpretations of the original Norman et al., (1989) findings.

Methods

Participants

The 30 experts were all nephrologists from various hospitals in the Netherlands. The 24 novices in this study were all year four preclinic students. The 24 intermediates were students who had completed their clerkships in internal medicine. The participants were randomly assigned to one of the three conditions and received a small compensation for their participation.

Material

Six patient cases were used in two different formats. In the control (laboratory-data-plus context) condition, the cases comprised of two sheets. The first sheet provided a clinical context (i.e. the medical history of the patient, the results of the physical examination and some additional findings). The clinical context was described in the type of language normally used by patients, so that medical interpretations were excluded. The second sheet contained 19 laboratory data, 8 blood chemistry, 7 urine chemistry, and 4 blood gases, including the normal range of values for each value presented in a standard medical format. These were the laboratory data cases used by Norman et al., (1989) The two following sheets were blank response sheets for respectively; diagnosis and recall. The clinical contexts used in addition to the laboratory data were identical to the cases used in the study of Verkoeijen et al., (2004). (The cases in the Norman et al., study used 20 laboratory data values. The present study and that by Verkoeijen et al., used 19 of these 20 values because the value of urine calcium is not normally presented to nephrologists in Dutch hospitals.) An example of a complete clinical case used in the laboratory-data-plus-context condition is presented in Appendix A.

The cases of the laboratory-data-only condition and the laboratory-data-only-plus-elaboration condition were similar to those in the laboratory-data-plus-context condition, but without the clinical context. The diagnoses related with the cases are listed in Table 1. The presentation sequence of the cases in all conditions was counterbalanced to control for order effects.

Table 1. *The diagnostics of the cases*

Case	Diagnoses
1	Addison's disease with tuberculosis
2	Diabetes mellitus type II
3	Exacerbation chronic obstructive pulmonary disease (COPD)
4	Tranxene intoxication induced respiratory acidosis
5	Hyperventilation induced respiratory alkalosis
6	Metabolic alkalosis instantiated by a milk-alkali syndrome

Procedure

All participants were informed that six cases would be successively presented and that they would have to read through each case to provide a diagnosis. Subsequently, they were told to write down, for each case, all the information they remembered. In the control condition (laboratory-data-plus-context), participants were also informed that they were allowed to turn pages back and forth between the clinical context and the laboratory data. In the laboratory-data-only-plus-elaboration condition the instructions were different. The participants were told that they had to read through each case and had to formulate the patient's problem and had to explain how the individual laboratory values were related to this problem. This was the same instruction Norman et al., (1989) had used. There were no time constrains for the procedure, although processing times were recorded from the time that participants started to read the case to the time a diagnosis was given. A sample case was given in advance to familiarize the participants with the procedure. All participants were tested separately.

Analysis

The analysis of the diagnoses and free recall protocols were essentially the same as in the study by Norman et al., (1989) in order to make comparisons between the studies as straightforward as possible. The criterion diagnoses were each subdivided into a number of diagnostic elements. Diagnoses produced by the participants were compared to the criterion diagnoses and awarded with points for each accurate diagnostic element. The resulting diagnostic accuracy score of the participant ranged from 0 to 10. For example, if the participant formulated: hypoxemia and chronic respiratory acidosis on the basis of exacerbation chronic obstructive pulmonary disease (COPD) caused by an infection of the bronchial tubes, then this diagnoses was broken down into four components: Hypoxemia; chronic respiratory acidosis; infection of the bronchial tubes; and Exacerbation COPD. The weights corresponding to these accurate components on the criterion diagnosis were respectively, 3, 2, 1, and 4 points (based on expert provided relative importance ratings). An accurate diagnosis could lead to a total score of 10 points.

The free recall protocols were segmented into idea units or propositions and scored by means of a proposition analysis method introduced by Patel and Groen (1986a). Propositions consist of two concepts connected by a qualifier, such as specification (spec) or temporal information (temp). For instance, the text fragment "A 45 year old man from Somalia complains about nausea, vomiting and abdominal pain. His complains are since three weeks" consists of six propositions: (1) man-spec (45 years old); (2) man-spec (from Somalia); (3) complaints-spec (nausea); (4) complaints-spec (vomiting); (5) complaints-spec (abdominal pain); (6) complaints-temp (three weeks). The collection of propositions forms the description of a case. Each proposition in free recall was matched against the propositions in the case. In the matching procedure the propositions recalled were dependent on their relationship with one or more text propositions, subdivided into four categories: literal propositions, paraphrased, low-level, and high-level inferred propositions. Low-level inferences are inferences based on only one proposition in the text, while high-level inferences (or "encapsulations," see Schmidt and Boshuizen, 1993a) are inferences based on more than one proposition from the text (Coughlin & Patel, 1987). High-level inferences were included in the analysis only if they could be matched to a number of propositions in the text. The total recall score was obtained by adding the number of literally recalled propositions, paraphrased propositions, low-level inferences, and high-level inferences.

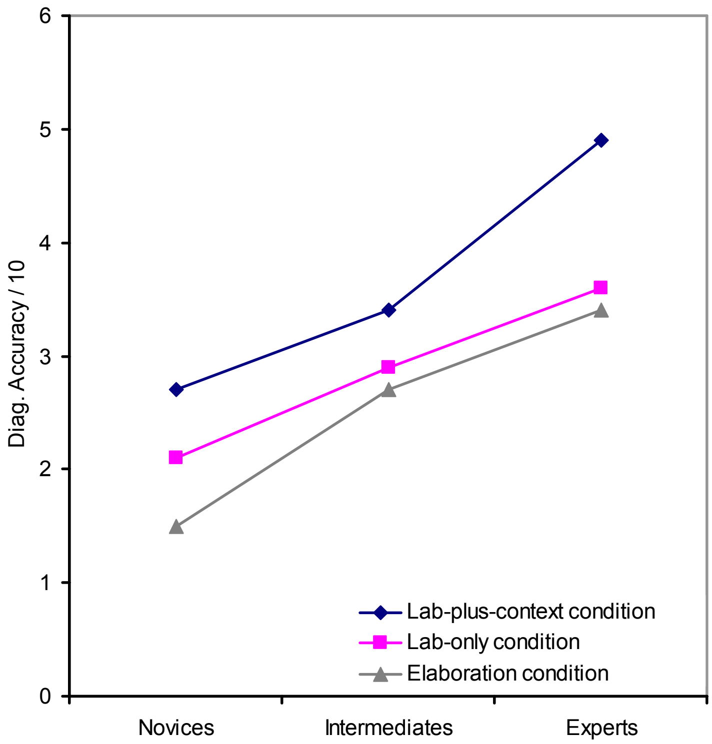
Two independent raters separately scored the diagnostic accuracy and free recall of a random subset of 12 protocols. A Pearson correlation of .86, $p < .01$ and .98, $p < .01$, respectively, were found between the raters. These correlations indicate that the procedure was reliable; therefore the remaining protocols were scored by one judge. Data were analyzed by using repeated measures (ANOVA), and Fisher's least significant difference (LSD) was used to make post hoc multiple pairwise comparisons between the different expertise groups. An alpha level of .05 was used for all statistical analyses.

Results

Processing time

Analysis of mean processing time of the six cases showed a significant effect for level of expertise, F (2, 69) = 5.79, MSE = 19872.23, p = .005. Pairwise comparisons revealed that mean laboratory data processing time did not differ significantly between novices and intermediates, but differed significantly between novices and experts, and intermediates and experts, the experts being faster. Significant differences between all conditions among the experts were also found: they were fastest in the laboratory-data-plus-context condition and slowest in the laboratory-data-only-plus-elaboration condition (experts' mean processing time in the laboratory-data-plus-context, laboratory-data-only, and laboratory-data-only-plus-elaboration conditions were, respectively, 59.3, 113.3 and 169.6 seconds, where the processing time of the laboratory-data-plus-context condition is corrected for the time it takes to read the context). Thus, having a clinical context available speeds up the processing of laboratory data considerably.

Figure 1. *Diagnostic accuracy as a function of level of expertise*



Diagnostic accuracy

Mean diagnostic accuracy increased with levels of expertise, F (2, 69) = 24.11, MSE = 5.70, p < .0001. Furthermore, a main effect of experimental condition was found, F (2, 69) = 9.58, MSE = 5.70, p < .0001, with no significant interaction. Pairwise comparisons of level of expertise revealed significant differences between all groups, indicating that experts provided more accurate diagnoses than intermediates, and intermediates provided more accurate diagnoses than novices. Pairwise comparisons of experimental condition revealed that mean diagnostic accuracy was highest in the laboratory-data-plus-context condition—providing a clinical context clearly improves diagnosis. Diagnostic accuracy as a function of level of expertise for all conditions is shown in Figure 1.

Free recall

Figure 2 shows the laboratory data recall as a function of level of expertise for all conditions. Analysis of laboratory data recall showed significant effects of experimental condition, F (2, 69) = 8.50, MSE = 38.89, p = .001, level of expertise, F (2, 69) = 5.81, MSE = 38.89, p = .005, and a significant interaction, F (2, 69) = 2.81, MSE = 38.89, p = .032. This demonstrates that there are differences within and across conditions and the slope within each condition is dependent on the quantity of laboratory recall. Between-condition comparisons of the novices of each condition revealed a significant effect of condition, F (2, 21) = 4.402, MSE = 43.458, p = .025. This significant effect of condition was caused by the laboratory-data-plus-context condition, which differed significantly from the laboratory-data-only and the laboratory-data-only-plus-elaboration conditions (recall amongst the novices in these 2 later conditions did not differ significantly). Furthermore, there was no significant effect of treatment between the intermediate groups.

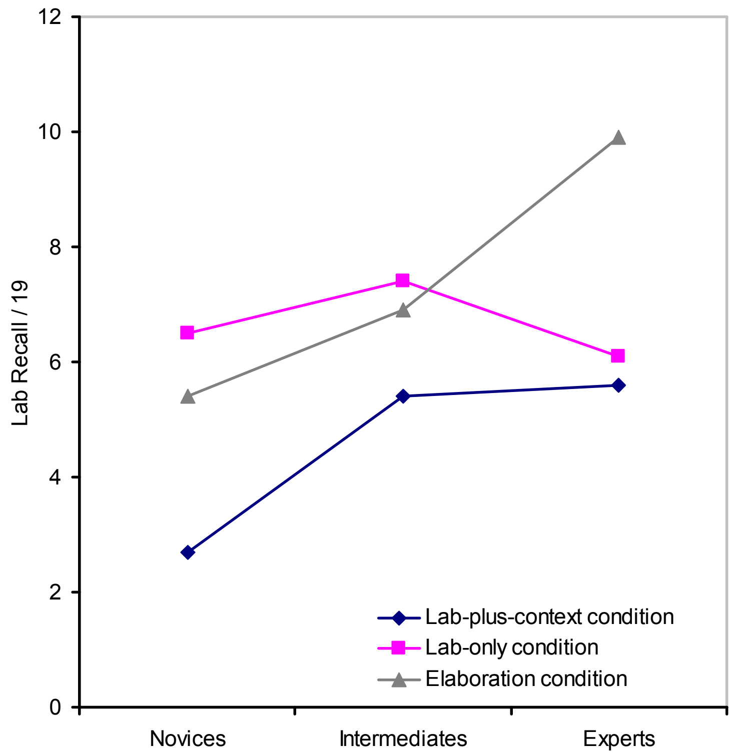


Figure 2. *Laboratory data recall as a function of level of expertise*

Analysis of recall pattern in the laboratory-data-only condition showed no significant effect of level of expertise, F (2, 24) = .635, MSE = 40.177, p = .539. These results are similar to those reported by Verkoeijen et al., (2004). Thus, unlike Norman et al., (1989) assumed, does processing laboratory data with the goal of arriving at a diagnosis not in itself produce an expertise effect in case recall. By contrast, the laboratory-data-only-plus-elaboration condition *did* produce a significant main effect of level of expertise, F (2, 22) = .367, MSE = 49,534, p = .013. Pairwise comparisons within the expert group showed no significant recall differences between the laboratory-data-only and the laboratory-data-plus-context conditions: only the laboratory-data-only-plus-elaboration condition led to recall of significantly more propositions. In fact, recall of the experts increased from 6 propositions in the laboratory-data-only condition to almost 10 propositions in the laboratory-data-only-plus-elaboration condition, an increase of 62%. These findings taken together seem to imply that the processing even of fairly limited and nonredundant materials such as sets of laboratory data in itself does not produce an expertise effect in case recall (as was implied by Norman et al., 1989). An expertise effect is only produced when doctors are required to elaborate explicitly on the material, a task in which they are unlikely to engage in everyday practice.

Total case recall (recall of context plus laboratory data) in the laboratory-data-plus-context condition showed a significant effect of condition, $F\ (2, 23) = 5.284$, $MSE = 271.873$, $p = .013$. Pairwise comparisons revealed significant differences between the novice and intermediate groups, and between the intermediate and expert groups, and no significant difference between the novice and expert groups suggesting that under natural processing conditions, involving both history and laboratory data, an intermediate effect in case recall is found (Figure 3).

Looking only at the high-level inferences (or encapsulations), a significant main effect of level of expertise is found, $F\ (2, 69) = 8.262$, $MSE = 4.058$, $p = .001$, and no further significant main effect or interaction. Pairwise comparisons revealed that mean number of encapsulations did differ significantly between novices and experts, and intermediates and experts. Experts in each condition produced more summaries than novices or intermediates. Although, there was no main effect of condition, a decreasing trend is shown in the number of summaries produced by the expert groups from the control condition, via the laboratory-data-only condition to the laboratory-data-only-plus-elaboration condition.

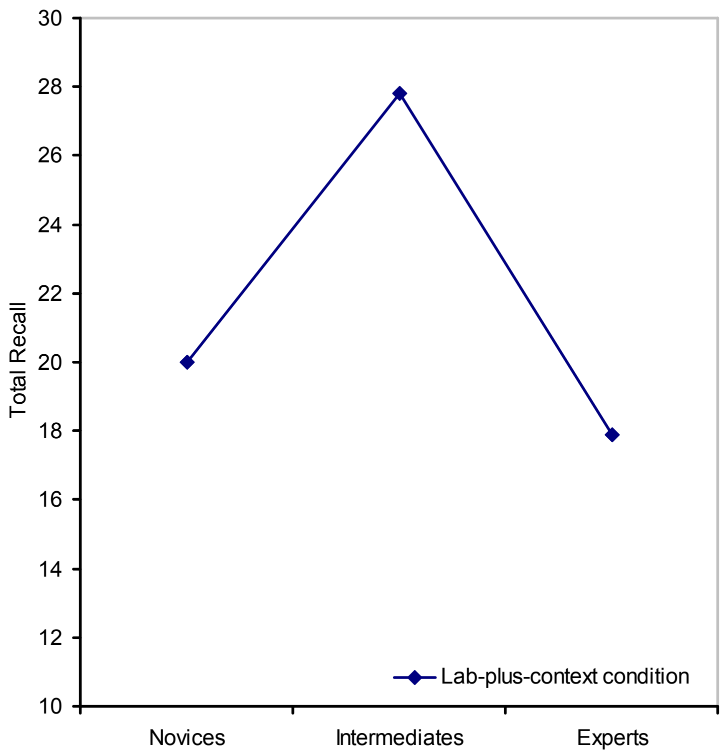


Figure 3. *Total case recall as a function of level of expertise in the laboratory-data-plus-context condition*

Taken together, the recall findings suggest that intermediate effects in clinical case recall only disappear if experts are forced to process medical information under particular constraints. Two of these constraints were demonstrated here. If experts do not have history or physical examination information available about their patients and have to rely on laboratory data only, they reproduce as much from the cases as intermediates. If they are instructed to think aloud and process the lab data in a piecemeal fashion, they even outperform intermediates, as was also demonstrated by Norman et al., (1989). Under natural processing conditions of the same materials, however, intermediates produce more recall than both experts and novices.

Discussion

The failure to demonstrate an expertise effect in clinical case recall has fascinated researchers for many years, in particular because this finding is at variance with extant literature from other domains that suggests that experts may be expected to outperform novices or advanced students when are asked to recall relevant domain-specific information (Charness, 1979; De Groot, 1946; Spillich et al., 1979). However, case recall studies in medicine consistently show advanced students to outperform experts. This intermediate effect in clinical case recall is a robust phenomenon (A common counterargument is that cases used in these studies may not be ecologically valid, although most of the studies showing intermediate effects in recall also display expertise effects in diagnostic accuracy and speed of processing.), (Boshuizen & Schmidt, 1992; Claessen & Boshuizen, 1985; Gilhooly et al., 1997; Muzzin et al., 1983; Patel & Groen, 1991; Rikers et al., 2000; Schmidt & Boshuizen, 1992, 1993a, 1993b; Schmidt et al., 1988). However, some published studies have, in fact, demonstrated an effect of expertise on case recall, the most notable being the study by Norman et al., (1989). The purpose of the present study was to explain these discrepant findings and reconcile them with results from previous studies.

To that end we required novice medical students, intermediates, and expert nephrologists to process nephrology cases under three conditions. One group was to diagnose and recall six cases each consisting of medical history and physical examination information followed by laboratory data; a second group processed only the laboratory data, and a third group processed only the laboratory data with the instruction to elaborate upon them.

The results of the present study can be summarized as follows: only if participants are required to elaborate on each of the information units presented to them does case recall show an expertise effect; that is, recall by experts is better than recall by advanced medical students or novice medical students. These findings are similar to those reported by Norman et al., (1989). If participants are not asked to elaborate but to study cases without any particular processing instructions (other than to come up with a diagnosis), the expertise effect in recall does not emerge. If laboratory data are framed within the context of a patient's history and physical examination data, the well-known intermediate effect appears: advanced students recall more information than either experts or novices.

What do these findings imply? Firstly, it seems that expertise effects in clinical case recall are not so much caused by the nature of the diagnostic task itself (the basic assumption of the Norman et al., study), as by the nature of the processing instructions. The instructions used in the laboratory-data-only-plus-elaboration condition (i.e., to formulate the patient problem and to explain how the individual laboratory values are related to this problem formulation) seem to have induced a deeper, more detailed, analysis of the patient case. Consequently, the information provided by the patient case might have been better remembered and better recalled (Craik, 1979; Craik & Lockhart, 1972; Craik & Tulving, 1973). It is interesting to note however that the instruction to elaborate affected the recall of the experts *only*. It had no effect on the novices' or intermediates' recall.

Secondly, the data show that recall by experts is not the result of some deliberate recall strategy. If intermediate effects in recall are to be explained by assuming that physicians provide only the important information from a case at the expense of the less important (as some critics have maintained), why then would these doctors have refrained from doing so in the laboratory-data-only-plus-elaboration condition? Rather than being the outcome of a recall strategy the findings point at a processing explanation. Schmidt and Boshuizen (1992, 1993a, 1993b) maintain that intermediate-level students and expert doctors operate upon different knowledge while processing a case. By the time they arrive at an advanced level, students have acquired detailed semantic networks that explain disease in terms of its underlying causal pathophysiological structure. These detailed networks enable students to interpret many of the signs and symptoms in a case in terms of these structures. As these signs and symptom are actively interpreted, they tend to be remembered as well. Experienced doctors by contrast have developed categorical knowledge (i.e., a generalization around similar instances, as cited by Brooks, Norman, and Allen, 1991) that enables them to interpret sets of signs and symptoms as unitary wholes rather than each of these separately. This is why, according to Schmidt and Boshuizen (1993b), medical experts' recall of a case contains fewer signs and symptoms and more high-level inferences or summaries. If, however, doctors are constrained in interpreting a case in a holistic fashion, for instance because they are forced to pay attention to each of the information units in front of them separately, then this automatic recoding of symptoms into their encapsulations may fail partially or even entirely.

Thirdly, the elaboration-inducing instructions affected the experts only—a strong argument in favor of the knowledge encapsulation theory, because only the experts are considered to have developed sufficient encapsulations to deal with everyday cases, and, therefore, only their performance is disrupted by the particular treatment. This disruption also manifested itself in the processing speed of the doctors (but not of the students); processing of the laboratory data in the laboratory-data-only-plus-elaboration condition took almost three times longer than processing the same data in the clinical context condition.

Of course, this disruption-of-processing hypothesis needs independent verification. One possibility is to require participants to elaborate on cases for which intermediate effects have been demonstrated in the past (Patel & Groen, 1991; Schmidt & Boshuizen, 1993b). The prediction would be that expertise effects in recall might be induced for these cases as well. A second possibility would be to ask participants to undertake additional and unrelated tasks while processing a case. This would limit working memory capacity and might also disrupt the skill of expert doctors to automatically translate sets of signs and symptoms into their underlying encapsulations.

References

Baramee, J., & Blegen, M. A. (2003). New graduate perception of clinical competence: Testing a causal model. International *Journal of Nursing Studies, 40*, 389-399.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science, 16*, 153-184.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). The role of specific similarity in a medical diagnostic task. Experimental Psychology: General, 120(3), 278-287.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology, 33*, 1-16.

Chi, M. T. H. (Ed.). (2000). *Self explaining: The dual processes of generating inference and repairing mental models* (Vol. 5). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Claessen, H. F. A., & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education, 19*, 61-67.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Medical Education, 62*, 818-828.

Craik, F. I. M. (1979). Levels of processing: Overview and closing comments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 447-459). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684.

Craik, F. I. M., & Tulving, E. (1973). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 102, 268-294.

De Groot, A. D. (1946). *Het denken van den schaker* [The thoughts of the chess player]. Den Haag: Noord Holland.

Eva, K. W., Norman, G. R., Neville, A. J., Wood, T. J., & Brooks, L. R. (2002). Expert-novice differences in memory: a reformulation. *Teaching and Learning in Medicine, 14*(4), 257-263.

Gilhooly, K. J., McGeorge, P., Hunter, J. R. W., Rawles, J. M., Kirby, I. K., & Green, C. (1997). Biomedical knowledge in diagnostic thinking: The case of electrocardiogram (ECG) interpretation. *European Journal of Cognitive Psychology, 9*(2), 199-223.

McKeithern, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organization and skills in computer programmers. *Cognitive Psychology, 13*, 307-325.

Muzzin, L. J., Norman, G. R., Feightner, J. W., & Tugwell, P. (1983). *Expertise in recall of clinical protocols in two specialty areas*. Paper presented at the Proceedings of the 22nd Conference on Research in Medical Education, Washington, DC.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by experts medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology; Learning, Memory and Cognition, 15*, 1166-1174.

Patel, V. L., & Groen, G. J. (1986). Knowledge-based solution strategies and the use of biomedical knowledge by medical students. *Medical Education, 24*, 129-136.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise, prospects and limits.* Cambridge: Cambridge University Press.

Rikers, M. J. P., Schmidt, H. G., & Boshuizen, H. P. A. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 25*(150-166).

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York, NY: Springer Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On Acquiring Expertise in Medicine. *Educational Psychological Review, 5*(3), 205-221.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*(3), 338-351.

Schmidt, H. G., Boshuizen, P. A., & Hobus, P. P. M. (1988, 17-19 August 1988). *Transitory stages in the development of medical expertise: The "intermediate effect" in clinical representations studies.* Paper presented at the 10th Annual Conference of the Cognitive Science Society, Montreal, Quebec, Canada.

Spillich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 275-290.

Verkoeijen, P. P. J. L., Rikers, R. M. J. P., Schmidt, H. G., Van de Wiel, M. W. J., & Kooman, J. P. (2004). Case representation by medical experts, intermediates and novices for laboratory data presented with or without a clinical context. *Medical Education, 38*, 617-627.

Appendix A

Case

A 45-year old man from Somalia feels an increasing fatigue and weakness. In addition, he complains about nausea, vomiting and abdominal pain. His complains are since three weeks even more progressive. He has no chest pain and has no shortage of breath during physical activities. The last couple of days he experiences dizziness. He also complains about fit of coughing, whereby he coughs out mucus. Occasionally, the mucus contains blood.

Patient history: unknown
Intoxication: none
Medication: none
Family history: unknown

Physical examination:
Moderately ill
Temperature: 101.3 F
Blood pressure: horizontal position 110/70; vertical position 85/55
Pulse rate 100/min
Central venous pressure normal
Abdomen: normal peristalsis, normal palpation, no tenderness
Heart: normal
Long: slight crepitations in left upper lobe
Extremities: normal
Pigmentations in the mouth and on the skin
Chest X-ray: cavernous abnormality in left upper lobe

Laboratory values:

			Normal Range			
Blood Chemistry						
	Sodium	132	132	-	145	mmol/l
	Potassium	5.6	3.60	-	5.00	mmol/l
	Calcium tot.	2.47	2.10	-	2.60	mmol/l
	Phosphorus	1.09	0.90	-	1.50	mmol/l
	Urea	15.0	3.0	-	7.0	µmol/l
	Creatinine	108	50	-	125	mmol/l
	Chloride	91	95	-	105	mmol/l
	Glucose	3.67	3.1	-	6.7	mmol/l
Urine chemistry						
	Protein	neg.				
	Glucose	neg.				
	Ketones	neg.				
	Blood	neg.				
	Sodium	310				mmol/l
	Potassium	17				mmol/l
	Urine pH	5.5				
Blood gases						
	pH	7.44	7.35	-	7.45	
	pCO_2	5.73	4.5	-	5.9	kPa
	pO_2	11.73	8.7	-	13.1	kPa
	HCO_3	23	22	-	28	mmol/l

CHAPTER SEVEN

Effects of Level of Expertise on Data-gathering Behavior during different Stages of the Diagnostic Process

This study is conducted at the University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. This project was funded in part by a National Board of Medical Examiners' (NBME) Medical Education Research Fund grant. The project does not necessarily reflect NBME policy, and support provides no official endorsement.
An earlier version of this paper is presented at the Northeast Group on Educational Affairs, Philadelphia, PA, USA, March 3-4 with the title: Expert-novice differences in data-gathering behavior during diagnoses.

Reference
Wimmers, P. F., & Kanter, S. L. (under review, 2006). Effects of level of expertise on data-gathering behavior during different stages of the diagnostic process.

Abstract

BACKGROUND Research in medical problem solving has been plagued by inconsistent findings about data-gathering behavior. There is uncertainty about the amount of information needed to mentally represent and solve a diagnostic problem. Some relate increasing expertise with being more efficient in data gathering and collecting less information, while others relate expertise with spending more time and selecting more information.

PURPOSE This study is an attempt to explain these seemingly contradictory findings and in particular, to discover how the amount of patient data gathered is related to the degree of experience of the participants.

METHODS Five computer-based cases were used to trace data-gathering behavior and its consistency among six stages of the diagnostic process (history, physical examination, laboratory tests, imaging, procedures, and consultations) for four different expertise groups.

RESULTS As expertise increases, participants select more information in the history and physical examination stages of the diagnostic process and less information in the laboratory, imaging, procedure, and consultation stages.

CONCLUSIONS Data-gathering behavior could be both a consistent and a reliable measure for performance during different stages of the diagnostic process. The amount of data gathered seems a more objective measure of expertise than accuracy of the diagnostic outcome. Data-gathering behavior can therefore function as an indicator of level of experience and case difficulty. A thorough understanding of data-gathering behavior helps understand the problem-solving process as well as the nature of expertise, and has implications for how we design instructional activities to teach novices to become better problem solvers.

THE DIAGNOSTIC REASONING PROCESS of expert physicians has fascinated educators and researchers for a long time, and along the way, many theories have been proposed to explain its complexity. One of the important aspects of this complicated problem-solving process is concerned with data collection from and about the patient. Data collection or data gathering usually takes place in several successive stages with the history and physical examination being the initial stages. During this cognitive process, diagnostic hypotheses are generated and evaluated against the newly requested and assessed information. During the latter stages of data gathering, the physician will select laboratory tests and/or imaging procedures to reject or confirm the generated diagnostic hypotheses (Elstein et al., 1978; Kassirer & Gorry, 1978). The physician's ability to reach an accurate diagnosis is related to efficiently collecting, integrating, and processing this patient information.

Pioneering research in the 1960s to explore the processes involved in medical problem solving was conducted by Horacio Rimoldi who developed clinical cases that consisted of a brief statement of a clinical problem and a set of 50-80 cards with questions on the visible side and answers on the other side (Rimoldi, 1955, 1960, 1961). The questions on the cards represented the successive stages of the diagnostic workup: history, physical examination, and laboratory procedures. When participants of different levels of expertise were instructed to select those questions that they thought would lead most directly to a solution of the problem, he found that with increasing levels of expertise fewer cards were requested. These differences between expertise groups were strongest for the interview and history stage. The number of questions requested for the physical examination stage showed a similar but less strong pattern. Laboratory procedures showed, however, almost no differences. The main conclusions of his work were that (1) since experts have greater certainty about the diagnostic outcome, they require less information than novices, and (2) experts are more efficient in gathering data with high information value early in the diagnostic process (Rimoldi, 1955, 1960, 1961; Rimoldi & Raimondo, 1998).

Later research into clinical reasoning conducted in the 1970s by Elstein, Shulman, and Sprafka (1978) found that high-scoring fourth-year medical students (i.e., students with more expertise than their peers) tended to collect more data than low-scoring students in order to develop an appropriate mental representation prior to generating a correct diagnosis. Low-scoring students, on the other hand, did not give much attention to the data they were gathering and apparently came to an (inaccurate) diagnosis too early. Elstein et al., (1978) concluded that high-scoring students were generally more efficient in data gathering, particularly in gathering cues likely to lead to a correct diagnosis. Conversely, they also found that the nature of the problem has a strong influence on how it is solved. The magnitude of the different process measures used was not consistent across problems.

Despite the inconsistency across problems of Elstein's findings, evidence from other fields than medicine indicates, in line with Elstein's findings, that this initial process of generating a good problem representation takes more time for experienced participants and more information is needed to shape further the mental representation. Chi et al., (1981) for example, showed with the use of physics problems that experts spend more time than non-experts on constructing good mental representations of problems. In addition, a recent study of Van Gog et al., (2005) showed with use of electrical circuits that higher expertise participants not only spend relatively more time during problem orientation but also focused more often on relevant information units in comparison to lower level expertise participants.

The question of interest is how we can understand and explain the seemingly contradictory findings. It is reasonable to conclude that data gathering plays an important role in establishing an appropriate mental representation and in generating diagnostic hypotheses during the patient workup. However, the amount of problem information required during this diagnostic process is not well established. From one perspective, expertise is related to being more efficient in data-gathering behavior, i.e., asking fewer questions during the initial stages of the diagnostic process (see the above mentioned studies of Rimoldi, 1955, 1960, 1961), and from the other perspective, expertise is related to spending more time and selecting more information in the initial problem solving process, as primarily found in other fields of expertise than medicine (Chi et al., 1981; Van Gog et al., 2005).

This study is an attempt to explain these contradictory findings and will in particular consider how the amount of patient-data gathered differs during subsequent stages (i.e., history, physical examination, laboratory tests, imaging, procedures, and consultations) of the diagnostic problem-solving process and how it is related to the degree of experience of the participants. In order to trace the process of data-gathering more precisely, patient cases were divided into smaller information units and in more specific stages, whereby participants of four different levels of expertise could select one information unit at the time. The number of information units selected and the consistency of this process between analogous stages across five different cases from the same discipline will be explored and discussed.

Methods

Participants

Eighty-one participants of four different expertise levels participated in this study. Expertise groups were defined based on their length of experience in medicine. The first group consisted of 30 undergraduate students with very little or no knowledge of medicine; the second group consisted of 33 fourth-year medical students; the third group consisted of nine pediatric residents; and the fourth group consisted of eight pediatric faculty physicians with several years of clinical experience. Participants received token compensation for completing all cases. Data for this study comes from a more comprehensive research project on case-based simulations conducted at the University of Pittsburgh School of Medicine.

Material

Five text-based simulated patient cases were presented on a computer screen. All cases were based on real patients from pediatrics. The correct diagnoses are listed in Table 1. The format was identical for all cases. The patient profile of each case contained a short description of the patient and his/her chief complaint (e.g., “An 8 month old white male infant is brought to your office for irritability and not feeding for one day. An hour ago, his mother noticed that he felt warm, and decided to bring him in.”). Further patient information was presented in the following main menus: history, physical examination, laboratory tests, imaging, procedures, and consultations. The user could access the information by selecting a main menu (i.e., “clicking” on it with the computer mouse). The main menu could contain information units and other menus items. For example, clicking on the main menu “history,” opened a subset of units witch contained information about history of present illness, past medical history menu, family history, social history, additional history, post-operative history, and post-partum history. In turn, the submenu item “past medical history” contained units with information about review of systems, early childhood history, hospitalizations, and allergies. Every main menu and its nested submenus were present in all cases whether they contained useful information or not. The seven main menus contained a total of 11, 17, 87, 21, 18, and 9 information units, respectively. For example, clicking on information unit CBC (complete blood count) within the hematology submenu of the laboratory tests menu provided the following information: WBC (white blood cell) 20.4 K/cu mm, Hematocrit 30%, Platelets 250 K/cu mm.

In addition to the menus, the user could access normal values, add a comment, or provide diagnoses at any time via “buttons” available at the top of the computer screen. A sample “screen capture” is provided in Appendix A.

Table 1. *Correct diagnoses associated with the five pediatric cases*

Case	Diagnoses
1	Haemophilus influenzae meningitis
2	Vitamin A intoxication
3	Cystic fibrosis
4	Tetralogy of fallot
5	Migraine

Procedure

Participants sat at a computer terminal with access to a keyboard and mouse. Although there were no time constrains on the procedure, participants were instructed to work through the cases as quickly as possible, taking into consideration the usual tradeoffs of cost and time efficiency versus safety and effectiveness for the simulated patient. Clicking on a particular main menu automatically opened a list with units that contained further information of the patient (as described above). It was possible to navigate back and forth between the menus and submenus. For each user, each item selected was recorded along with a date and time stamp. The presentation sequence of the cases was fixed.

Analysis

For each participant, counts were recorded for the total number of items selected as well as for the number of items selected within each main menu (i.e., history, physical examination, laboratory test, imaging, procedures, and consultations). Analogous main menu scores, as total scores of all menus were compared for consistency across cases. Different menu combinations were used to analyze and compare group differences across cases on information-gathering behavior. The main menus were vertically listed in a fixed order starting with history and ending with consultations.

Each user's diagnosis was scored for correctness according to the following scale: two points for a correct diagnosis, 1 point for a partially correct diagnosis, and 0 points for an incorrect diagnosis whereby only one score was given per user per case. For example, a correct diagnosis for case 1 of "Haemophilus influenzae meningitis" was awarded 2 points. One point was given for the diagnosis of "bacterial meningitis." Diagnoses like "sepsis" or "fever" received no points.

SPSS 14.0 for Windows is used for all analyses. Data were analyzed using repeated measures (ANOVA). Fisher's least significant difference (LSD) was used to make multiple pair-wise comparisons between the different expertise groups. Intra- and intercase correlations for the individual menus were calculated. Reliability analysis (Cronbach's alpha) was applied to analogous menus (e.g., history, physical examination) across cases. An alpha level of .05 was used for all statistical analyses.

Results

Of the 81 participants, two were removed partly from analyses, a fourth-year medical student who did not complete the last case and a physician who stopped after finishing the first case.

Diagnostic accuracy

Mean diagnostic accuracy increased significantly with levels of expertise, F (3, 75) = 69.124, $p < .0005$. The main effect for case, F (4, 296) = 72.979, $p < .0005$, and the group by case interaction were also significant, F (12, 296) = 8.304, $p < .0005$. Pair-wise comparisons indicated that mean diagnostic accuracy only differed significantly between the group of undergraduate students and all the other expertise groups. Based on level of diagnostic accuracy there can be concluded that cases differ in difficulty. For example, the Migraine case 5 was diagnosed correctly by almost all participants, while almost none of the participants had diagnosed the Vitamin A intoxication case 2 correctly. Average intercase correlation of diagnostic accuracy for the whole group was .27 (range from -.09 to .66). The intern al consistency (Cronbach's alpha) of the diagnostic outcome of the 5 cases was .65. Level of diagnostic accuracy is not very consistent across cases; this supports the differences in difficulty. Figure 1 shows diagnostic accuracy as a function of level of expertise for all cases. Fourth-year medical students were familiar with the cases from second-year coursework, while other participants were not.

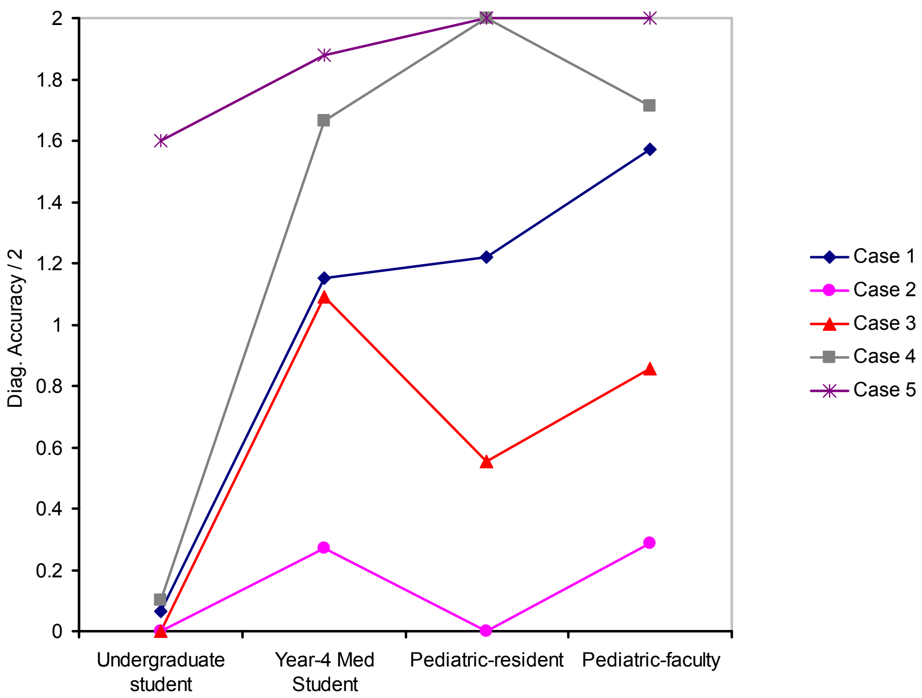


Figure 1. *Diagnostic accuracy as a function of level of expertise for all cases*

Processing time

Analysis of total mean processing time showed that the main effect for level of expertise was not significant, F (3, 74) = 1.775, $p = .159$. The main effect for case was significant, F (4, 296) = 93.081, $p < .0005$, and the group by case interaction was significant, F (12, 296)

= 6.038, $p < .0005$. Pair-wise comparisons of the cases revealed that the difficult case 2 (Vitamin A intoxication) shows a deviant pattern in comparison with the other cases (which show all a declining slope, see Figure 2). Rerunning the statistics without this case showed that the main effect for level of expertise was significant, F (3, 74) = 6.206, p = .001. Furthermore, the main effect for case, F (3, 222) = 83.558, $p < .0005$, and the group by case interaction, F (9, 222) = 2.762, p = .004, were both significant. Pair-wise comparisons revealed that mean processing time only differed significantly between the group of undergraduate students and the fourth-year medical students, and residents. Individual analyses of the history, physical examination and laboratory test menus yielded no significant main effects. However, it is interesting to note that the expert-physicians group took more time (16 %) in the history menu but less time (66 %) in the laboratory tests menu in comparison with novices.

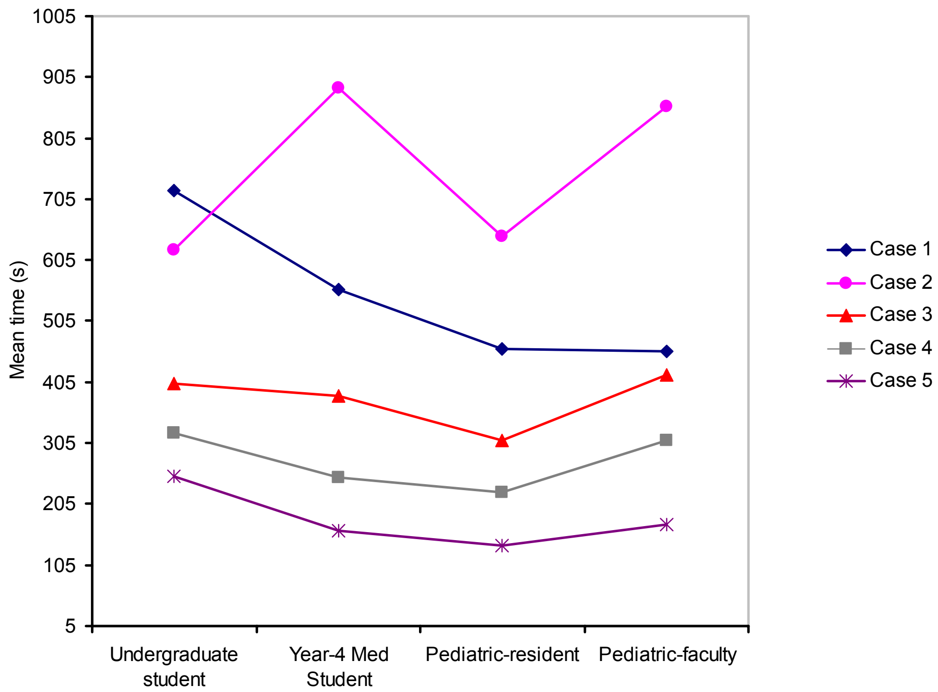


Figure 2. *Processing time as a function of level of expertise for all cases*

Total number of selected information units

Analysis of the total number of selected units of all menus together did not yield a significant main effect for level of expertise, F (3, 74) = .815, p = .489, and no group by case interaction. The main effect for case was significant, F (4, 296) = 57.810, $p < .0005$, indicating that total number of selected units per case is different. Mean number of total selected units for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were, 41.90, 36.91, 37.31, and 38.17, respectively. Average intercase correlation of total number of selected units of the 5 cases was .49 (range from .32 to .67). The internal consistency (Cronbach's alpha) was .83, meaning that participants are firmly consistent in the number of information units selected across the 5 pediatric cases (.70 is often used as an acceptable cutoff value).

History

The history menu showed monotonically increasing linear trends for all cases in number of selected units with increasing levels of expertise with no significant group by case interaction. This main effect of group was, however, not significant, $F(3, 74) = 2.112$, $p = .106$. The main effect of cases was significant, $F(4, 300) = 8.975$, $p < .0005$. Mean number of selected units in the history menu for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were, 8.22, 9.04, 8.98, and 1.11, respectively. Average intercase correlation of total number of selected units of the 5 cases was .41 (range from .15 to .70), which corresponds with an internal consistency of .78. Participants are firmly consistent across different cases on selecting units within the history menu.

Physical examination

The physical examination menu also showed monotonically increasing linear effects in number of selected units with increasing levels of expertise for all cases. The main effect of level of expertise was significant, $F(3, 74) = 3.132$, $p = .031$, as the main effect of cases, $F(4, 296) = 18.400$, $p < .0005$, and the group by case interaction, $F(12, 300) = 2.107$, $p = .017$. Pair-wise comparisons revealed that faculty physicians selected significantly more units then undergraduate students, or fourth-year medical students. Mean number of selected units in the physical examination menu for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 10.89, 11.89, 12.22, and 14.60, respectively. Average intercase correlation of total number of selected units of the 5 cases was .38 (range from .16 to .52). This yields an internal consistency of .75, indicating that participants are also consistent on the amount of information selected in the physical examination menu across cases.

Laboratory tests

In contrast to the former menus of the diagnostic process (history and physical examination), a decreasing number of selected units are associated with an increasing level of expertise in the laboratory test menus for all cases. The laboratory tests menu showed a significant main effect for level of expertise, $F(3, 74) = 3.308$, $p = .025$, a significant main effect for case, $F(4, 296) = 25.643$, $p < .0005$, and no significant group by case interaction. Pair-wise comparisons revealed that the undergraduate students selected significantly more information of the laboratory tests menus then the other groups. Mean number of selected units in the laboratory tests menu for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 10.51, 5.94, 7.16, and 5.46, respectively. Average intercase correlation of total number of selected units of the laboratory menu was .44 (range from .18 to .66). The internal consistency was .80 and this indicates that participants are firmly consistent in gathering information of laboratory tests across cases.

Imaging

The imaging menu showed a decreasing trend in number of selected units with increasing levels of expertise for all cases. However, the main effect of level of expertise was not significant, F (3, 74) = 1.907, p = .136, as the group by case interaction. The main effect of cases was significant, F (4, 296) = 12.338, p < .0005. Mean number of selected units in the imaging menu for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 3.38, 2.34, 2.98, and 2.51, respectively. Average intercase correlation of total number of selected units of the 5 cases was .28 (range from .13 to .48). This corresponds with an internal consistency of .67: an acceptable reliability.

Procedures

The procedures menu showed, like the results of the laboratory tests and imaging menus, a decreasing trend in number of selected units with increasing levels of expertise for all cases. The main effect of level of expertise was marginally significant, F (3, 74) = 2.718, p = .051. The main effect of cases was significant, F (4, 296) = 18.245, p < .0005, and there was no group by case interaction. Mean number of selected units in the procedures menu for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 1.25, .84, 1.00, and .77, respectively. Average intercase correlation of total number of selected units of the 5 cases was .13 (range from -.04 to .31). This corresponds with an internal consistency of .42, what is a low reliability.

Consultations

The consultations menu showed similar patterns; a decreasing number of selected units are associated with increasing levels of expertise. A significant main effect of level of expertise, F (3, 75) = 9.627, p < .0005, a significant case effect, F (4, 300) = 19.504, p < .0005, and a significant group by case interaction, F (12, 300) = 2.725, p = .002. However, 47 percent of the participants did not use this menu at all.

Taken together, the findings indicate that although there are significant case differences, the overall pattern is the same: increasing levels of expertise are related to selecting more information in the first two menus of the diagnostic process: history and physical examination, and less information in the latter four menus: laboratory tests, imaging, procedures, and consultations. Furthermore, the observed intercase correlations with accompanying reliability estimates across analogous menus of the different cases indicate that the participants are firmly consistent in data-gathering. Data-gathering behavior, during the subsequent stages, seems to make a shift between physical examination and laboratory tests menu. We therefore defined an initial stage of the diagnostic process consisting of history and physical examination, and a latter stage consisting of laboratory tests, imaging, procedures, and consultations. Both stages, the initial and latter, will be analyzed and compared in the following two sections.

Initial stage of the diagnostic process

The combination of both menus, history and physical examination, showed a significant main effect for level of expertise, F (1, 74) = 3.11, p = .031, a significant main effect for cases, F (4, 300) = 17.33, $p < .0005$, and no significant group by case interaction. Pair-wise comparisons revealed that faculty physicians selected significantly more units then undergraduate students, or fourth-year medical students. Mean number of selected units for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 19.13, 20.93, 21.20, and 24.71, respectively. Average intercase correlation of total number of selected units was .44 (range from .14 to .66) and the internal consistency estimate is .80. In the initial stage of the diagnostic process, an increasing number of selected units is related to an increasing level of expertise. Number of selected units in the initial stage of the diagnostic process as a function of level of expertise for all cases is shown in Figure 3.

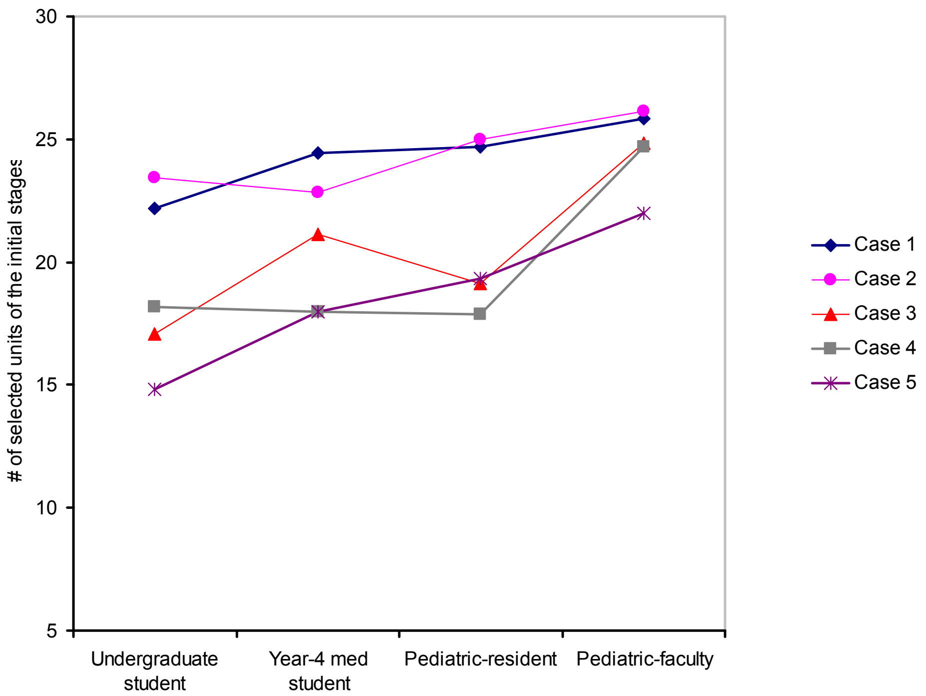


Figure 3. *Number of selected units of the initial stages (history and physical examination) as a function of level of expertise for all cases*

The later stage of the diagnostic process

Combining the latter menus, laboratory tests, imaging, procedures, and consultations, revealed a significant main effect for level of expertise, F (1, 74) = 4.388, p = .007, a significant main effect for cases, F (4, 300) = 28.272, p < .0005, and no significant group by case interaction. Pair-wise comparisons revealed that faculty physicians selected significantly less units then undergraduate students, and undergraduate selected significantly more units then fourth-year medical students and faculty physicians. Mean number of selected units for the 4 groups—undergraduate students, fourth-year medical students, residents, and faculty physicians—were 17.24, 9.81, 12.42, and 9.94, respectively. Average intercase correlation of total number of selected units was .52 (range from .29 to .69) and the internal consistency .84. In the later stage of the diagnostic process, a decreasing number of selected units is associated with an increasing level of expertise. Number of selected units in the later stage of the diagnostic process as a function of level of expertise for all cases is shown in Figure 4.

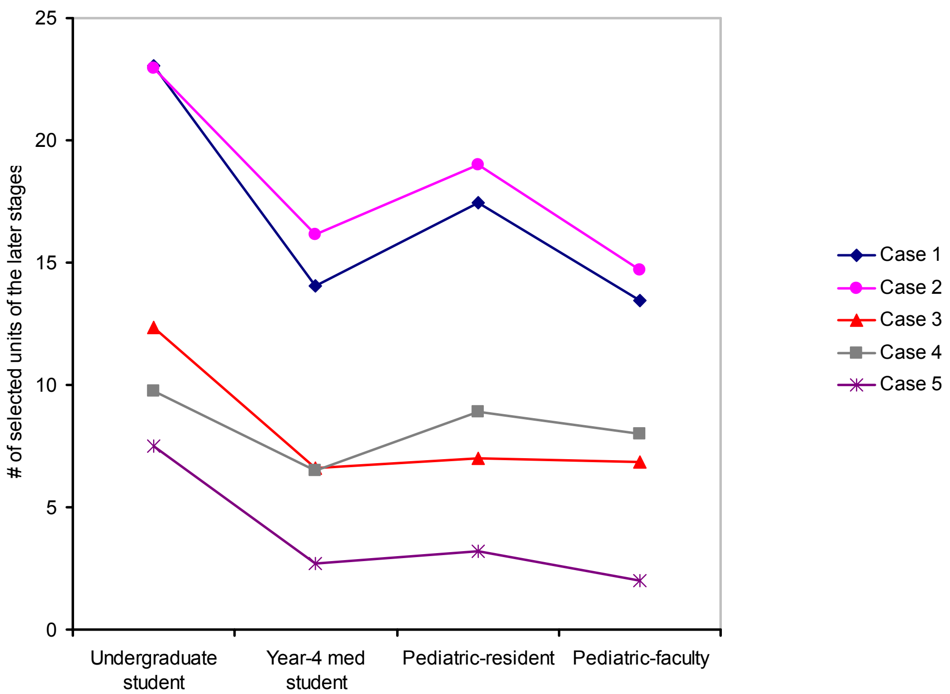


Figure 4. *Number of selected units as a function of the later stages of the diagnostic process (laboratory test, imaging, and procedures) as a function of level of expertise for all cases*

Discussion

The present study is an attempt to understand better the diagnostic reasoning process by examining differences in the ways experts and novices gather information by explicitly tracking the units of information selected at successive stages of the diagnostic process accessed in five computer-based patient-case simulations by participants at four different levels of expertise. More specifically, the number of information units selected, and the consistency of these selections among analogous stages (i.e., history, physical examination, laboratory test, imaging, procedures, and consultations) across cases were explored and analyzed.

The main findings can be summarized as follows: (1) diagnostic accuracy increased with increasing levels of expertise, (2) the number of information units selected is dependent on the particular stage of the case and is a function of level of expertise, (3) the number of information units selected is highly consistent among the various stages across cases, and (4) case difficulty seems to affect the amount of data-gathering.

The total number of units selected seems to be an insufficient measure for distinguishing levels of expertise as evidenced by a lack of a significant difference among the four groups of participants studied. However, a closer look at the individual stages of the diagnostic process revealed why. With increasing expertise, participants selected more information in the initial stages (history and physical examination) and less information in the later stages of the diagnostic process (laboratory test, imaging, and procedures) than less experienced participants. With increased medical training, an important change occurs in diagnostic ability characterized by a change in the selection of information during the various stages of problem solving.

These findings seem to disagree with the original Rimoldi studies which indicate that experts selected fewer “question cards” than novices (Rimoldi, 1955, 1960, 1961). Rimoldi’s results could be interpreted to indicate that the highly organized and easily activated knowledge structures of experts allow them to be more efficient in asking questions and processing these. An important difference in design between our study and that of Rimoldi’s was that the cases used in the present study consisted of 164 information units while the Rimoldi cases consisted of 50-80 questions. It could be that the ‘right’ questions (asked by the experts) contained more relevant patient information or at least, gave more robust cues. In one of Rimoldi’s later papers, while summarizing his own work of the last four decades, he stated that: “experts will be better at selecting data with high potential to reduce uncertainty.” The study of Van Gog et al., (2005) for example, showed that more experienced participants focused more often on relevant information units of a problem than did less experienced participants—the information units that obviously gave helpful cues for diagnosis. A study of Coughlin and Patel (1987) showed, with the use of free recall, that physicians indeed used significantly more critical cues from clinical problems than medical students. Selecting more information units early in the problem-solving process seems to be a strategy to reduce the problem space by eliminating a great number of (incorrect) explanations or diagnostic hypotheses. In other words, a physician uses the information of the case to narrow the range of possible disorders.

The importance of the history and the physical examination to the diagnostic process deserves special emphasis. The usefulness of history and physical examination may be dependent on the medical discipline, however, most clinicians would confirm the primary role of those stages in arriving at a correct diagnosis. A study by Peterson and colleagues of 80 internal medicine patients, for example, showed evidence that a thorough history led to the final diagnosis for 76% of the cases, whereas physical examination and laboratory data contributed for only 12% and 11%, respectively (Peterson, Holbrook, Hales, Smith, & Staker, 1992). A study of Schmitt et al., (1986) showed that history-based diagnoses could predict final diagnoses of 146 patients with dyspnea 74 % of the time. Furthermore, Sills (1978) stated that history and physical examination are the most valuable diagnostic tools, and laboratory data provided hardly any diagnostic valuable information and was useful in only 1.4% of 186 pediatric “failure-to-thrive” cases. In our study, experts indeed seem to support this notion in their diagnostic process by spending relatively more time than novices to collect more information during history and physical examination in contrast to laboratory data. Nevertheless, teaching history-taking skills may not get the attention it deserves. Many students do not improve their interviewing skills through medical school, (see for example, Hasnain, Bordage, Connell, & Sinacore, 2001) and Pfeiffer and colleagues proved that these interviewing skills of students even decline during the last two clinical years of medical school (Pfeiffer, Madray, Ardolino, & Willms, 1998).

Our results not only indicate that expertise can be characterized, in part, by process indicators of problem solving as indicated by differences in information-gathering behavior among groups at different levels of expertise, but also show that this behavior is notably consistent among the analogous stages across different cases (Cronbach’s alpha from .75 to .83). The consistencies across cases are remarkable, especially because the consistency of data-gathering behaviors are greater than of the diagnostic accuracy measure ($\alpha = .65$). Amount of data gathered may be a more objective measure of expertise than accuracy of the diagnostic outcome. Consistencies in data-gathering behavior across cases, however, are not well documented in the literature and most studies that advocate consistency in data gathering made use primarily of factor analyses (e.g., Berner et al., 1977; Harasym et al., 1980). Rimoldi (1963) was able to replicate previous results when the same participants were examined in two successive years, but he never mentioned consistencies across cases in a single study. Norman et al., (1982) compared performance of 10 residents on four real and four simulated patients and found no difference between real and simulated patients in the amount of data gathering in history taking and physical examination.

The primary goal of this study was not to analyze the structure of the mental representation of the problem solver, but to examine how collecting information differs among levels of expertise and among different stages of the diagnostic process. Although some inferences about the initial representation can be garnered from the processes studied. A computational view of cognition considers thinking as the manipulation of an internal representation (Hunt, 1989). If one considers problem solving to be a form of thinking, then two processes seem to interact during the first stages of the problem-solving process. First a process whereby information units in the case are used to build an initial mental representation of the problem and formulate diagnostic hypotheses, and second, a process whereby the potential solutions are narrowed down by the information given (Coughlin & Patel, 1987). It is suggested by the data in this study, that experts take more time with these processes and require more case information, especially during history taking and physical examination of the patient workup. The latter stages may be primarily used by experts to verify the most correct diagnostic hypothesis. Novices, on the other hand, miss knowledge to elaborate upon the information given during history and physical examination and continue to gather information during subsequent stages of the cases: laboratory tests, imaging, procedures, and consultations.

The observations of the current study indicate that data-gathering behavior could be both a consistent and a reliable measure for performance during the diagnostic process. Furthermore, there is a relationship between case difficulty and amount of data gathered. Specifically, in comparing the relatively simple migraine case with the much more difficult Vitamin A intoxication case, it could be concluded that a higher level of diagnostic accuracy is related to faster processing times and to less data gathering. Our process indicator of performance is related to level of experience and dependent on level of case difficulty. Without a doubt, the nature of the problem seems to have a strong effect on how it is approached and processed (Elstein et al., 1978). There is evidence that case processing of experts changed under the demands of the task, like an unusual case format or being forced to pay attention to the information units of the case (Norman et al., 1989; Wimmers et al., 2005). Consequently, "diagnostic skill implies more than the mere accumulation of factual knowledge, (Rimoldi, 1963);" it implies, in addition, efficiently collecting, integrating, and processing patient information.

Some limitations of this study should be considered. All cases came from the same clinical discipline (i.e., pediatrics). Thus the consistency in data-gathering behavior found in this study may not generalize across other clinical domains. In addition, differences in data-gathering behavior as found across cases and that were addressed to difficulty differences could be (at least partly) related to the fixed order of the presentation of the cases. It is reasonable to assume that the participants have to get used to the computer set up and design of the cases.

Despite the fact that many stages of the diagnostic process showed clear trends it did not always translate to statistical group differences. The group sizes of the residents and faculty physicians were relatively small and responsible for the lack of statistical power. Furthermore, as mentioned before, the fourth-year medical students saw the cases before during second-year course work. The higher diagnostic accuracy of this group could be based on recognition. On the other hand, this confounder in our study raises an interesting question for further study: Why might prior exposure to the same patient cases affect diagnostic accuracy but not data-gathering behavior?

This study generates additional interesting implications for future research. Does the amount of information collected serve a role in generating a more elaborate knowledge representation or only to eliminate solutions no longer deemed relevant? And how do those two processes interact with each other? Combining data-gathering analysis with commonly used methods such as free recall or verbal think-aloud protocols could be of assistance. Would experts' free recall outperform students' free recall of information in the first stages of the case in contrast to latter stages? An analysis of think-aloud protocols, comparing and contrasting different stages of the case might be worth considering. From an educational perspective, the question of to what extent the knowledge representation and/or diagnostic outcome would change if students were 'forced' to spend more time on history and physical examination of a case would be worth considering. Finally, the observation that changes in data-gathering behaviors that take place as a function of the demand of the task (level of case difficulty) may be an indicator for a shift in processing needs further clarification. Elstein et al., (1978) already noticed that the nature of the task had a strong effect on how it is approached, and called this phenomenon "content-specificity." However, it could be possible, like Alexander and Judy (1988) proposed in their review article, that experts are weighting their content knowledge against the demands of the task and then bring in their appropriate form of strategic knowledge to perform on that task, hence the variation in several process indicators, like hypotheses generation and data-gathering behavior.

References

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

Berner, E. S., Bligh, T. J., & Guerin, R. O. (1977). An indication for a process dimension in medical problem solving. *Medical Education, 11*, 324-328.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Medical Education, 62*, 818-828.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. London, Cambridge, Mass: Harvard University Press.

Harasym, P., Baumber, J., Bryant, H., Fundytus, D., Preshaw, R., Watanabe, M., et al. (1980). An evaluation of the clinical problem-solving process using a simulation technique. *Medical Education, 14*(6), 381-386.

Hasnain, M., Bordage, G., Connell, K. J., & Sinacore, J. M. (2001). History-taking behaviors associated with diagnostic competence of clerks: An exploratory study. *Academic Medicine, 76*(Suppl 10), S14-S17.

Hunt, E. (1989). Cognitive Science: Definition, Status, and Questions. *Annual Review of Psychology, 40*, 603-629.

Kassirer, J. P., & Gorry, A. G. (1978). Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine, 89*, 245-255.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by experts medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology; Learning, Memory and Cognition, 15*, 1166-1174.

Norman, G. R., Tugwell, J. P., & Feighter, J. W. (1982). A comparison of resident performance on real and simulated patients. *Journal of Medical Education, 57*, 708-715.

Peterson, M. C., Holbrook, J. H., Hales, D., Smith, N. L., & Staker, L. V. (1992). Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Western Journal of Medicine, 156*, 163-165.

Pfeiffer, C., Madray, H., Ardolino, A., & Willms, J. (1998). The rise and fall of students' skill in obtaining a medical history. *Medical Education, 32*(3), 283-288.

Rimoldi, H. J. A. (1955). A technique for the study of problem solving. *Educational and Psychological Measurement, 15*, 450-461.

Rimoldi, H. J. A. (1960). Problem solving as a process. *Educational and Psychological Measurement, 20*(3), 449-460.

Rimoldi, H. J. A. (1961). The test of diagnostic skills. *Journal of Medical Education, 36*(1), 73-79.

Rimoldi, H. J. A. (1963). Evaluation and training in clinical diagnostic skills. *Medical and Biological Illustrations, 13*, 186-191.

Rimoldi, H. J. A., & Raimondo, R. (1998). Assessing the Process of Clinical Problem Solving. *Advances in Health Sciences Education, 3*(3), 217.

Schmitt, B. P., Kushner, M. S., & Wiener, S. L. (1986). The diagnostic usefulness of the histroy of the patient with dyspnea. *Journal of General Internal Medicine, 1*, 386-393.

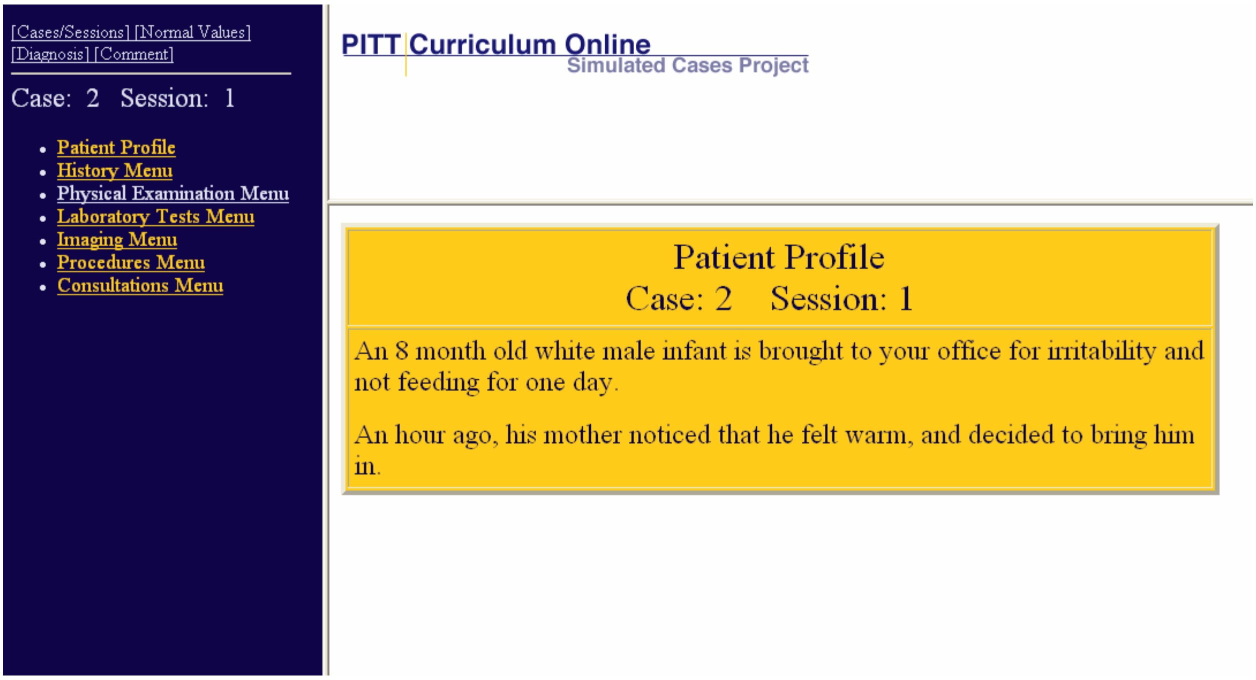
Sills, R. H. (1978). Failure to Thrive: The role of clinical and laboratory evaluation. *Am J Dis Child, 132*, 967-969.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205-221.

Wimmers, P. F., Schmidt, H. G., Verkoeijen, P. P. J. L., & van de Wiel, M. W. J. (2005). Inducing expertise effects in clinical case recall. *Medical Education, 39*(9), 949-957.

Appendix A

An example of a short patient description and the menu layout of a computer case



CHAPTER EIGHT

Summary and Conclusions

THE DEVELOPMENT OF clinical competence is the main purpose of medical education. The long road to become clinically competent starts on the first day of medical school, and every institution strives to select the best students. The responsibility of medical schools is to train those students to become clinically competent physicians who are highly qualified and specialized professionals able to function in a constantly changing society that puts continuously growing demands on the medical profession. Understanding clinical competence is therefore crucial not only for medical education, assessment, and licensing examinations, but also for society and its responsibility for the quality of health care. However, there is little consensus about what clinical competence is and how to measure it. A proper definition and a better understanding of clinical competence and its components would serve as a criterion for validating medical educational programs and would assure a minimum level of competency at the end of medical school and beyond during residency.

Each of the previous chapters of this thesis highlights different aspects of clinical competence and its development. In the introductory chapter the lack of consistency in definitions of clinical competence is emphasized and a historical overview is given to shed light on the variety of definitions used over the last decades. The study presented in Chapter 2 explores the effect of level of pre-university performance on preclinical and clinical performance. The study in Chapter 3 addresses the impact of clinical training during clerkships on students' learning. Specifically, the relationship among the nature and volume of patient encounters, quality of supervision, and the learning outcomes were explored. The study of Chapter 4 is directed at the interaction between knowledge and a problem-solving ability during students' diagnostic problem solving at end-of-clerkship examinations. Chapter 5 focuses on clinical competence during clerkships and end-of-clerkship examinations. In this study we explored whether clinical teachers place different values on individual components of competence while students work on the wards or while students are examined during clerkship examinations. The study reported in Chapter 6 investigates how recall of case information changed for different expertise groups under different conditions, while research discussed in Chapter 7 was conducted to find out how the selection of case information by students and doctors during different stages of the diagnostic process is related to different expertise groups. In the present and final chapter of this thesis the main findings of all previous chapters are summarized and discussed. At the end of this chapter, conclusions and suggestions for further research are given.

Summary of the main findings

Part I: Determinants of clinical competence development

During medical school, students go through several stages of development (Boshuizen, 2005; Schmidt & Boshuizen, 1992). The first preclinical phase of medical education emphasizes the acquisition of basic biomedical science knowledge, which serves as a solid foundation for the clinical phase. In this later phase the emphasis shifts towards practice and the application of knowledge in a clinical setting. Learning in a clinical environment is centered on real patient encounters, and a wider set of characteristics are needed and developed during this period of medical education. The study reported in Chapter 2 focuses on the transitions from public preparatory school to medical school and within medical school from preclinical education to clinical education. Of particular interest is whether students are progressing at a consistent level throughout medical school, independent of the phase transitions of the curriculum, and whether this consistency is related to their level of performance before medical school. For this purpose, medical students were classified into three groups—low, intermediate, and high achievers—based on level of performance before entering medical school. Preclinical and clinical performance were determined by students' mean grades over all course examinations within that phase. Performance levels before entering medical school were used as predictors for success in medical school for the whole cohort and the different grade-subgroups, respectively. For the whole cohort the results indicated that level of performance before entering medical school was highly related to level of preclinical performance but much less strong related to level of clinical performance. However, the individual grade-subgroups showed different patterns. For the high-achieving students, medical school performance of both phases, preclinical and clinical, was highly predictable and this subgroup progressed at a consistent performance level throughout medical school. They seemed better equipped for curriculum transitions than low-achieving students. Performance of low-achieving students was least consistent and hence almost not predictable at all. Performance of the intermediate grade subgroup was predictable only during preclinical years, not for the clinical years. It could be concluded that prior knowledge, as represented by level of pre-university performance, may be a necessary prerequisite of clinical competence, but it is far from sufficient. Therefore, perhaps non-cognitive variables, like a professional attitude, interpersonal skills, or communication skills are better indicators of performance in medical school, especially as early indicators for clinical performance (Albanese, Snow, Skochelak, Hugett, & Farell, 2003; Moulaert, Verwijnen, Rikers, & Scherpbier, 2004; Stern, Frohna, & Gruppen, 2005; Webb et al., 1997).

Preclinical education reflects domain knowledge, cognitive abilities, and to some extent, diagnostic problem solving. Clinical education, on the other hand, places more emphasis on diagnostic problem-solving skills and integrates the experiences of the preclinical years with clinical practice and encountering real patients. It is generally assumed that a sufficient number of patient encounters is essential for the development of clinical competence (Neufeld & Norman, 1985; Snell, Battles, Bedford, & Washington, 1998; Witzke, Koff, McGeagh, & Skinner, 1990). Studies that examined the relative growth in knowledge during clerkships reported significant increases (Butterfield & Libertin, 1993; Schwartz, Donnelly, Sloan, & Young, 1994). However, many previous studies were not able to identify a direct relationship between the number of clinical encounters and growing clinical competence (Châtenay et al., 1996; Gruppen, Wisdom, Anderson, & Woolliscroft, 1993; McManus, Richards, Winder, & Sproston, 1998; Van Leeuwen et al., 1997).

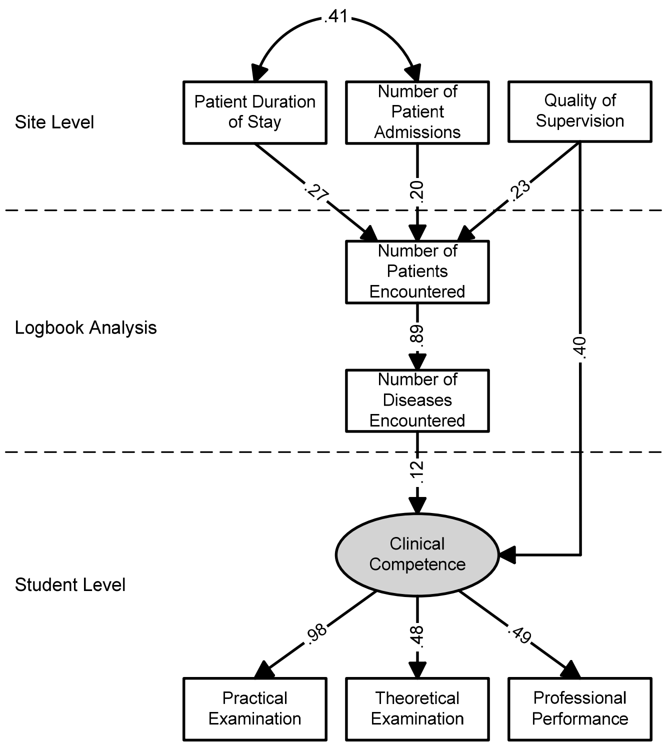


Figure 1. *Clerkship Competency Model*

The study reported in Chapter 3 focuses on learning in the clinical setting. It addresses in particular the impact of patient encounters during clerkships on the development of clinical competence. Clerkship students at several internal medicine sites recorded their patient encounters in logbooks and evaluated the quality of supervision they received. Student competence, the outcome measure, was determined by three independent indicators: practical end-of-clerkship examination, theoretical end-of-clerkship examination, and evaluation of professional performance of the students. Site characteristics that might influence the variation in patient encounters were collected. Analysis of the logbooks confirmed that differences between hospital sites are bigger than expected given the differences within sites. Hospital sites characteristics account for a large amount of the variation in the number of patients encountered by students during their clerkships. The variation in the number of patients seen, however, did not directly affect the development of clinical competence. A finding of particular interest is, as our clerkship competency model (Figure 1) showed, that the quality of the supervision in clinical education counts more than the numbers of patients seen or the variety of diseases encountered. It appears that the clinical supervisor has a stimulating effect on student learning and student learning environment, resulting in more patient encounters. A long period of supervised training is crucial to gradually acquire an increased responsibility in treating patients (Ericsson, 2004; Ericsson & Charness, 1994).

Part II: The nature of clinical competence

Clinical competence in general and diagnostic problem solving in particular appear out to be dependent on either the content or context of the problem (e.g., Elstein, Shulman, & Sprafka, 1978). This discovery in the late seventies had a huge impact on medical education, expertise research, and assessment. It was a largely laboratory-based finding and the question we posed ourselves therefore was: To what extent does this conclusion hold in the practice of clinical training. The study in Chapter 4 contrasts the finding of content specificity in competence with the existence of a general clinical problem-solving ability in end-of-clerkship examinations.

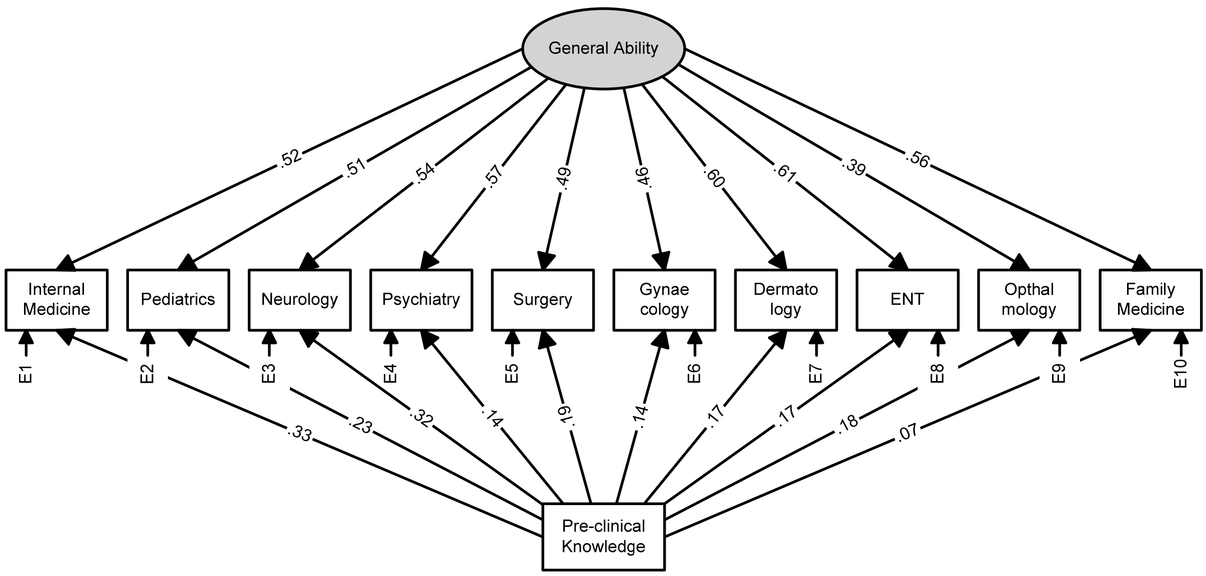


Figure 2. *The one-factor model with preclinical knowledge directly influencing clerkship final scores*

The cohort studied consisted of medical students who successfully completed their preclinical years at medical school and rotated through all clerkships. A correlation matrix was calculated with clerkship final scores from 10 disciplines to examine the magnitude of the interrelations. A confirmatory factor analysis was applied to the corresponding covariances using structural equation modeling to investigate whether scores on finals shared any common variance across clerkships. In the first modeling step of the study three models were analyzed and compared: (1) a general factor model, (2) an independence model, and (3) a combined general/specific factor model. In the next modeling step, two additional models were tested to examine the nature of the relationship between preclinical knowledge and problem solving. In testing and comparing the first two models, we concluded that neither a pure content-specificity assumption nor a pure general-ability assumption could explain the data sufficiently. The performances on clerkship finals were neither fully independent, as would be expected if performance on each final was purely dependent on specific knowledge of that clerkship, nor could a single underlying factor completely explain performances on all clerkship finals, as would be expected if a general ability was responsible for performances on clerkship finals. It seemed that a combination model represented clerkship performance best. In this model clerkship-specific knowledge and a single underlying factor were both related to clinical problem solving. In an attempt to understand these findings in more depth, the second modeling step was taken to determine the nature of the relationship between preclinical knowledge and this general ability. A model relating the preclinical knowledge indicator directly to each of the clerkships was superior (Figure 2). In this particular model, both specific subject-matter preclinical knowledge and an underlying general ability influenced performance on the clerkship finals independently.

So, some aspects of problem-solving in a clinical context seemed, unlike what the early findings of Elstein et al., (1978) suggest, to have general, consistent characteristics across different problems. For example, in the study reported in Chapter 7 we were able to show, with the use of computer-based clinical cases, high levels of consistency across cases in information-gathering behavior in similar stages of the diagnostic process (i.e., history taking, physical examination, laboratory tests, imaging, and procedures). Alexander and Judy (1988) suggest that during problem solving, competent learners weigh their content knowledge against the demands of the problem and then bring in their appropriate form of strategic knowledge in order to solve the problem. It seems logical to conclude that strategies can be executed only in relation to domain-specific knowledge. However, when domain-specific knowledge develops, domain-specific problem solving will become easier and therefore less dependent on general strategies or abilities. Diagnostic problem-solving contains both specific aspects and general aspects, which explain the findings discussed in Chapter 4. For example, diagnoses can be considered a highly specific aspect of problem solving while history taking can be considered a more general skill, as indeed is more often suggested in the literature (Berner, Bligh, & Guerin, 1977; Donnelly, Gallagher, Hess, & Hogan, 1974; Juul, Noe, & Nerenberg, 1979).

Chapter 5 reported the results of a survey given to clinical teachers from the various medical specialties. The survey was intended to explore what competencies were considered important during adequate student performance on the wards and to what extent the same competencies were assessed at examinations. Thus, each individual competency was addressed twice resulting in two separate data sets. Two independent factor analyses were conducted using SEM software to uncover underlying latent relationships among the different competencies for what was considered important for performance on the wards and what was considered important during clinical examinations.

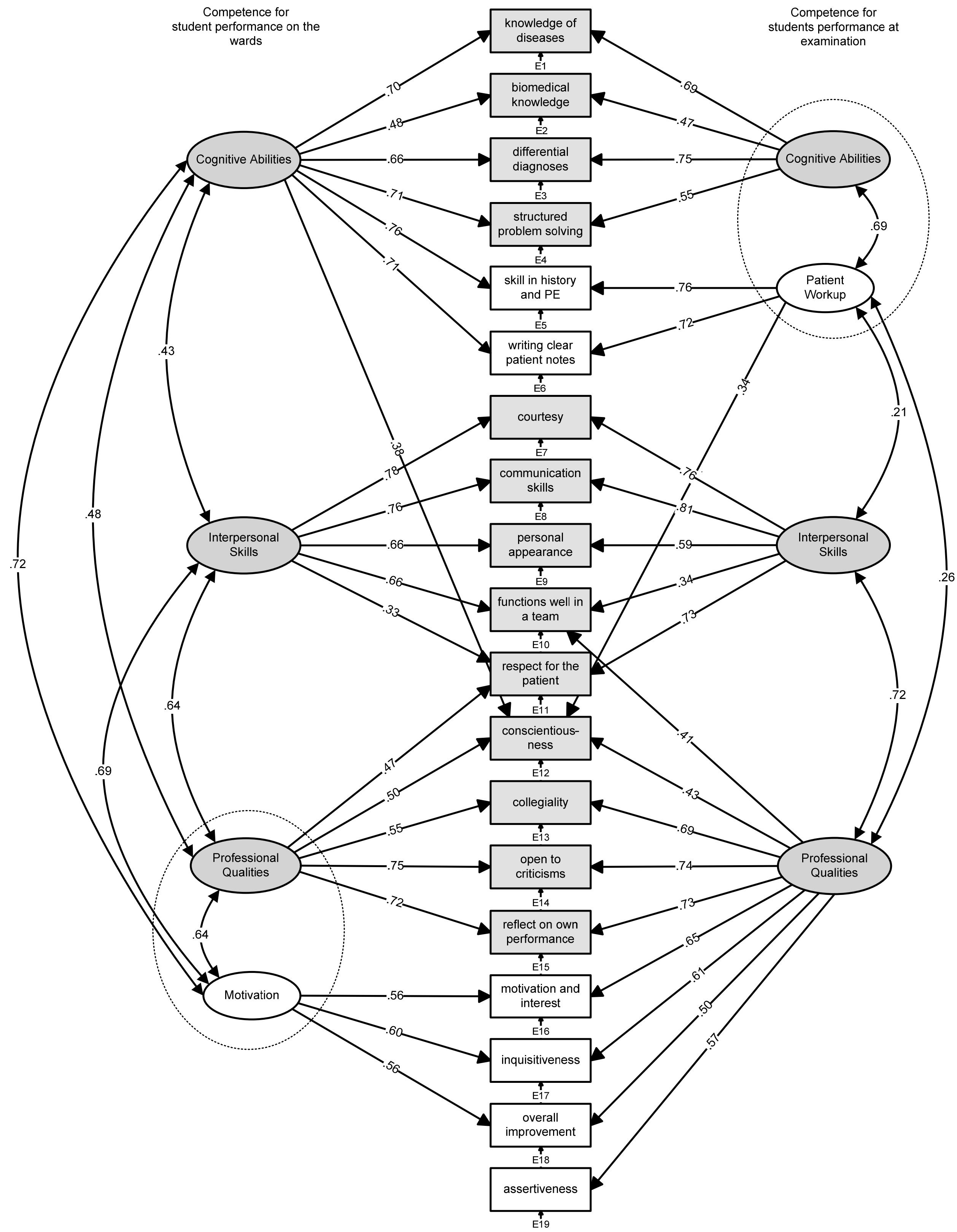


Figure 3. *Conceptualized model of the two four-factor models of the components of competence for student performance on the wards and for student performance at examinations*

Factor analysis revealed a four-factor structure for both data sets (see Figure 3). Comparing the two sets revealed that three of the four factors showed almost identical item structures between both factor models. These three components of clinical competence were in line with the classical categorization into cognitive abilities, interpersonal skills, and professional qualities (Epstein & Hundert, 2002; Forsythe, McGaghie, & Friedman, 1986). However, the results also indicated that different competencies seemed important for student performance while being examined as indicated by level of ranked importance of the individual components of competence and their underlying factor structure. Cognitive abilities were valued as relatively more important for student performance at examinations and interpersonal skills and professional qualities were seen as relatively more important for student performance on the wards. Further differences between the two models seemed to be in the fourth factor in each model: a “motivational” factor during ward performance and a “patient workup” factor for performance on student examination. Student examination seemed to be centered on cognitive abilities and diagnostic problem solving, which were only slightly related to the other factors: interpersonal skills and professional qualities. Competence as perceived during performance on the wards, on the other hand, showed moderate to high correlations among all factors. It was concluded that, based on the perception of clinical educators, what seems important for adequate performance on the ward is not necessarily in alignment with what is required for examination.

Part III: The development of clinical competence

Expertise studies in medical education research serve two main goals. First, understanding how experts solve problems provides information about the representation of knowledge, in particular how their knowledge is organized and structured. Second, comparing experts’ problem-solving capability with that of novices gives information about how expertise develops (Chi, 2006). Medical experts seem, while processing a case, to chunk relevant information into higher level summarizing concepts (or “encapsulations”). When asked to recall a case, experts tend, in contrast to less experienced participants, to recall these condensed concepts rather then the “raw” information. Therefore, less experienced participants seemed to recall more case information after processing a case than experts (or novices): this phenomenon has been termed the “intermediate effect” (Schmidt & Boshuizen, 1992, 1993a/b; Schmidt, Boshuizen, & Hobus, 1988). The study in Chapter 6 was conducted to explore expertise differences in the amount of clinical case recall. For this purpose, participants from three different levels of expertise (i.e., preclinical medical students, clinical students, and expert nephrologists) were divided into three different conditions. In the first condition, cases consisted of a context, history and physical examination, and laboratory data; in the second condition, the context was removed and the cases consisted of laboratory data only; in the third condition, instructions to elaborate on the cases was added to the laboratory data only cases. It was shown that the amount of case information recalled by experts could be influenced by requiring them to elaborate on each of the information units presented in the case. In that particular condition, case recall showed a monotonically increasing effect with increasing levels of expertise instead of an intermediate effect. In addition, removing the context lowered the diagnostic accuracy, but did not affect free recall. It was concluded that instructions to elaborate in combination with laboratory data only cases seem to have induced a deeper, more detailed, analysis of the patient case. Expertise effects in clinical case recall were produced only when the normal processing of patient information was disrupted. Case processing by experts can change under the demands of the task, like by an unusual case format or being forced to pay attention to the information units of the case by giving elaborative instructions. This study uses case recall as research method to contribute to the understanding of how knowledge is mentally represented and organized and how knowledge of experts differs from that of novices.

Another way to get insight into how knowledge is mentally represented is by observing how information is gathered from a computer-based case. Data collection or gathering normally takes place in several successive stages; the history and physical examination are the initial stages. On the basis of this information, a mental representation of the patient's problem is constructed. This process is dependent on his/her scientific and clinical knowledge and previous experience with patients. At the same time, during this cognitive process, a multitude of diagnostic hypotheses are generated and evaluated against new requested and assessed information (Elstein et al., 1978; Kassirer & Gorry, 1978). The study discussed in Chapter 7 investigated expertise differences in data-gathering behavior. In order to trace the process of data-gathering more precisely, patient cases were divided into smaller information units and into more specific stages (i.e., history, physical examination, laboratory tests, imaging, procedures, and consultations); participants of four different levels of expertise (i.e., undergraduate students, fourth-year medical students, pediatric residents, and pediatric faculty physicians) participated in this study. The number of information units selected and the consistency of this process between analogous stages across five cases were explored, and we found that experts spent more time and selected more information during history and physical examination and less information in the later stages of the diagnostic process (laboratory tests, imaging, etc) than the less experienced participants. It could be concluded that selecting more information in the initial stages of problem solving by experts seemed a strategy to reduce the problem space by eliminating a great number of (incorrect) explanations or diagnostic hypotheses. In other words, an experienced physician uses the information of the case to narrow the range of possible disorders. Taken together, two processes seemed to interact during the first stages of the diagnostic process; first, a process whereby information units in the case were used to build an initial mental representation of the problem and formulate diagnostic hypotheses, and second, a process whereby the mental representation was narrowed down by the information given (Coughlin & Patel, 1987). The latter stages (laboratory test, imaging, procedures) were primarily used to verify the diagnostic hypothesis. The less experienced participants, on the other hand, missed the necessary background knowledge to make use of the information given during history and physical examination and continued to gather information during subsequent stages of the cases: laboratory tests, imaging, procedures, and consultations. Another remarkable finding was the consistency of this behavior across five different cases. History taking and physical examination were the most general aspects of the diagnostic process and generating diagnoses the most content (or case) specific.

Conclusions

The main conclusions in this thesis are summarized here. The first conclusion pertains to the role of knowledge in medical education. In Chapter 2 we saw that for the complete cohort the correlation between pre-university performance and preclinical performance was high (r = .54), but the correlation between pre-university performance and clinical performance was much lower (r = .25). In Chapter 4 we saw that the average correlation between preclinical knowledge and performance on individual clerkship finals of 10 different disciplines is .19 (it ranged from .07 to .33). This suggests that the relationship between the preclinical knowledge variable and each of the clerkship finals is low to moderate, and preclinical knowledge plays an inconsistent role with regard to performance in each of the clerkship finals. In fact, the influence of preclinical knowledge seems to decrease over time, as the correlations seem to drop for the later clerkships. Assessment of undergraduate medical students during the preclinical phase has focused mostly on recall of factual knowledge and the application of this knowledge in problem solving. However, clinical competence is more than a process of extending causal knowledge about a domain. And even though there is enough evidence that pre-admission grades (that represented level of prior knowledge) are insufficient for predicting clinical performance, they are still the most frequently used prerequisites for admission to medical school. In addition, examinations based on recall of factual knowledge and the application of this knowledge may fail to document what students will do when faced with a real patient.

A second conclusion pertains to the importance of seeing patients in a clinical setting. The patient encounter is thought to be crucial during clinical education. There is ample evidence about what is exactly learned and how it is learned during clerkships, but more specifically, no previous studies have found a *direct* relationship between the number of clinical encounters during rotations and clinical competence (Châtenay et al., 1996; Gruppen et al., 1993; McManus et al., 1998; Van Leeuwen et al., 1997). The overall assumption that sufficient exposure to patients will in itself lead to clinical competence is not justified by our data (Chapter 3). The number of diseases seen (which is highly related with the number of patients encountered) had a small but nonsignificant influence on clinical competence (identified by practical and theoretical end-of-clerkship examinations, and professional performance), while the quality of supervision had a correlation (i.e., standardized regression coefficient) of .40 with clinical competence and a correlation of .23 with number of patient encountered by students. Clinical supervisors seem to have a stimulating effect on students' learning and students' learning environment, resulting in better performance. The quality of supervision could therefore compensate for a small number of patients (Dolmans, Wolfhagen, Essed, Scherpbier, & Van der Vleuten, 2002); proper guidance and feedback are stimulating factors for learning to occur.

A third conclusion is related to the finding of a general problem-solving ability during clinical problem solving. The study presented in Chapter 4 showes that for performance on clinical examinations a combination of both is needed: content-specific knowledge and general problem-solving skill. Chapter 7 specifies further the generalizability of some aspects of the diagnostic process. The quality of data-gathering behavior seems a consistent aspect of diagnostic problem solving across cases, in contrast to the level of diagnostic accuracy that is more related to a particular case. We reasoned that problem solving is an interaction between content or domain-specific knowledge and general abilities and takes place on a continuum between both. Problem-solving ability is dependent on case-difficulty and level of experience of the participants. The use of general processes may be more important when an individual is confronted with difficult problems as found in fields other than medicine (Glaser, 1984). If problems are too easy, then they may be little more than recall tasks for experts, and one would guess that problem-solving processes would not be generated (see for example, Schuwirth, Verheggen, van der Vleuten, Boshuizen, & Dinant, 2001).

A fourth conclusion is related to the results of the study discussed in Chapter 5. What is important for adequate student performance on the wards is, according to the perceptions of clinical educators, considered less important for examinations. For daily performance of the student on the wards, characteristics like motivation, interpersonal skills, and other professional qualities (e.g., collegiality, working in teams, respect for the patient, communication skills) are more important than cognitive abilities (e.g., knowledge of diseases, problem-solving skills, history taking and physical examination), while for clerkship examination the opposite is true. If we take into consideration that proper assessment is one of the most difficult aspects of clinical education and that faculty skills in assessment methods are often lacking, then every attempt to make more reliable and valid measurement methods will have less chance of succeeding if clinical educators lack the necessary skills or think differently about student examination in contrast to what is required for adequate daily performance on the wards. Taking into account the effects of assessment on learning, this phenomenon could jeopardize the development of clinical competence.

A final concluding remark relates to the importance of history-taking and physical examination skills. The teaching of history-taking and physical examination skills may not always get the attention it deserves (Hasnain, Bordage, Connell, & Sinacore, 2001; Pfeiffer, Madray, Ardolino, & Willms, 1998). Clinical educators very rarely directly observe students taking a history or doing a physical examination on the wards (Pulito, Donelly, Plymale, & Mentzer Jr, 2006). In Chapter 5 we saw that interpersonal skills and professional qualities are considered relatively more important for student performance on the wards than the patient workup, which is relatively more important for student performance at examinations. Nevertheless, the study discussed in Chapter 7 shows that experts are spending relatively more time than novices to collect more information during history and physical examination in contrast to laboratory data. Thus for experts, history and physical examination are considered valuable diagnostic tools. Several studies confirm the importance of history and physical examination. A study by Schmitt et al., (1986), for example, showed that history-based diagnoses could predict final diagnoses 74 % of the time. Laboratory data, on the other hand, provided hardly any diagnostic valuable information (Sills, 1978). The importance of the history and the physical examination to the diagnostic process needs special attention early in the medical curriculum, and those skills needs to be properly observed by supervisors during clinical clerkships.

Suggestions for further research

The studies presented in this thesis lead to several suggestions for further research. In Chapter 2 we saw that subgroups based on differences in pre-admission grades behave differently throughout medical school. Probably, factors other than cognitive abilities are playing an important role during the development of competence within medical school. Therefore, it would be of interest to focus research on which factors or combination of factors could be used to define groups of students with specific behaviors and study patterns. We could, for example, think of personality characteristics that contribute to clinically competent behavior, like emotional stability, empathic ability, adaptability, ability to be self-critical, or conscientiousness (Chapter 5). These personality characteristics might be particularly important in the later clinical years of medical school and residency. We saw in Chapter 3 that simple exposure to patients is not enough to acquire the essential competencies. A combination of factors related to the learning environment seems important (e.g., patient encounters in combination with proper feedback). Of interest is what and how do students learn from patient encounters, and more specifically, what other aspects of clinical supervision are needed for an optimal learning effect? More information is also needed about which other components of the hospital environment contribute to the development of students' clinical competence.

Another finding with consequences for medical education research is the content-specificity of problem solving. We saw in Chapter 4 that performance seems to contain both a general element, related to a problem-solving ability, and specific elements, related to domain knowledge. Clinical problem solving is probably an interaction between a general ability and content knowledge depending on the level of experience of the student, level of difficulty of what is assessed, and which aspect(s) of the diagnostic process or clinical competence is emphasized. Research should be focused on the critical analysis of the interaction between domain-specific knowledge and domain-general strategies or abilities. The models tested in Chapter 4 could be applied to participants with different levels of expertise. Doing so could verify whether an increased level of experience is related to a decreased application of a general problem-solving ability. Furthermore, it would be important to provide participants with problems from different levels of difficulty because it is reasonable to assume that if content knowledge is sufficient to solve a problem, the need for applying general strategies decreases.

Studies emphasizing processing differences between novices and experts have given important insights into the development of competence. In a laboratory setting many kinds of task variations and manipulations of conditions and instructions are possible. In Chapter 6, manipulation of case format and instruction had an effect on the way experts process a case—a strong argument in favor of the knowledge encapsulation theory because only the experts are considered to have developed sufficient encapsulations to deal with everyday cases, and therefore, only their performance is disrupted by the particular treatment. This disruption also manifested itself in the processing speed of the doctors (but not of the students); processing of the laboratory data in the laboratory-data-plus-elaboration condition took almost three times longer than processing the same data in the clinical context condition. One possibility to verify this disruption-of-processing hypothesis is to require participants to elaborate on cases for which intermediate effects have been demonstrated in the past (Patel & Groen, 1991; Schmidt & Boshuizen, 1993b). The prediction would be that expertise effects in recall might be induced for these cases as well. A second possibility would be to ask participants to undertake additional and unrelated tasks while processing a case. This would limit working memory capacity and might also disrupt the skill of expert doctors to automatically translate sets of signs and symptoms into their underlying encapsulations.

Data-gathering behavior, as described in the study of Chapter 7, generated additional interesting suggestions for future research. This study looked at data-gathering differences between novices and experts during different stages of the diagnostic process. Experts in contrast to novices collected more information in the initial stages of processing (i.e., history and physical examination). This phenomenon suggests two possible explanations that are both testable: The amount of information collected serves a role in generating a more elaborate knowledge representation, or it only eliminates solutions no longer deemed relevant. Of particular interest is how those two processes interact with each other. From an educational perspective, the question of to what extent the knowledge representation and/or diagnostic outcome would change if students were 'forced' to spend more time on history and physical examination of a case would be worth considering.

Combining the methodologies used in the studies of Chapter 6 and 7 may lead to new insights. The study in chapter 6 made use of "free recall," while the study in chapter 7 attempted to "track" processing directly. Both methods have advantages and disadvantages. Direct tracking of data-gathering behavior could be combined with indirect methods such as free recall or verbal think-aloud protocols. Could we expect, in this particular case, that experts' free recall would outperform students' free recall of information in the first stages of the case in contrast to later stages? An analysis of think-aloud protocols, comparing and contrasting different stages of the case might be worth considering. The findings of Chapter 5, 6, and 7 have in common that case processing changes under the demands of the task (or demands of the task in combination with instructions). A further specification of what particular demands change case processing is needed. Whether level of case difficulty might be an indicator for a shift in processing needs more clarification.

Chapter five highlights the perceived difference (according clinical educators) between what competencies are considered important for adequate performance during clerkships and what are required to do well on examinations. Cognitive abilities were relatively more important for student performance at examinations and interpersonal skills and professional qualities were relatively more important for student performance on the wards. This difference, for example, could refer to a more commonly addressed problem that clinical educators rarely observe students taking a history or doing a physical examination on the wards. Future research should focus on an explanation for this discrepancy in perception and what the consequences are of this discrepancy in perception for clinical education and assessment. A start can be made by using a qualitative approach to look more deeply into the perceptions of clinical educators.

References

Albanese, M. A., Snow, M. H., Skochelak, S. E., Hugett, K. N., & Farell, P. M. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine, 78*(3), 313-321.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

Berner, E. S., Bligh, T. J., & Guerin, R. O. (1977). An indication for a process dimension in medical problem solving. *Medical Education, 11*, 324-328.

Boshuizen, H. P. A. (2005). Does practice make perfect? A slow and discontinues process. In H. P. A. Boshuizen, R. Bromme & H. Gruber (Eds.), *Professional Learning: Gaps and Transitions on the way from Novice to Expert* (Vol. 2). Dordrecht / Boston / London: Kluwer Academic Publishers.

Butterfield, P. S., & Libertin, A. G. (1993). Learning outcomes of an ambulatory care rotation in internal medicine for junior medical students. *Journal of General Internal Medicine, 8*(4), 189-192.

Châtenay, M., Maguire, T., Skakun, E., Chang, G., Cook, D., & Warnock, G. L. (1996). Does the volume of clinical experiences affect performance of clinical clerks on surgery exit examinations? *The American Journal of Surgery, 172*(4), 366-372.

Chi, M. T. H. (2006). Methods to access the representation of experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance*: Cambridge University Press.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Medical Education, 62*, 818-828.

Dolmans, D. H. J. M., Wolfhagen, I. H. A. P., Essed, G. G. M., Scherpbier, A. J. J. A., & Van der Vleuten, C. P. M. (2002). The impacts of supervision, patient mix and numbers of students on the effectiveness of clinical rotations. *Academic Medicine, 77*(4), 332-335.

Donnelly, M. B., Gallagher, R. E., Hess, J. W., & Hogan, M. J. (1974). *The dimensionality of measures derived from complex clinical simulations.* Paper presented at the Thirteenth Annual Conference on Research in Medical Education, Chicago, IL, US.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning*. London, Cambridge, Mass: Harvard University Press.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA, 287*(2), 226-235.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10 Suppl), S70-S81.

Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychology, 49*(8), 725-747.

Forsythe, G. B., McGaghie, W. C., & Friedman, C. P. (1986). Construct validity of medical clinical competence measures: a Multitrait-Multimethod Matrix study using confirmatory factor analysis. *American Educational Research Journal 23*(2), 315-336.

Glaser, R. (1984). Education and Thinking: The role of knowledge. *American Psychologist, 39*, 93-104.

Gruppen, L. D., Wisdom, K., Anderson, D. S., & Woolliscroft, J. O. (1993). Assessing the consistency and educational benefits of students' clinical experiences during an ambulatory care internal medicine rotation. *Academic Medicine, 68*(9), 674-680.

Hasnain, M., Bordage, G., Connell, K. J., & Sinacore, J. M. (2001). History-taking behaviors associated with diagnostic competence of clerks: an exploratory study. *Academic Medicine, 76*(Suppl 10), S14-S17.

Juul, D. H., Noe, M. J., & Nerenberg, R. L. (1979). A factor analytic study of branching patient management problems. *Medical Education., 13*(3), 199-205.

Kassirer, J. P., & Gorry, A. G. (1978). Clinical problem solving: a behavioral analysis. *Annals of Internal Medicine, 89*, 245-255.

McManus, I. C., Richards, P., Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ, 316*, 345-350.

Moulaert, V., Verwijnen, M. G. M., Rikers, R. M. J. P., & Scherpbier, A. J. J. A. (2004). The effects of deliberate practice in undergraduate medical education. *Medical Education, 18*, 1044-1052.

Neufeld, V. R., & Norman, G. R. (1985). *Assessing Clinical Competence*. New York: Springer-Verlag.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise, prospects and limits*. Cambridge: Cambridge University Press.

Pfeiffer, C., Madray, H., Ardolino, A., & Willms, J. (1998). The rise and fall of students' skill in obtaining a medical history. *Medical Education, 32*(3), 283-288.

Pulito, A. R., Donelly, M. B., Plymale, M., & Mentzer Jr, R. M. (2006). What do faculty observe of medical students' clinical performance. *Teaching and Learning in Medicine, 18*(2), 99-104.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York, NY: Springer-Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On Acquiring Expertise in Medicine. *Educational Psychological Review, 5*(3), 205-221.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*(3), 338-351.

Schmidt, H. G., Boshuizen, P. A., & Hobus, P. P. M. (1988, 17-19 August 1988). *Transitory stages in the development of medical expertise: The "intermediate effect" in clinical representations studies.* Paper presented at the Tenth Annual Conference of the Cognitive Science Society, Montreal, Quebec, Canada.

Schmitt, B. P., Kushner, M. S., & Wiener, S. L. (1986). The diagnostic usefulness of the histroy of the patient with dyspnea. *Journal of General Internal Medicine, 1*, 386-393.

Schuwirth, L. W. T., Verheggen, M. M., van der Vleuten, C. P. M., Boshuizen, H. P. A., & Dinant, G. J. (2001). Do short cases elicit different thinking processes than factual knowledge questions do? *Medical Education, 35*(4), 348-356.

Schwartz, R. W., Donnelly, M. B., Sloan, D. A., & Young, B. (1994). Knowledge gain in a problem-based surgery clerkship. *Academic Medicine, 69*, 148-151.

Sills, R. H. (1978). Failure to Thrive: The role of clinical and laboratory evaluation. *Am J Dis Child, 132*, 967-969.

Snell, L. M., Battles, J. B., Bedford, J. A., & Washington, E. T. (1998). Verifying the curriculum of a family medicine clerkship. *Medical Education, 32*, 370-375.

Van Leeuwen, Y. D., Düsman, H., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R. P. T. M., et al. (1997). Factors Influencing the Growth in Knowledge of Trainees in General Practice. *Advances in Health Sciences Education, 2*(1), 61-70.

Webb, C. T., Sedlacek, W., Cohen, D., Shields, P., Gracely, E., Hawkins, M., et al. (1997). The impact of nonacademic varables on performance at two medical schools. *Journal of the National Medical Association, 89*(3), 173-180.

Witzke, D. B., Koff, N. A., McGeagh, A. M., & Skinner, P. D. (1990). Developing a computer-based system to document and report students' clinical encounters. *Academic Medicine, 65*(7), 440-441.

Samenvatting [Summary in Dutch]

DE PRIMAIRE DOELSTELLING van medisch onderwijs is het opleiden van studenten tot klinische bekwaamheid of competentie. De lange weg naar klinische competentie begint voor de medische student op de eerste dag van de opleiding. Geneeskundeopleidingen staan daarna voor de uitdaging om studenten zich te laten ontwikkelen tot hoog gekwalificeerde en gespecialiseerde artsen die goed voorbereid zijn op een constant veranderende maatschappij die hoge eisen stelt aan de gezondheidszorg. Het concept "klinische competentie" is daarom niet alleen essentieel voor het definiëren van de eindtermen van de geneeskundeopleiding en de toetsing daarvan, maar ook voor de maatschappij en kwaliteit van de gezondheidszorg. Er is echter weinig consensus over de vraag wat klinische competentie is en hoe deze gemeten kan worden. Een heldere definitie en een beter inzicht in de betekenis van klinische competentie en haar componenten zouden als criterium voor het visiteren van medische onderwijsprogramma's kunnen dienen en als kwaliteitsbewaking voor een minimumniveau van competentie van de student aan het einde van de geneeskundeopleiding en de medische vervolgopleidingen.

De hoofdstukken van dit proefschrift benadrukken elk een verschillend aspect van klinische competentie en haar ontwikkeling. In het inleidende hoofdstuk wordt de verscheidenheid van in omloop zijnde definities beschreven vanuit een historisch perspectief. In hoofdstuk 2 wordt een studie besproken waarin het effect van prestatieniveaus van de VWO-vooropleiding op prestatieniveaus van preklinisch onderwijs en klinisch onderwijs werd onderzocht. De studie in hoofdstuk 3 gaat in op het effect van klinisch onderwijs tijdens co-schappen op de ontwikkeling van competentie bij studenten. Met name de relaties tussen het patiëntencontact, de kwaliteit van begeleiding en de leerresultaten werden onderzocht. In hoofdstuk 4 wordt een studie besproken die focuste op de interactie tussen kennis en medisch probleemoplossen tijdens klinische examens. Hoofdstuk 5 gaat verder in op klinische competentie tijdens het dagelijks functioneren in de co-schappen versus competentie tijdens expliciete toetsmomenten in de co-schappen. In deze studie werd onderzocht of componenten van klinische competentie een verschillende waarde krijgen toebedeeld tijdens beide genoemde situaties. In hoofdstuk 6 wordt een studie beschreven waarin groepen medici met een verschillend expertiseniveau de opdracht kregen om onder verschillende omstandigheden klinische casussen te diagnosticeren. Vervolgens werd hun gevraagd om op te schrijven wat zij zich nog van de casussen herinnerden. De studie in hoofdstuk 7, tot slot, beschrijft hoe verschillende deskundigheidsgroepen patiëntgegevens verzamelden tijdens het diagnosticeren. Met name de hoeveelheid benodigde informatie tijdens verschillende specifieke stappen in het diagnostische proces en het onderling verband daartussen, zoals de anamnese en het lichamelijk onderzoek, werden vergeleken.

In dit afsluitende hoofdstuk worden de belangrijkste bevindingen per deel samengevat en besproken.

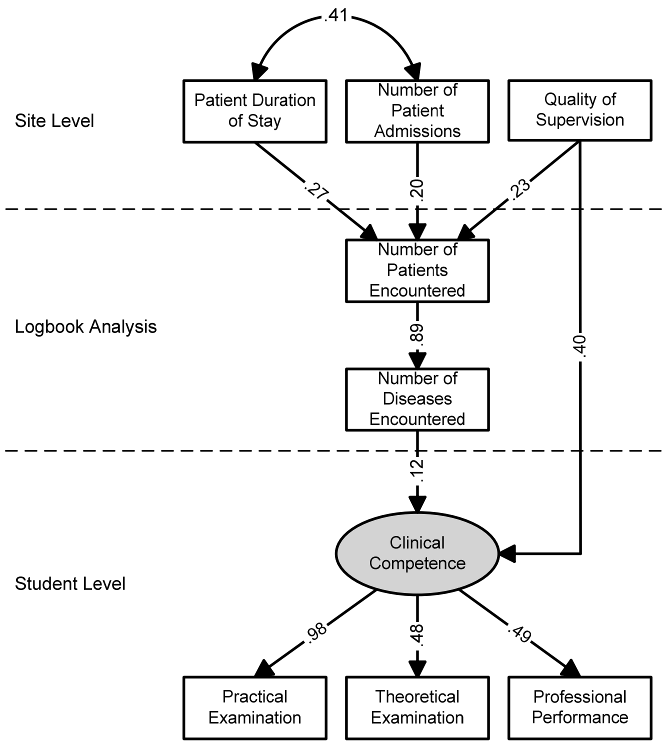
Samenvatting van de belangrijkste bevindingen

Deel I: Determinanten van klinische competentieontwikkeling

De studenten doorlopen tijdens hun geneeskundeopleiding verschillende stadia van ontwikkeling van medische kennis (Boshuizen, 2005; Schmidt & Boshuizen, 1992). De eerste, preklinische fase van het medisch curriculum benadrukt de ontwikkeling van fundamentele biomedische wetenschapskennis, die als basis dient voor de daaropvolgende klinische fase. Deze klinische fase vindt plaats in de praktijk waarin de student eerder opgedane kennis in een klinische leeromgeving leert toepassen. Het leren is nu gecentreerd om de 'echte' patiënt en een breder kennis- en vaardighedenpakket is nodig tijdens deze periode van het medisch curriculum (Metz, Verbeek-Weel, & Huisjes, 2001). De studie die in hoofdstuk 2 wordt beschreven, concentreerde zich op de overgangen van voorgezet wetenschappelijk onderwijs (VWO) naar de geneeskundeopleiding én binnen de geneeskundeopleiding van preklinisch onderwijs naar klinisch onderwijs. Van bijzonder belang is de vraag of studenten op een gelijk niveau door de faseovergangen van de medische opleiding stromen. Daarbij is ook de vraag van belang of de doorstroming afhankelijk is van hun VWO-prestaties. De medische studenten waren geclassificeerd in drie groepen op basis van hun VWO-prestaties (laag, midden en hoog prestatieniveau). In deze studie zijn VWO-prestaties gebruikt als voorspeller voor succes tijdens de geneeskunde opleiding voor, respectievelijk, het hele cohort en de verschillende prestatieniveau-subgroepen binnen dit cohort, waarbij preklinische en klinische prestatieniveaus werden bepaald door het gemiddelde cijfer tijdens deze fasen. Voor het hele cohort bleek dat de VWO-prestaties hoog gecorreleerd waren met de preklinische prestatieniveaus, maar veel zwakker gecorreleerd met de klinische prestatieniveaus. Een nadere analyse van de afzonderlijke subgroepen onthulde echter een ander beeld. Voor studenten uit de hoog prestatieniveaugroep bleek dat VWO-prestaties goede voorspellers waren voor zowel de preklinische als de klinische fase van de geneeskundeopleiding. Studieprestaties van de studenten uit de midden prestatieniveaugroep waren vanuit de VWO-prestaties alleen voorspelbaar voor de preklinische fase. Studieprestaties van de laag prestatieniveaugroep bleken echter voor beide fasen in het geheel niet voorspelbaar vanuit hun VWO-prestaties. Samenvattend kan worden gesteld dat de hoog prestatieniveaugroep consistent presteert in alle fasen van het onderwijs: in het VWO en in de preklinische en klinische fase. De laag prestatieniveaugroep presteert daarentegen instabiel en niet voorspelbaar. Studenten met hoge prestaties tijdens de vooropleiding blijken dus beter toegerust te zijn voor faseovergangen in het medisch curriculum dan studenten die lager presteren. Het kennisniveau zoals weergegeven door het prestatieniveau van de student tijdens de vooropleiding, is nog steeds een belangrijk selectiecriterium voor toelating tot de medische opleiding. Hoewel een bepaald kennisniveau een belangrijk fundament vormt voor de verdere ontwikkeling van klinische competentie is het echter, zoals blijkt uit bovengenoemde resultaten, een verre van afdoende voorspeller voor de prestaties tijdens de geneeskundeopleiding. Dit geldt in het bijzonder voor de klinische fase waarin, in vergelijking tot de preklinische fase, een groter beroep gedaan wordt op andere eigenschappen en vaardigheden van de student, zoals bijvoorbeeld een professionele houding, goede interpersoonlijke vaardigheden en communicatievaardigheden (Stern, Frohna, & Gruppen, 2005). Wellicht vormen deze niet-cognitieve variabelen betere indicatoren, vooral voor voorspelling van prestatieniveaus in de klinische praktijk. (Albanese, Snow, Skochelak, Hugett, & Farell, 2003; Moulaert, Verwijnen, Rikers, & Scherpbier, 2004; Webb et al., 1997).

Tijdens het preklinisch onderwijs bouwen studenten fundamentele biomedische wetenschapskennis op en ontwikkelen hun cognitieve capaciteiten en probleemoplossende vaardigheden. Hoewel veel geneeskundeopleidingen het belang inzien om studenten eerder in contact te brengen met de kliniek, geldt in het algemeen dat pas in het klinische onderwijs de nadruk wordt gelegd op het oplossen van 'echte' patiëntgerelateerde problemen en de integratie van eerder opgedane medische kennis en vaardigheden met de klinische praktijk. Het zien van een ruim aantal patiënten en hun klachten wordt essentieel geacht voor de ontwikkeling van klinische competentie (Neufeld & Norman, 1985; Snell, Battles, Bedford, & Washington, 1998; Witzke, Koff, McGeagh, & Skinner, 1990). Studies naar de relatieve groei van kennis tijdens de co-schappen tonen inderdaad significante verschillen aan (Butterfield & Libertin, 1993; Schwartz, Donnelly, Sloan, & Young, 1994). Veel studies waren echter niet in staat om een direct verband aan te tonen tussen het aantal patiëntcontacten en het niveau van het veelomvattender begrip klinische competentie (Châtenay et al., 1996; Gruppen, Wisdom, Anderson, & Woolliscroft, 1993; McManus, Richards, Winder, & Sproston, 1998; Van Leeuwen et al., 1997).

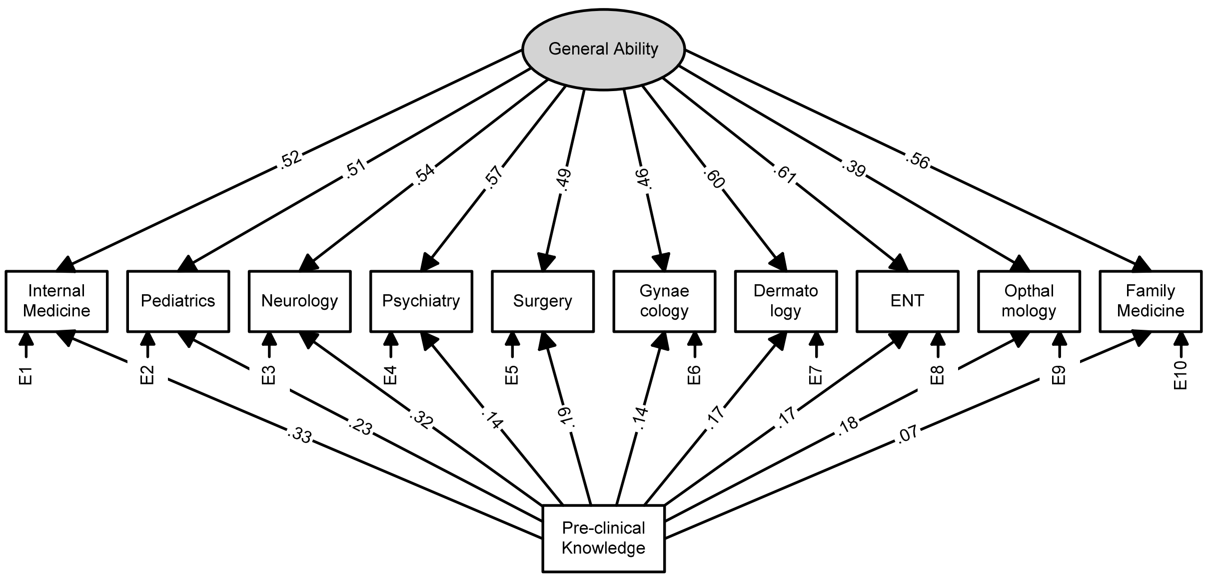
De studie die in hoofdstuk 3 wordt besproken, gaat in op het leren tijdens de klinische praktijk. In het bijzonder wordt aandacht geschonken aan het effect van het aantal patiëntcontacten van de co-assistenten op de ontwikkeling van klinische competentie. Bij verschillende ziekenhuizen registreerden co-assistenten tijdens een interne geneeskunde co-schap al hun patiëntcontacten in logboeken en evalueerden de door hun ervaren kwaliteit van de begeleiding tijdens dit co-schap. Het competentieniveau van de co-assistenten werd bepaald door drie onafhankelijke indicatoren: een theoretische en een praktische toets, aangevuld met een evaluatie van het professionele gedrag door de klinisch begeleider. Alle mogelijke ziekenhuisinformatie die van invloed zou kunnen zijn op de variatie in het aantal patiëntcontacten werd verzameld. De analyse van de logboeken bevestigde dat er grote verschillen waren tussen de ziekenhuislocaties. Ziekenhuisspecifieke kenmerken zijn dus deels verantwoordelijk voor de hoge variatie in het aantal patiënten dat de co-assistent ziet. Deze variatie beïnvloedde echter *niet* direct de ontwikkeling van klinische competentie. Het door ons voorgestelde competentiemodel (zie figuur 1) laat zien dat de kwaliteit van de begeleiding belangrijker is voor de ontwikkeling van klinische competentie dan het aantal patiënten dat een student tijdens deze periode ziet. De literatuur ondersteunt dat goede begeleiding en constructieve feedback essentieel blijken voor het verwerven van competentie (Ericsson, 2004; Ericsson & Charness, 1994).



Figuur 1. *Het klinische competentie model van de co-schappen*

Deel II: De aard van klinische competentie

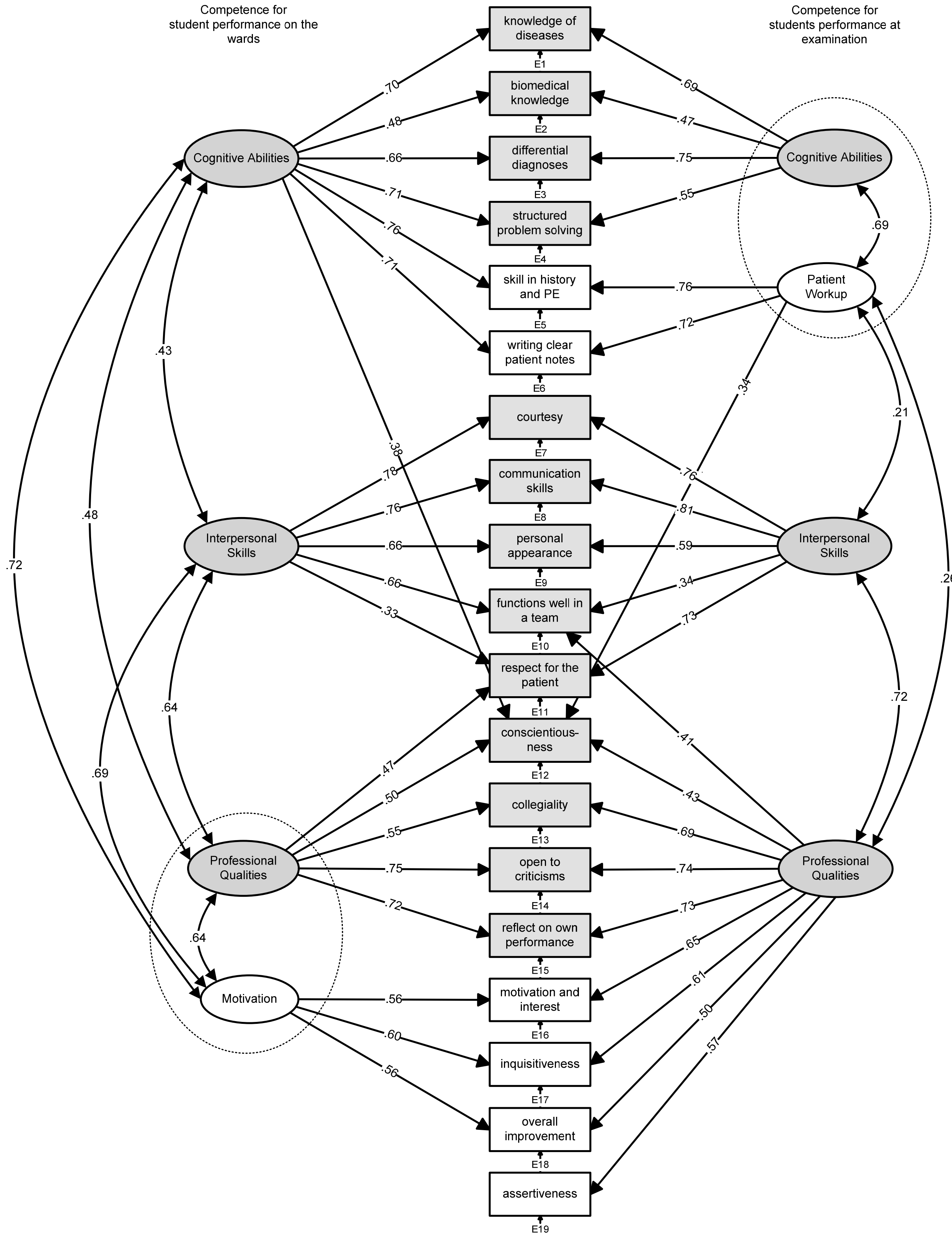
Klinische competentie in het algemeen en diagnosticeren in het bijzonder blijken afhankelijk te zijn van de inhoud of de context van het op te lossen probleem (e.g., Elstein, Shulman, & Sprafka, 1978). Deze bevinding, die "inhoudsspecificiteit" wordt genoemd, werd gedaan aan het einde van de jaren '70 en had een belangrijke invloed op het medisch onderwijs en de wijze van toetsing binnen dit onderwijs. Deze inhoudsspecificiteit was echter grotendeels gebaseerd op studies die plaatsvonden in een gecontroleerde omgeving. De vraag die daarom gesteld kan worden is in hoeverre deze conclusie ook houdbaar is in de praktijk van de medische opleiding. Deze vraag stond centraal in de studie die in hoofdstuk 4 wordt besproken. In deze studie werd inhoudsspecificiteit van competentie tegenover een algemeen klinisch probleemoplossende vaardigheid gesteld die zichtbaar werd in de afsluitende toetsen van de afzonderlijke co-schappen. Het cohort uit deze studie bestond uit medisch studenten die succesvol hun preklinische jaren voltooid hadden en vervolgens alle co-schappen hadden doorlopen. Om de samenhang tussen toetsresulaten van de tien co-schappen vanuit de verschillende medische disciplines te onderzoeken werd een correlatiematrix opgesteld. Vervolgens werd op de corresponderende covarianties een factoranalyse toegepast met behulp van "structural equation modeling" (SEM) software om te onderzoeken of de afsluitende toetsen gemeenschappelijke variantie deelden. In de eerste modelleringsvergelijkingsstap werden drie modellen voorgesteld en getest: (1) een één-factor model (representeert een algemeen probleemoplossende vaardigheidshypothese); (2) een onafhankelijk model (representeert een puur inhoudsspecificiteitshypothese); en (3) een gecombineerd algemeen/specifiek model. In de tweede modelleringsvergelijkingsstap werden twee structurele modellen vergeleken om de aard van de samenhang tussen preklinische kennis en een disciplineonafhankelijke, probleemoplossende vaardigheid te onderzoeken. Bij het testen en vergelijken van de eerste twee structurele modellen werd geconcludeerd dat noch de aanname van een zuivere inhoudsspecificiteit noch die van een zuiver algemeen probleemoplossende vaardigheid de toetsresultaten afdoende konden verklaren. De toetsprestaties waren enerzijds niet volledig onafhankelijk van specifieke kennis behorend bij dat betreffende co-schap. Anderzijds waren zij ook niet volledig onafhankelijk van één enkele onderliggende factor, zoals verwacht kan worden als uitgegaan wordt van een algemeen probleemoplossende vaardigheid die verantwoordelijk is voor goede toetsresultaten. Het door ons voorgesteld combinatiemodel bleek het beste de eindtoetsresultaten te representeren. In dit model zijn co-schapspecifieke kennis *en* één enkele onderliggende factor beide verantwoordelijk voor klinisch probleemoplossen. In de tweede modelleringsvergelijkingsstap genoot het model waarbij preklinische kennis de eindtoetsresultaten van de co-schappen rechtstreeks beïnvloedt de voorkeur boven het model waarbij preklinische kennis de eindtoetsresultaten indirect, via de factor, beïnvloedt. In het model van onze keuze beïnvloedden zowel specifieke kennis als een algemene factor onafhankelijk van elkaar de prestaties op de eindtoetsen (zie figuur 2).



Figuur 2. *Structureel model waarbij preklinische kennis de eindtoetsresultaten van de co-schappen rechtstreeks beïnvloedt*

Het blijkt dus dat sommige aspecten van probleemoplossen in een klinische context algemene vergelijkbare kenmerken bezitten, in tegenstelling tot de vroegere bevindingen van Elstein et al., (1978). De studie uit hoofdstuk 7 laat met behulp van computerondersteunde casussen zien dat er grote overeenkomsten zijn in de manier waarop proefpersonen verschillende casussen oplossen. Deze overeenkomsten bleken uit de hoeveelheid geselecteerde patiëntgegevens tijdens verschillende stadia van het diagnostische proces (anamnese, lichamelijk onderzoek, laboratoriumonderzoek, aanvullend onderzoek, diagnose en beleid), waarbij de hoeveelheid geselecteerde patiëntgegevens gerelateerd was aan de expertisegroep. Alexander en Judy (1988) veronderstellen dat tijdens probleemoplossen bekwame studenten hun inhoudsspecifieke kennis afwegen tegen de complexiteit van het probleem. Vervolgens worden algemene strategieën toegepast om het probleem op te lossen. Het lijkt logisch om te concluderen dat de probleemoplosser algemene strategieën pas kan uitvoeren wanneer deze beschikt over domeinspecifieke kennis. Het is echter te verwachten dat het probleemoplossen minder afhankelijk wordt van algemene strategieën wanneer inhoudsspecifieke kennis groeit. De resultaten van het onderzoek beschreven in hoofdstuk 4 tonen inderdaad aan dat klinisch probleemoplossen zowel specifieke aspecten als algemene aspecten bevat. Daarbij kan diagnosticeren als een specifiek aspect van probleemoplossen worden beschouwd, terwijl het afnemen van een anamnese als een meer algemene vaardigheid kan worden beschouwd, zoals inderdaad vaker wordt aangegeven in de literatuur (Berner, Bligh, & Guerin, 1977; Donnelly, Gallagher, Hess, & Hogan, 1974; Juul, Noe, & Nerenberg, 1979).

Hoofdstuk 5 omvat een studie die het belang van specifieke competenties tijdens het dagelijks functioneren in de co-schappen vergelijkt met het belang van dezelfde specifieke competenties tijdens de eindtoetsen in deze co-schappen. De vraag die hierbij gesteld wordt, is of de componenten van klinische competentie (specifieke competenties) een verschillende waarde krijgen toebedeeld tijdens beide genoemde situaties. Klinische begeleiders uit diverse medische disciplines werd een vragenlijst voorgelegd en gevraagd aan te geven welke competenties zij belangrijk achtten voor het dagelijks functioneren van co-assistenten. Vervolgens gaven zij aan hoe belangrijk zij dezelfde competentie achtten voor de toetsmomenten in de co-schappen. Elke afzonderlijke competentie werd dus tweemaal gescoord, resulterend in twee afzonderlijke datareeksen. Twee onafhankelijke factoranalyses werden uitgevoerd met behulp van SEM software om de onderliggende latente verhoudingen tussen de verschillende competenties aan het licht te brengen en daarna te vergelijken. Voor beide datareeksen bleek een vierfactormodel de covariantiestructuur van de data het beste te verklaren, zoals afgebeeld is in figuur 3. Bij vergelijking van de twee afzonderlijk factoranalyses bleek dat drie van de vier factoren bijna identieke itemstructuren (de individuele competenties) tussen beide factormodellen vertoonden. Deze drie componenten van klinische competentie kwamen overeen met de klassieke categorisatie in cognitieve capaciteiten, interpersoonlijke vaardigheden en professionele kwaliteiten (Epstein & Hundert, 2002; Forsythe, McGaghie, & Friedman, 1986).



Figuur 3. *Samengesteld model van de twee vierfactorenmodellen. Linkerzijde representeert de specifieke competenties en hun factoren tijdens het dagelijks functioneren in de co-schappen en de rechterzijde representeert dezelfde competenties en hun factoren tijdens de eindtoetsen van de co-schappen*

Beide factormodellen brachten ook verschillen in klinische competentie aan het licht, enerzijds binnen de context van het dagelijks functioneren in de co-schappen en anderzijds in de context van expliciete toetsmomenten hierin. Cognitieve capaciteiten blijken belangrijker voor toetsmomenten, terwijl interpersoonlijke vaardigheden en professionele kwaliteiten belangrijker blijken tijdens het dagelijks functioneren op de afdelingen. Een verder verschil tussen de twee modellen was de vierde factor van beide afzonderlijke modellen (zie figuur 3). Toetsing van studenten in hun co-schappen lijkt te zijn gecentreerd om cognitieve capaciteiten en probleemoplossende vaardigheden, die slechts een zwakke relatie vertonen met andere factoren, namelijk interpersoonlijke vaardigheden en professionele kwaliteiten. De competenties die nodig zijn voor het dagelijks functioneren op de afdelingen vertoonden middelmatige tot hoge correlaties onder alle factoren. Geconcludeerd kan worden dat competenties waarvan verondersteld wordt dat zij nodig zijn voor het dagelijks functioneren tijdens de co-schappen niet noodzakelijkerwijs overeenkomen met competenties die als belangrijk worden verondersteld tijdens toetsing. Dit terwijl één van de eisen van toetsing is dat de toetssituatie de context van het dagelijks functioneren representeert.

Deel III: De ontwikkeling van klinische competentie

Studies waarin onderzoek wordt gedaan naar de ontwikkeling van medische expertise hebben twee belangrijke doelstellingen. Ten eerste kan onderzoek naar probleemoplossende vaardigheden bij experts in een specifiek domein informatie verschaffen over de mentale representatie van kennis, met name hoe deze is georganiseerd en gestructureerd. Ten tweede kan dit type onderzoek informatie verschaffen over de ontwikkeling van expertise (Chi, 2006). Medische deskundigen (experts) blijken bij het oplossen van een casus relevante kennis samen te vatten in meer omvattende, klinische begrippen—hun biomedische kennis is geïntegreerd geraakt in hun klinische kennis. Dit integratieproces wordt door Schmidt en Boshuizen "kennisencapsulatie" genoemd (Schmidt & Boshuizen, 1992, 1993a; Schmidt, Boshuizen, & Hobus, 1988). Wanneer, na het diagnosticeren van een casus, aan experts wordt gevraagd wat zij zich nog kunnen herinneren, blijken experts, zich juist deze gecondenseerde begrippen te herinneren in plaats van de 'ruwe' informatie van de casus. Dit in tegenstelling tot minder ervaren probleemoplossers. Dit fenomeen, waarbij gevorderde probleemoplossers meer casuselementen herinneren dan beginnende of experts wordt het "iddeneffect" (*intermediate effect*, zie Schmidt & Boshuizen, 1993b) genoemd en is een vaak voorkomend, robust verschijnsel. De studie beschreven in hoofdstuk 6 focust op dit iddeneffect. Verdeeld over drie verschillende condities, kregen proefpersonen van drie verschillende expertiseniveaus (i.e., preklinische medische studenten, klinische studenten, en ervaren nefrologen) een casus voorgelegd met het verzoek een diagnose te stellen en op te schrijven wat ze zich nog konden herinneren van de casus. In de eerste conditie (de controlegroep) omvatte de casussen een medische context (beknopte gegevens over de medische voorgeschiedenis van de patiënt en het lichamelijk onderzoek) en laboratoriumgegevens. In de tweede conditie (de experimentele groep I) werd de medische context weggelaten en omvatte de casussen alleen laboratoriumgegevens. In de derde conditie (de experimentele groep II) werden de casussen uit de tweede conditie gegeven, maar nu met een speciale instructie om de proefpersonen uitvoerig in te laten gaan op de afzonderlijke elementen van de casus. De resultaten toonden aan dat de hoeveelheid casusinformatie die de experts zich herinnerden, beïnvloed kon worden door de trigger om uit te weiden over de afzonderlijke elementen binnen de laboratoriumgegevens in de casus. In de conditie met de speciale instructie vertoonde de hoeveelheid herinnerde casuselementen, afgezet tegen een toenemend niveau van expertise, een monotoon stijgende lijn. Dit kan beschreven worden als een expertise-effect in tegenstelling tot een middeneffect. Wanneer de andere twee condities werden vergeleken, bleek dat het verwijderen van de context de diagnostische nauwkeurigheid verminderde, maar dat dit geen effect had op de hoeveelheid herinnerde casuselementen. Er kan geconcludeerd worden dat een instructie om specifiek in te gaan op de details van de casus tot diepere verwerking van de casus kan leiden en de normale casusverwerking verstoort.

De hoeveelheid casuselementen die experts zich herinneren na het diagnosticeren van een casus wordt vaak gebruikt als methode om meer inzicht te krijgen in mentale representaties van medische en klinische kennis én in de organisatie hiervan. Inzicht in de ontwikkeling van deze mentale representaties wordt afgeleid door de verschillen tussen de afzonderlijke expertisegroepen te analyseren. Een andere manier om inzicht te krijgen in de mentale kennisrepresentatie van experts is door te onderzoeken hoeveel patiëntinformatie nodig is om een diagnose te kunnen stellen. Het verzamelen van patiëntinformatie gebeurt in verschillende, opeenvolgende stappen, waarbij de anamnese en het lichamelijk onderzoek de logische eerste stappen zijn. Op basis van deze informatie construeert de arts een mentale representatie van het ziektebeeld van de patiënt. Dit proces is afhankelijk van de biomedische en klinische kennis van de arts en van de vroegere ervaring met patiënten. Bijna gelijktijdig aan dit cognitieve proces worden door de arts verschillende diagnostische hypothesen gegenereerd en getoetst aan de nieuw verzamelde casusinformatie (Elstein et al., 1978; Kassirer & Gorry, 1978). De studie uit hoofdstuk 6 onderzoekt expertiseverschillen in de verschillende stappen binnen het diagnostisch proces. Om de wijze van gegevensverzameling door de probleemoplosser nauwkeurig te kunnen analyseren zijn de verschillende stappen of stadia van het diagnostische proces (zoals, anamnese, lichamelijk onderzoek en laboratoriumgegevens) verder verdeeld in kleinere informatie-elementen. Proefpersonen van vier verschillende expertiseniveaus namen deel aan deze studie: studenten met weinig medische voorkennis, klinische studenten, artsen in opleiding tot medisch specialist kindergeneeskunde, en stafleden kindergeneeskunde. Het aantal geselecteerde informatie-elementen en de consistentie van dit gedrag over alle analoge stadia van de vijf gebruikte casussen werd onderzocht. De resultaten toonden aan dat experts, in tegenstelling tot minder ervaren proefpersonen, meer tijd bleken te besteden en meer informatie-elementen bleken te selecteren gedurende de eerste stadia (anamnese en lichamelijk onderzoek) en minder in de latere stadia van het diagnostisch proces (laboratoriumonderzoek, aanvullend onderzoek waaronder beeldvormende technieken en verzoek voor eventuele consultaties). Het selecteren van meer casusinformatie in de initiële periode van het diagnostisch proces kan duiden op een strategie om de mentale, door de casusinformatie geactiveerde probleemruimte te reduceren door middel van het elimineren van (incorrecte) verklaringen of diagnostische hypothesen. Met andere woorden, een ervaren arts selecteert informatie uit de casus om de hoeveelheid mogelijke differentiaal diagnoses te reduceren. Kortom, twee verschillende processen blijken te interacteren gedurende de eerste stadia van het diagnostische proces. Ten eerste een proces waarbij informatie-elementen van de casus worden gebruikt om een mentale representatie op te bouwen en om diagnostische hypothesen te formuleren en ten tweede een proces waarbij de mentale representatie wordt gereduceerd door de nieuw verzamelde informatie (Coughlin & Patel, 1987). De latere stadia van het diagnostische proces worden klaarblijkelijk voornamelijk gebruikt om de diagnostische hypothesen te verifiëren. De minder ervaren proefpersonen missen echter de noodzakelijke medische en klinische kennis om op een juiste wijze gebruik te maken van de informatie uit de casus. Zij gaan door met het verzamelen van informatie vanuit laboratoriumgegevens, aanvullend onderzoek en procedures. Een andere opmerkelijke bevinding in deze studie is de consistentie van dit gedrag gedurende het diagnosticeren van de verschillende casussen. Het patroon van gegevens verzamelen tijdens de anamnese en het lichamelijk onderzoek waren de meest algemene aspecten van het diagnostische proces en het genereren van een correcte diagnose bleek de meest inhouds- of casus specifieke te zijn.

References

Albanese, M. A., Snow, M. H., Skochelak, S. E., Hugett, K. N., & Farell, P. M. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine, 78*(3), 313-321.

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*(4), 375-404.

Berner, E. S., Bligh, T. J., & Guerin, R. O. (1977). An indication for a process dimension in medical problem solving. *Medical Education, 11*, 324-328.

Boshuizen, H. P. A. (2005). Does practice make perfect? A slow and discontinues process. In H. P. A. Boshuizen, R. Bromme & H. Gruber (Eds.), *Professional Learning: Gaps and Transitions on the way from Novice to Expert* (Vol. 2). Dordrecht / Boston / London: Kluwer Academic Publishers.

Butterfield, P. S., & Libertin, A. G. (1993). Learning outcomes of an ambulatory care rotation in internal medicine for junior medical students. *Journal of General Internal Medicine, 8*(4), 189-192.

Châtenay, M., Maguire, T., Skakun, E., Chang, G., Cook, D., & Warnock, G. L. (1996). Does the volume of clinical experiences affect performance of clinical clerks on surgery exit examinations? *The American Journal of Surgery, 172*(4), 366-372.

Chi, M. T. H. (2006). Methods to access the representation of experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge Handbook of expertise and expert performance.* Cambridge University Press.

Coughlin, L. D., & Patel, V. L. (1987). Processing of critical information by physicians and medical students. *Medical Education, 62*, 818-828.

Donnelly, M. B., Gallagher, R. E., Hess, J. W., & Hogan, M. J. (1974). *The dimensionality of measures derived from complex clinical simulations.* Paper presented at the Thirteenth Annual Conference on Research in Medical Education, Chicago, IL, US.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning.* London, Cambridge, Mass: Harvard University Press.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA, 287*(2), 226-235.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10 Suppl), S70-S81.

Ericsson, K. A., & Charness, N. (1994). Expert performance: its structure and acquisition. *American Psychology, 49*(8), 725-747.

Forsythe, G. B., McGaghie, W. C., & Friedman, C. P. (1986). Construct validity of medical clinical competence measures: a Multitrait-Multimethod Matrix study using confirmatory factor Analysis. *American Educational Research Journal 23*(2), 315-336.

Gruppen, L. D., Wisdom, K., Anderson, D. S., & Woolliscroft, J. O. (1993). Assessing the consistency and educational benefits of students' clinical experiences during an ambulatory care internal medicine rotation. *Academic Medicine, 68*(9), 674-680.

Juul, D. H., Noe, M. J., & Nerenberg, R. L. (1979). A factor analytic study of branching patient management problems. *Medical Education, 13*(3), 199-205.

Kassirer, J. P., & Gorry, A. G. (1978). Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine, 89,* 245-255.

McManus, I. C., Richards, P., Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ, 316*, 345-350.

Metz, J. C. M., Verbeek-Weel, A. M. M., & Huisjes, H. J. (2001). *Raamplan 2001 artsopleiding: bijgestelde eindtermen van de artsopleiding*. Den Haag: VSNU.

Moulaert, V., Verwijnen, M. G. M., Rikers, R. M. J. P., & Scherpbier, A. J. J. A. (2004). The effects of deliberate practice in undergraduate medical education. *Medical Education, 18*, 1044-1052.

Neufeld, V. R., & Norman, G. R. (1985). *Assessing Clinical Competence*. New York: Springer-Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In A. E. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York, NY: Springer-Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychological Review, 5*(3), 205-221.

Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition, 21*(3), 338-351.

Schmidt, H. G., Boshuizen, P. A., & Hobus, P. P. M. (1988, 17-19 August 1988). *Transitory stages in the development of medical expertise: The "intermediate effect" in clinical representations studies.* Paper presented at the 10th Annual Conference of the Cognitive Science Society, Montreal, Quebec, Canada.

Schwartz, R. W., Donnelly, M. B., Sloan, D. A., & Young, B. (1994). Knowledge gain in a problem-based surgery clerkship. *Academic Medicine, 69*, 148-151.

Snell, L. M., Battles, J. B., Bedford, J. A., & Washington, E. T. (1998). Verifying the curriculum of a family medicine clerkship. *Medical Education, 32*, 370-375.

Stern, D. T., Frohna, A. Z., & Gruppen, L. D. (2005). The prediction of professional behaviour. *Medical Education, 39*(1), 75-82.

Van Leeuwen, Y. D., Düsman, H., Mol, S. S. L., Pollemans, M. C., Drop, M. J., Grol, R. P. T. M., et al., (1997). Factors Influencing the Growth in Knowledge of Trainees in General Practice. *Advances in Health Sciences Education, 2*(1), 61-70.

Webb, C. T., Sedlacek, W., Cohen, D., Shields, P., Gracely, E., Hawkins, M., et al., (1997). The impact of nonacademic varables on performance at two medical schools. *Journal of the National Medical Association, 89*(3), 173-180.

Witzke, D. B., Koff, N. A., McGeagh, A. M., & Skinner, P. D. (1990). Developing a computer-based system to document and report students' clinical encounters. *Academic Medicine, 65*(7), 440-441.

Slotwoord

[Afterword in Dutch]

EEN FREUDIAANSE INTERESSE in de menselijke psyche en een wellicht deels genetisch bepaalde interesse in onderwijs en kennisverwerving resulteerde in de keuze om psychologie te gaan studeren aan de Universiteit Maastricht (UM). Een verdere afstudeerspecialisatie in cognitieve psychologie en onderwijspsychologie was geen moeilijke keuze net zoals de keuze om met Henk Schmidt aan zijn gloednieuwe psychologie opleiding aan de Erasmus Universiteit Rotterdam (EUR) onderzoek te gaan doen voor mijn afstudeerscriptie. Hoewel ik toen al de smaak te pakken had van leven in de VS leek mij een kort uitstapje naar Rotterdam geen bezwaar. De warmte die Henk als persoon kan uitstralen had ook een positieve invloed op zijn nieuwe medewerkers. Ik werd welkom onthaald en gelijk opgenomen als lid van het splinternieuwe psychologieteam. Een bed, bureau, kantoor, computer en zelfs een onderzoeksassistent (Nadja Strekalova) stonden klaar. In het jonge en energieke psychologieteam kon ik mijn liefde voor onderwijs en psychologie geheel kwijt en leerde ik in een prettige omgeving de eerste kneepjes van onderzoek doen. Een goed gesprek in de wandelgangen, samen lunchen of het binnenhoppen bij collega's bleken vaak de cruciale leermomenten te zijn. Het fundament van de wetenschappelijke kennis dat ik vandaag bezit rust dan ook gedeeltelijk op de schouders van Henk Schmidt en zijn team: kamergenoot Wilco's luisterend oor en onuitputtelijke computerkennis, Peter's kennis van statistiek en Remy's (maar ook Anique's) kennis van medische expertise. Mijn afstudeerscriptie, *Expertise in medicine: a cognitive-psychological approach*, leverde me ook m'n eerste presentatie op bij het massale onderwijscongres in Chicago (AERA). Dat Nederland doorkruizen opzoek naar nefrologen de eerste start werd tot het uiteindelijke product dat u nu in handen heeft, kon ik toen niet weten. Het korte uitstapje naar Rotterdam liep echter wel wat uit en het duurde nog bijna 3 jaar voordat ik echt afscheid van Rotterdam kon nemen. In die 3 jaar werkte ik voor het overgrote deel bij de Medische Faculteit van de EUR om de ingeslagen onderzoeksweg in de vorm van een promotietraject te vervolgen en deels bij psychologie voor het geven van onderwijs. Onder de invloed van de rijke psychologische (onderzoeks-) kennis van Henk en de rijke medische kennis van Ted Splinter werden nog een aantal studies uitgevoerd in een zoektocht om medische expertise en klinische competentie beter te begrijpen. Het opleidingsinstituut van de medische faculteit, waar ik deel van uit maakte, bestond ook hier voornamelijk uit jonge en energieke leden. De werksfeer was dan ook erg aangenaam en stimulerend. Met veel plezier kijk ik terug op de tijd dat ik mijn kamer deelde met Anneke. We waren de enige twee psychologen op de gang en we schroomden dan ook niet om banale dagelijkse, vaak persoonlijke beslommeringen tot in alle diepte en openheid te bespreken. Daar ik graag urenlang achter mijn computer bleef zitten werken waren Marije's frequente theemomenten altijd erg welkom. De ontelbare buitenwerktijdse activiteiten variërend van borrels tot dansavonden met Anneke, Dirk, Frederica, Loes, en Marije zal ik zeker niet gauw vergeten. Ik ben dan ook heel dankbaar en trots dat Marije van Mannekes en Anneke Zanting me als paranimfen op 19 oktober ter zijde staan. De combinatie van werken aan twee verschillende faculteiten had buiten een dubbel kerstpakket ook ander voordelen: het alles overtreffende en zeer succesvolle SEM project met Sofie in de 100 meter hoge Euromast was toch letterlijk het hoogtepunt, we waren het beste team. Het vertoeven in de multiculturele wereldstad Rotterdam gaf meer dan alleen leuk werk. Inwonen bij Wilco, een kamer in Schiedam, tijdelijk wonen in Gerard Holthuis' atelier aan de 2de Middellandsstraat en mijn appartement in Delfshaven zijn nu nog enkel herinneringen, net zoals een wandeling langs de Maas, het uitzicht vanaf de Holland-Amerika kade of een nachtje doorhalen met Melo en Stijn in de Now & Wow. In 2005 kwam echter een einde aan mijn Rotterdams bestaan. Mijn promotietraject was nog niet ten einde toen een aantrekkelijk aanbod om aan de Medische Faculteit van de University of Pittsburgh te gaan werken me deed besluiten voor het grote avontuur te kiezen. Ten eerste bracht het me terug in de VS en ten tweede kon ik bij een goed instituut mijn werk in medische expertise voortzetten. Afscheid nemen deed pijn, maar Pittsburgh gaf veel terug. Werken met de Vice Dean Steve Kanter bracht me in contact met nieuwe vrienden (Maurice en John) en andere interessante ervaringen kruisten mijn pad. De studie beschreven in hoofdstuk 7 is een voorbeeld van mijn werk in Pittsburgh. Goede herinneringen zijn Henk Schmidt's bezoek aan ons huis in Squirel Hill, Pittsburgh en zijn presentatie over medische expertise aan de medische faculteit en natuurlijk het verblijf van mijn vader en later Eric. Echter het afronden van mijn promotiewerk in combinatie met een fulltime baan bleken niet altijd even makkelijk. Hoewel het meetkundig onmogelijk is, wegen de laatste loodjes inderdaad het zwaarst. Een promotietraject is dan niet alleen een leertraject maar ook een gevecht met jezelf, afzondering en het afdwingen van motivatie en discipline. Je leert dat kennis en wijsheid twee verschillende dingen zijn, dat wie je kent belangrijker kan zijn dan wat je kent of doet en dat de balans tussen ontspanning en inspanning vaak moeilijk te bepalen is. Gelukkig kan je op die momenten terugvallen op de ondersteuning van je omgeving. Valerie's geduld en liefde voor mij kenden dan ook geen grenzen, vooral als ik terugdenk aan een 'onaanspreekbare' persoon (ik) die achter gesloten deur, nachtenlang, statistische modellen aan het vergelijken en beschrijven was.

Dat alles toch goed afloopt blijkt uit het feit dat ik nu onder het genot van Mozart's Requiem, een frisse oceaanwind en uitzicht over de palmbomen van Santa Monica het slotwoord schrijf. En dat alles in het bezit van een nieuwe en uitdagende baan bij een van de beste instituten ter wereld: the University of California, Los Angeles (UCLA).

Tot slot wil ik graag mijn helden Gregory Hancock en Ralph Mueller bedanken voor hun vertrouwen en inspiratie. Anju Relan en Sally Krasne wil ik bedanken voor de constructieve commentaren op voorgaande versies van hoofdstukken. En natuurlijk mijn vader, moeder en familie én mijn goede vrienden bedanken voor het geloof in mij en de nimmer aflatende steun. Ik houd van jullie.

Paul, Santa Monica, 19 Augustus 2006

Curriculum vitae

PROFESSIONAL BIOGRAPHY Paul Wimmers started his career as a Technical Engineer and Instructor and made a career change in 1998 to pursue his interests in psychology and education. He received his Master's degree at Maastricht University, the Netherlands, in cognitive and educational psychology in 2002. He took additional courses at the University of California, Los Angeles (UCLA), to expand his knowledge of education and educational research. He was awarded a position in the Psychology Department of Erasmus University Rotterdam as teaching and research assistant. In September 2002, he transferred to the Medical Department of Erasmus University to complete his Ph.D in psychology. In June 2005 he accepted a faculty position in the Office of the Vice Dean at the University of Pittsburgh School of Medicine where he could continue to work on research in medical expertise and medical education. In June 2006 Paul relocated to David Geffen School of Medicine at UCLA to accept a position as assistant professor in the Center for Education Development and Research (ED&R). His professional interests include the acquisition of expertise, clinical learning, medical problem solving, evaluation development, admission policies, and constructivist teaching strategies (e.g., Problem-based Learning).

Publications

Wimmers, P. F., Splinter, T. A. W., Hancock, G. R., & Schmidt, H. G. (in press 2006). Clinical competence: General ability or case-specific? *Advances in Health Sciences Education.*

Kanter, S. L., **Wimmers, P. F.**, & Levine, A. S. (in press 2006). In-depth learning experiences enable full integration of ethics, values, and the human dimensions of medicine. *Academic Medicine*.

Krasne, S. J., **Wimmers, P. F.**, Relan, A., & Drake, T. A. (2006). Differential effects of two types of formative assessments on performance of first-year medical students. *Advances in Health Sciences Education, 11*, 155-171.

Wimmers, P. F., Schmidt, H. G., Verkoeijen, P. P. J. L., & van de Wiel, M. W. J. (2005). Inducing expertise effects in clinical case recall. *Medical Education, 39*(9), 949-957.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (2006). Influence of clerkship experiences on clinical competence. *Medical Education, 40*(5), 450-458.

Paper presentations

Baillie, S., Crooks, K. A., & **Wimmers, P. F.** (May 21-24, 2006). *Evaluation of the quality and appropriateness of elective clinical rotations.* Paper presented at the 12th International Ottawa Conference on Medical Education, New York, NY, US.

Clifton, M. S., **Wimmers, P. F.**, & Mahoney, J. F. (May 21-24, 2006). *Can factor analysis in OSCEs help us learn about clinical skills?* Paper presented at the 12th International Ottawa Conference on Medical Education, New York, NY, US.

Wimmers, P. F., & Kanter, S. L. (May 21-24, 2006a). *Could sequence of diagnostic processing be an indicator for level of performance?* Paper presented at the 12th International Ottawa Conference on Medical Education, New York, NY, US.

Wimmers, P. F., & Kanter, S. L. (March 3-4, 2006b). *Expert-novice differences in data-gathering behavior during diagnoses.* Paper presented at the Northeast Group on Educational Affairs, Philadelphia, PA, US.

Wimmers, P. F., Staszewski, J. J., & Kanter, S. L. (April 8-12, 2006). *Effect of level of expertise on the selection of knowledge during the diagnostic process.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, US.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (Nov 11-12, 2004). *Klinische competentie: inhoudsspecifiek of een algemene vaardigheid?* [Clinical competence: Content specific or a general ability]. Paper presented at the Nederlandse Vereniging voor Medisch Onderwijs [Dutch Association of Medical Education], Egmond aan Zee, the Netherlands.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (Sept 5-8, 2004). *Evaluating clinical education with the use of patient logbooks: A literature review.* Paper presented at the Annual Meeting of the Association for Medical Education in Europe, Edinburgh, UK.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (July 6-8, 2004). *Pre-medical achievement as predictor for success in medical school.* Paper presented at the 11th International Ottawa Conferences on Medical Education, Barcelona, Spain.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (April 12-16, 2004). *How content specific is clinical competence?* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, US.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (Nov 21-22, 2003). *Verschillen in kwaliteit en kwantiteit van problemen/ziektebeelden tijdens een co-schap interne geneeskunde* [Quantitative and qualitative differences of problems and diseases seen during an internal medicine clerkship]. Paper presented at the Nederlandse Vereniging voor Medisch Onderwijs [Dutch Association of Medical Education], Egmond aan Zee, the Netherlands.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (Nov 21-22, 2003). *De rol van klinische ervaring en begeleiding op competentie: Analyse van een co-schap interne geneeskunde* [The influence of clinical experience and supervision on clinical competence: Analysis of an internal medicine clerkship]. Paper presented at the Nederlandse Vereniging voor Medisch Onderwijs [Dutch Association of Medical Education], Egmond aan Zee, the Netherlands.

Wimmers, P. F., Splinter, T. A. W., & Schmidt, H. G. (Aug 31-Sept 3, 2003). *Modeling clinical competence in a medical internship: The impact of variation in actual experiences.* Paper presented at the Annual Meeting of the Association for Medical Education in Europe, Bern, Switzerland.

Wimmers, P. F., Verkoeijen, P. P. J. L., van de Wiel, M. W. J, & Schmidt, H. G. (April 21-25, 2003). *Inducing shifts in clinical case processing by manipulation of format and instruction.* Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, US.

Poster presentations

Relan, A., Krasne, S. J., & **Wimmers, P. F** (2006, April 30 - May 3). *Medical students' perceptions of cognitive and motivational effects of weekly formative assessments.* Presented at the Western Group Educational Affairs (WGEA) Regional Conference, Asilomar, Pacific Grove, California, US

Baillie, S., Relan, A., & **Wimmers, P. F**. (July 6-8, 2004). *Student perceived effectiveness of clerkship quality: What makes a difference?* Presented at the 11th International Ottawa Conference on Medical Education, Barcelona, Spain.

Krasne, S. J., Relan, A., **Wimmers, P. F.**, Fung, C.-C., & Drake, T. A. (July 6-8, 2004). *The use of on-line formative assessments to enrich learning in an integrated medical school curriculum.* Presented at the 11th International Ottawa Conference on Medical Education, Barcelona, Spain.