Strategic choices in curriculum design to facilitate knowledge and competency development
Kerdijk, Wouter

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date:
2014

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Strategic choices in curriculum design to facilitate knowledge and competency development

Dissertation for the University of Groningen, the Netherlands, with references and summary in Dutch. The studies presented in this thesis were carried out at the Center for Research and Innovation in Medical Education within the Graduate School of Medical sciences and Research institute SHARE of the University of Groningen.

Address for correspondence:
Wouter Kerdijk
University Medical Center Groningen
Antonius Deusinglaan 1, FB 21
9713 AV Groningen, The Netherlands
W.kerdijk@umcg.nl

Design and Layout: Rob de Kort/kortsluiting.com & Erik Noppert
Cover illustration: Peter Kerdijk

© Kerdijk, W., 2014
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the author.

ISBN 978 90 367 7253 2
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>The effect of implementing undergraduate competency-based medical education on students' knowledge acquisition, clinical performance and perceived preparedness for practice: a comparative study</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>Identifying essential clinical rotations for a competency-based curriculum</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>Acquisition and retention of discipline-specific knowledge during clerkships</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>Cumulative assessment: strategic choices to influence students’ study effort</td>
<td>97</td>
</tr>
<tr>
<td>6</td>
<td>Cumulative assessment versus assessment at the end of a course: effects on self-study time and test performance</td>
<td>121</td>
</tr>
<tr>
<td>7</td>
<td>General Discussion</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>Samenvatting</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>Dankwoord</td>
<td>193</td>
</tr>
</tbody>
</table>
COMPETENCY-BASED EDUCATION has become the foundation for many medical curricula all over the world.\textsuperscript{1-10} Even though it has been a topic of discussion since the 1970s and 1980s,\textsuperscript{11-16} over the past decade the concept has rapidly gained wide acceptance. Competency frameworks such as the CanMEDS, Tomorrow's doctor, the Scottish doctor and the ACGME general competencies framework have been implemented widely in both postgraduate and undergraduate education.\textsuperscript{1,17-21}

Competency-based medical education has been defined as “an approach to preparing physicians for practice that is fundamentally oriented to graduate outcome abilities and is organized around competencies derived from an analysis of societal and patient needs. It de-emphasizes time-based training and promises greater accountability, flexibility, and learner centeredness”.\textsuperscript{2,22} In short, it is an outcome-based approach to curriculum design using an organizing framework of competencies.\textsuperscript{2} In this definition, competencies are defined as observable abilities students should develop, integrating multiple components such as knowledge, skills and behaviour.\textsuperscript{2,3,23,24} As such, competencies are student characteristics that should be defined clearly, unambiguously, and sufficiently general.\textsuperscript{3,23-25} Competencies should reflect the view and mission of a medical school's stakeholders, address predefined areas of competence and inform the learners about what is expected of them.\textsuperscript{3,24,26} Furthermore, competencies should reflect the ultimate goal of the curriculum and be formulated in terms of assessable behaviour.\textsuperscript{24}

The growing support for competency-based medical education is associated with changes in society. A major impetus for the competency movement has been a widening call from society for more accountability and professionalism of medical schools and healthcare professionals.\textsuperscript{1,11,27,28} In competency-based medical education, it is clearly defined what is expected at the end of the curriculum. Ideally, only professional doctors and students who have mastered all necessary competencies will be able to pass the exams.\textsuperscript{29} Consequently, competency-based medical education increases the accountability of medical schools. Furthermore, the developers of competency frameworks, such as the CanMEDS, specifically took into account the interests of nonmedical stakeholders like public organizations and patients.\textsuperscript{1,2,20} Competency-based education has been receiving increasing support from legislators and curriculum developers, since proper implementation warrants more accountability to and focus on the needs of society.\textsuperscript{1,9,22,30}

Another factor that facilitated the adoption and dissemination of competency-based medical education in Europe is the Bologna Declaration.\textsuperscript{31} The Bologna Declaration is a document that aims at European harmonisation of higher education in order to facilitate international student mobility and was signed by the ministers of education of the European Union member states. This agreement resulted in the Tuning Project, which aims to develop a framework of comparable learning outcomes as the basis for all undergraduate medical curricula in Europe.\textsuperscript{32-34} Competency-based education offers curriculum designers a way to achieve this goal, because it offers clear-cut guidelines for formulating learning outcomes.
There is also an educational reason behind the widespread implementation of competency-based medical education. In traditional medical curricula, emphasis is mainly placed on knowledge acquisition, while skills and attitudes receive relatively less attention, even though the latter two are also highly relevant for doctors’ performance in practice. Over the past decades, skills training for medical students has already been receiving much attention in many medical schools. This development is evidenced by an impressive body of research into learning skills through simulation and assessing clinical skills through tools such as the OSCE and the Mini-CEX. Implementing competency-based education ensures that professional behaviour also receives sufficient attention and that a holistic approach is adopted to what makes a competent doctor.

A curriculum will be most effective when there is a structural alignment between its educational goals, the teaching formats and the assessment program. Therefore, curriculum design requires strategic choices to benefit student learning. Most of the literature about the implementation of competency-based education addresses educational goals. These articles mainly focus on how to formulate learning outcomes and what competency-based medical education is expected to achieve. However, very little is known about which teaching formats and assessment methods are optimal for competency-based medical education, especially at the undergraduate level. In addition, it is unknown how changing the curriculum towards competency development influences its outcomes. For example, if considerable time of the medical curriculum is devoted to competency development, how does this influence student learning and development? In former curricula the main focus was on the acquisition of basic knowledge and skills. How does the shift towards competency-based curricula impact students’ knowledge and skills development? In this thesis some pieces of this puzzle are addressed.
OUTLINE OF THE THESIS

This thesis investigates the competency-based undergraduate medical curriculum that was implemented at the University of Groningen in 2003. More specifically, the research described in this thesis focuses on the effects of and risks associated with several strategic choices in curriculum design that aim to facilitate student knowledge and competency development.

In Chapter 2 the design choices in a competency-based undergraduate curriculum and their educational effects are addressed. We investigated differences concerning three educational outcomes between the last two cohorts before and two cohorts after the implementation of a competency-based curriculum. We studied students' knowledge growth, clinical performance and perceived preparedness for practice.

Subsequently two lines of inquiry will be further explored, focusing on two specific strategic choices aimed to benefit competency development and knowledge development, respectively. The first line of inquiry concerns the clinical phase, at the end of which students have to fully master the required competencies for graduation. It has been suggested that longer clerkship rotations are a beneficial choice in curriculum design that should provide the optimal context for students to develop their competencies. Given the limited duration of the clinical phase, longer rotations will lead to less variety in disciplines students rotate through. This raises the question whether variety of disciplines offers students specific educational benefits. Are there benefits associated with rotating through specific disciplines?

The study described in Chapter 3 aimed to explore the suitability of various disciplines for students to master a set of prescribed competencies. First, a focus group of medical experts and students rated the suitability of 12 disciplines for students to master each of 177 competencies. Subsequently, we explored which clerkships could be considered as mandatory.

Rotating through specific disciplines, might influence students' knowledge about those disciplines. In Chapter 4 we investigated how rotating through a certain discipline during clerkships affects students' discipline-specific knowledge, as measured by the Dutch interuniversitary progress test. For five disciplines we compared students' discipline-specific knowledge before, during and after a clerkship and, if applicable, after a second clerkship in that discipline. Furthermore, we analysed to what extent discipline-specific knowledge declines from the moment a clerkship rotation in a specific discipline has ended.

The second line of inquiry regards how assessment can be strategically used to benefit students' knowledge development. Implementation of competency-based medical education means that competency development becomes a formal curricular goal. Often, this goal is added to the already existing goals of knowledge and skills development. A
likely consequence is that some of the former curriculum time aimed at knowledge development will be reserved for competency development in the new curriculum. Therefore, less time will be available for knowledge acquisition. This may become a problem, because medical knowledge plays a key role in developing medical expertise. Assessment may offer alternative means of supporting students' knowledge development. The central issue in our research is cumulative assessment. The purpose of cumulative assessment is to keep students studying by applying principles such as repeated testing, repetition of content and compensation among tests.

Chapter 5 presents our research into how cumulative assessment influences students' test scores throughout a 10-week course. We investigated two second-year courses and two third-year courses. We compared tests scores of initially low and initially high scoring students over the course, while correcting for regression to the mean and test difficulty.

In Chapter 6 a random controlled study is described in which the difference between cumulative assessment and end-of-course assessment was investigated. We analysed self-study time students spent throughout a course and students' performance at the end of the course in two conditions, cumulative and end-of-block assessment. Furthermore, we measured how students perceived both assessment methods to influence their study behaviour.

Chapter 7 provides a brief summary of the main findings of this thesis and includes a general discussion in which our findings are considered in the light of knowledge development and competency development in undergraduate medical education. We discuss several methodological considerations and implications for medical education practice. Suggestions are done for future research and the general direction of research on competency-based medical education.
REFERENCES

The effect of implementing undergraduate competency-based medical education on students’ knowledge acquisition, clinical performance and perceived preparedness for practice: a comparative study

Wouter Kerdijk
Jos W Snoek
Elisabeth A van Hell
Janke Cohen-Schotanus

Published in: BMC Medical Education 2013;13:76
ABSTRACT

Background Little is known about the gains and losses associated with the implementation of undergraduate competency-based medical education. Therefore, we compared knowledge acquisition, clinical performance and perceived preparedness for practice of students from a competency-based active learning (CBAL) curriculum and a prior active learning (AL) curriculum.

Methods We included two cohorts of both the AL curriculum (n=453) and the CBAL curriculum (n=372). Knowledge acquisition was determined by benchmarking each cohort on 24 interuniversity progress tests against parallel cohorts of two other medical schools. Differences in knowledge acquisition were determined comparing the number of times CBAL and AL cohorts scored significantly higher or lower on progress tests. Clinical performance was operationalized as students’ mean clerkship grade. Perceived preparedness for practice was assessed using a survey.

Results The CBAL cohorts demonstrated relatively lower knowledge acquisition than the AL cohorts during the first study years, but not at the end of their studies. We found no significant differences in clinical performance. Concerning perceived preparedness for practice we found no significant differences except that students from the CBAL curriculum felt better prepared for ‘putting a patient problem in a broad context of political, sociological, cultural and economic factors’ than students from the AL curriculum.

Conclusions Our data do not support the assumption that competency-based education results in graduates who are better prepared for medical practice. More research is needed before we can draw generalizable conclusions on the potential of undergraduate competency-based medical education.
BACKGROUND

IN RESPONSE to societal concerns about the role of doctors in contemporary healthcare, competency-based medical education is receiving increasing attention worldwide. Its underlying assumption is that competency-based medical education results in doctors who are better prepared for medical practice. In Canada and the United States, the national accreditation councils have implemented competency-based criteria for postgraduate medical education. Additionally, a competency framework has been proposed and guidelines have been developed for undergraduate competency-based medical education. In the European Union, as part of the Bologna process, all medical schools are required to base their undergraduate curricula on a clear and well-defined set of competencies. A major focus of competency-based curricula is to facilitate students’ development of competencies, demonstrable abilities consisting of knowledge, skills and professional behaviour. Consequently, when implementing competency-based medical education, curriculum time has to be reserved for students’ competency development. This means there will be less time available for existing activities of preceding curricula. Therefore, such a reallocation of time may not only result in the facilitation of competency development but may also impair students’ development in other areas. To our knowledge, the gains and losses associated with implementing undergraduate competency-based curricula are still unknown. Therefore, we examined undergraduate medical students’ knowledge acquisition, clinical performance and perceived preparedness for medical practice for two curricula – a competency-based active learning (CBAL) curriculum and its predecessor, a regular active learning (AL) curriculum.

Undergraduate medical curricula usually have a set duration. When implementing competency-based education, curriculum time has to be reserved so students can develop their competencies. The time reserved for activities aimed at competency development will usually come at the expense of time previously reserved for knowledge acquisition. This reallocation of time may negatively affect students’ knowledge acquisition in a competency-based curriculum. Although medical students’ knowledge has not been found to be an immediate predictor of clinical performance, it does impact clinical performance indirectly. To allow for well-informed decisions about curriculum innovations, it should be clear whether implementing undergraduate competency-based medical education leads to knowledge loss among medical students.

One of the key forces behind competency-based medical education is the public call for medical curricula to reflect the needs of contemporary medical practice. Therefore, competency frameworks comprehensively reflect what a competent doctor should be able to demonstrate in practice, and should benefit students’ preparation for medical practice. Throughout competency-based curricula, relevant competencies and their relation with practice are continuously emphasized which helps students to understand what is expected of
them during medical training and in medical practice. Consequently, students should feel better prepared for practice which, in turn, is a prerequisite for self-efficacy—the extent to which a person believes that he or she can successfully fulfil a specific task in a specific context. Self-efficacy is of key importance for developing competence and autonomy in practice. We expected students from a CBAL curriculum to feel better prepared for medical practice and to perform better during clerkships than students from an AL curriculum, where the presence of an underlying competency framework is less explicit.

The possible gains and losses associated with the implementation of an undergraduate competency-based medical curriculum provide valuable information for future curriculum development and add to the theory of competency-based medical education. Therefore, we examined the influence of implementing a competency-based curriculum on medical students’ knowledge acquisition, clinical performance and perceived preparedness for medical practice.

METHODS

Context
The AL and the CBAL curriculum were developed and implemented at the University of Groningen, The Netherlands. Characteristics of both curricula are presented in Table 1. The CBAL curriculum was implemented in September 2003 and focuses on seven areas of competence: communication, clinical problem-solving, using basic knowledge and science, patient investigation, patient management, social and community contexts of health care and reflection.

In both curricula, active learning principles are applied to facilitate knowledge acquisition. Students learn in small groups, collaborate with their peers and engage in self-directed learning. Teachers and tutors fulfil a coaching and facilitating role.

Learning methods and the amount of time reserved for skills training are similar in both curricula. However, in the AL curriculum skills training is divided over smaller courses throughout the preclinical phase, whereas skills training in the CBAL curriculum is concentrated in the first year of the clinical phase. During this year, five-week periods of skills training in the clinical training centre are alternated with five-week clerkship rotations. The purpose of this alternation is to ease the transition from the preclinical to the clinical phase by helping students...
develop their skills, just in time, to apply them in practice and to further integrate them with knowledge and professional behaviour.\textsuperscript{24}

Table 1. Characteristics of the Active Learning and Competency-Based Active Learning curriculum at the UMCG.

<table>
<thead>
<tr>
<th></th>
<th>AL curriculum</th>
<th>CBAL curriculum</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active learning</strong></td>
<td>Emphasis on active learning in small groups</td>
<td>Emphasis on active learning in small groups</td>
</tr>
<tr>
<td><strong>Focus on competencies</strong></td>
<td>0% curriculum time allocated specifically for competency development</td>
<td>15% curriculum time allocated specifically for competency development</td>
</tr>
<tr>
<td></td>
<td>No portfolio or small group sessions aimed at competency development</td>
<td>Portfolio and small group sessions aimed at competency development</td>
</tr>
<tr>
<td></td>
<td>Purpose of a course is communicated</td>
<td>Purpose of a course and the related competencies are communicated</td>
</tr>
<tr>
<td><strong>Clerkships</strong></td>
<td>80 weeks of clinical experience</td>
<td>80 weeks of clinical experience</td>
</tr>
<tr>
<td>Rotational duration is</td>
<td>1-8 weeks</td>
<td>4-5 weeks</td>
</tr>
<tr>
<td></td>
<td>22 rotations</td>
<td>15 rotations</td>
</tr>
<tr>
<td>Last clerkship rotation entails an elective of 13 weeks</td>
<td>Last clerkship rotation entails an elective of 20 weeks</td>
<td></td>
</tr>
</tbody>
</table>

The main difference between the two curricula lies in the emphasis on competency development. In the CBAL curriculum, the link between the purpose of each course and relevant competencies are clearly communicated throughout the course. This is not the case in the AL curriculum. Furthermore, 15\% of the total CBAL curriculum time is reserved specifically for small group sessions aimed at competency development. Time for these sessions is created by diminishing the number of small group sessions originally aimed at knowledge acquisition in the AL curriculum. The total curriculum time remains the same. Throughout the preclinical phase of the CBAL curriculum, small group sessions for competency development are based on students’ experiences in practice and assignments related to each area of competence. An example of such an assignment is that first-year students, unfamiliar with medical practice, have to describe the qualities of a good doctor. In their third study year the students have to repeat this assignment, and reflect on what they have learnt and experienced in the meantime. Other assignments are related to activities in intramural or extramural practice – for example an internship in a nursing home or consecutive interviews with a chronically ill patient. The competency development sessions are facilitated by a senior faculty member and are scheduled six to eight times a year. Additionally, students have to collect their assignments in a portfolio, on which they receive feedback bi-annually.

During the clinical phase sessions aimed at competency development are scheduled 24 times a year. During these sessions students discuss their own experiences and certain themes in relation to their development (for example cultural diversity or dealing with death). In addition to
assignments related to these meetings, students have to keep track of a personal development plan in their portfolio in which they formulate learning goals based on the areas of competence. During the clinical phase the portfolio is evaluated twice a year in an interview with a senior staff member.

The curriculum time available for clerkships is 80 weeks in both curricula. Students in the CBAL curriculum rotate through fewer disciplines than students in the AL curriculum. In the AL curriculum, clerkship duration varies between one and eight weeks and students rotate through 22 disciplines. When designing the CBAL curriculum we felt that the aim of clerkships shifted from experiencing as many disciplines as possible towards a balance between diversity and the stability of surroundings to support students' competency development. Consequently, in the CBAL curriculum, the minimum duration for clerkship rotations was extended to 4 weeks to allow sufficient time for students to work on their competencies. Consequently, the number of clerkship rotations was reduced to 15. Furthermore, the last clerkship rotation entails a clinical elective of which the duration was increased from 13 weeks in the AL curriculum to 20 weeks in the CBAL curriculum.

**Participants**
Undergraduate medical education in The Netherlands lasts 6 years. We included students who graduated within 7 years from the start of the last 2 cohorts of the AL curriculum ((2001/2002 and 2002/2003; N=453) and the first 2 cohorts of the CBAL curriculum (2003/2004 and 2004/2005; N=372).

**Ethical statement**
Data were gathered during the time that, under Dutch law, educational studies were exempt from Institutional Board Review. At that time, no ethical review board for medical educational research existed in the Netherlands. However, data gathering was carried out in accordance with established ethical standards and the Declaration of Helsinki. The privacy policy of the University of Groningen states that student records can be used for research purposes, as long as reports cannot be traced back to individual students. In accordance with this privacy policy, anonymized data were derived from the university administration.

**Instruments**
Knowledge acquisition was assessed by benchmarking our cohorts' scores on the Dutch interuniversity progress test (IPT) against those of parallel cohorts from two other Dutch medical schools with similar cohort sizes (approximately 250 students per cohort). All cohorts sat the IPT four times per year at the same time, i.e. 24 tests per cohort. The IPT is based on the Dutch National Blueprint for the Medical Curriculum, and is designed to assess "the end objectives of undergraduate medical training as far as knowledge is concerned". Each progress test contains 200 multiple choice questions and is constructed to reflect the entire domain of medical knowledge. The IPT is not related to the curriculum of one particular institution. The reason for benchmarking against two other medical schools was that all students sat exactly the same tests at the same point in their education. IPT benchmarking is especially
suitable for analysing effects of curriculum changes, because, at the time of our study admittance to medical schools in the Netherlands was still primarily determined by a national lottery system. This system guarantees an intake of first-year students which is very similar across medical schools with regard to past performance, age, gender and motivation to study medicine. During the time of our study the medical schools used for comparison had not changed their curricula.

Clinical performance was operationalized as students’ average clerkship grade. In both curricula clinical assessment was identical: each clerkship grade was based on several mini-CEX scores. Mini-CEX scores are sufficiently reliable to estimate clinical competence. In both curricula, grades were given on a 10-point scale.

To measure perceived preparedness for medical practice, we used data from an internal quality control survey, measuring how prepared students feel in each area of competence. Perceived preparedness was measured for 33 competencies (Table 2), using a 5-point scale (1 = ‘insufficiently prepared’, 5 = ‘excellently prepared’). Our medical school considers a mean score between 4 and 5 as excellently prepared, between 3 and 4 as well-prepared and below 3 as insufficiently prepared.

**Analysis**

To analyse students’ knowledge acquisition, we used a method based on the first steps in the longitudinal benchmarking methods described by Muijtjens et al. We compared our students’ average score to those of the students from the other medical schools, using t-tests. The 24 means were plotted in a graph for each cohort. When our students scored significantly higher or lower, a ↑ or ↓ was drawn in the graph, respectively. A Bonferroni correction was used to compensate for the high number of tests and effect sizes were calculated.

We compared clinical performance and perceived preparedness for medical practice in the CBAL and AL curriculum using independent sample t-tests. With regard to perceived preparedness for medical practice, we first calculated the internal consistency of the scales using Cronbach’s α. Subsequently, curricula were compared on the mean scores for both items and scales using an α of 0.01 and effect sizes were calculated.
<table>
<thead>
<tr>
<th>Table 2. Means, standard deviations and t-statistics for perceived preparedness of graduates from two curricula.</th>
<th>Active learning curriculum</th>
<th>Competency-based active learning curriculum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>(SD)</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Communication</strong> (α = 0.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicating with a patient</td>
<td>4.29 (0.44)</td>
<td>4.33 (0.45)</td>
</tr>
<tr>
<td>Treating a patient with respect and confidentiality</td>
<td>4.51 (0.56)</td>
<td>4.62 (0.5)</td>
</tr>
<tr>
<td>Working together with colleagues</td>
<td>4.36 (0.59)</td>
<td>4.32 (0.54)</td>
</tr>
<tr>
<td>Accepting the expertise of others</td>
<td>4.31 (0.56)</td>
<td>4.36 (0.54)</td>
</tr>
<tr>
<td>Building and maintaining a doctor-patient relationship</td>
<td>4.00 (0.74)</td>
<td>3.98 (0.76)</td>
</tr>
<tr>
<td>Efficiently consulting with colleagues and other health care professionals</td>
<td>4.08 (0.63)</td>
<td>4.14 (0.55)</td>
</tr>
<tr>
<td><strong>Clinical problem-solving</strong> (α = 0.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a systematic approach to a patient problem</td>
<td>3.98 (0.48)</td>
<td>3.90 (0.4)</td>
</tr>
<tr>
<td>Interpreting problem descriptions, patient history, physical examinations and other findings</td>
<td>4.15 (0.6)</td>
<td>4.09 (0.54)</td>
</tr>
<tr>
<td>Making a differential diagnosis</td>
<td>3.99 (0.57)</td>
<td>3.88 (0.54)</td>
</tr>
<tr>
<td>Deciding which information about treatment should be provided to the patient</td>
<td>3.88 (0.61)</td>
<td>3.77 (0.57)</td>
</tr>
<tr>
<td><strong>Using basic knowledge and science</strong> (α = 0.86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conducting scientific research</td>
<td>3.53 (0.67)</td>
<td>3.62 (0.62)</td>
</tr>
<tr>
<td>Approaching scientific information critically</td>
<td>3.45 (0.86)</td>
<td>3.63 (0.77)</td>
</tr>
<tr>
<td>Converting scientific information into effective policy</td>
<td>3.58 (0.85)</td>
<td>3.68 (0.71)</td>
</tr>
<tr>
<td>Justifying conduct based on a scientific argumentation</td>
<td>3.52 (0.75)</td>
<td>3.53 (0.73)</td>
</tr>
<tr>
<td><strong>Patient investigation</strong> (α = 0.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosing a patient problem</td>
<td>3.91 (0.48)</td>
<td>3.86 (0.4)</td>
</tr>
<tr>
<td>Documenting relevant information</td>
<td>3.99 (0.7)</td>
<td>4.02 (0.59)</td>
</tr>
<tr>
<td>Performing a physical examination</td>
<td>4.02 (0.63)</td>
<td>3.99 (0.58)</td>
</tr>
<tr>
<td>performing of medical skills expected from an MD</td>
<td>3.73 (0.64)</td>
<td>3.61 (0.6)</td>
</tr>
<tr>
<td><strong>Patient management</strong> (α = 0.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determining a founded and suitable treatment</td>
<td>3.77 (0.55)</td>
<td>3.68 (0.52)</td>
</tr>
<tr>
<td>Executing a treatment plan</td>
<td>3.69 (0.66)</td>
<td>3.59 (0.61)</td>
</tr>
<tr>
<td>Monitoring the effects of a treatment plan</td>
<td>3.58 (0.71)</td>
<td>3.48 (0.78)</td>
</tr>
<tr>
<td>Adjusting a treatment plan</td>
<td>3.56 (0.74)</td>
<td>3.39 (0.71)</td>
</tr>
<tr>
<td>Hold an effective and respectful consultation with a patient</td>
<td>4.35 (0.61)</td>
<td>4.42 (0.56)</td>
</tr>
<tr>
<td><strong>Social and community contexts of health care</strong> (α = 0.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placing a patient problem in a broad context of political, sociological, cultural and economic factors</td>
<td>3.66 (0.49)</td>
<td>3.77 (0.48)</td>
</tr>
<tr>
<td>Being aware of the consequences of the patient problem for the patients environment</td>
<td>3.75 (0.77)</td>
<td>3.97 (0.66)</td>
</tr>
<tr>
<td>Having knowledge of factors that influence health and disease at societal level</td>
<td>4.03 (0.64)</td>
<td>4.11 (0.57)</td>
</tr>
<tr>
<td>Promoting health of patient and society as a whole</td>
<td>3.56 (0.68)</td>
<td>3.72 (0.66)</td>
</tr>
<tr>
<td>Following relevant legal regulations</td>
<td>3.61 (0.71)</td>
<td>3.69 (0.71)</td>
</tr>
<tr>
<td><strong>Reflection</strong> (α = 0.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognizing and acknowledging one’s own shortcomings</td>
<td>3.33 (0.77)</td>
<td>3.37 (0.75)</td>
</tr>
<tr>
<td>Combining work life with private life</td>
<td>3.78 (0.67)</td>
<td>3.82 (0.63)</td>
</tr>
<tr>
<td>Dealing with ethical dilemmas</td>
<td>3.72 (0.79)</td>
<td>3.79 (0.77)</td>
</tr>
<tr>
<td>Reflecting on the conduct of colleagues</td>
<td>3.72 (0.67)</td>
<td>3.78 (0.62)</td>
</tr>
</tbody>
</table>

* = significant at the α = .01 level
RESULTS

Knowledge acquisition

The AL cohorts scored significantly higher on 10 (2001-2002; ES 0.30-0.57) and 14 progress tests (2002-2003; ES 0.27-0.66) and significantly lower on 1 progress test (2001-2002; ES 0.31) than cohorts from the other two medical schools. The CBAL cohorts scored significantly higher on 2 progress tests (2003-2004; ES 0.30 and 0.34) and significantly lower on 2 (2003-2004; ES 0.24 and 0.27) and 4 progress tests (2004-2005; ES 0.23-0.44) than cohorts from the other two medical schools. (Figure 1). None of the 4 cohorts scored significantly different on the last three tests of the final year.

Clinical Performance

We did not find a significant difference between the clinical performance of students from the CBAL curriculum (Mean = 7.91; SD = 0.28) and the AL curriculum (Mean = 7.87; SD = 0.35; t(823) = -1.540; p = 0.124).

Perceived preparedness for medical practice

Of the CBAL and AL curriculum, 177 (48%) and 172 students (46%) completed the survey, respectively. Respondents and non-respondents were similar in gender distribution (74% and 70% female respondents, respectively) and mean clinical performance (Mean = 7.89; SD = 0.29 and Mean = 7.88; SD = 0.33, respectively). The internal consistency of the scales ranged from 0.70 to 0.86 (Table 2). Graduates from the CBAL curriculum felt excellently prepared for 10 and well prepared for 23 competencies. Graduates from the AL curriculum felt excellently prepared for 11 and well prepared for 22 competencies. Students from both curricula felt best prepared to treat a patient with respect and confidentiality (Mean_{AL} = 4.51; Mean_{CBAL} = 4.62) and felt worst prepared for following relevant legal regulations (Mean_{AL} = 3.33; Mean_{CBAL} = 3.37). At scale level, students felt excellently prepared for communication and well prepared in the other areas of competence. We found no significant differences at scale level. At item level, students from the CBAL curriculum felt better prepared for putting a patient problem in a broad context of political, sociological, cultural and economic factors (t(347) = -2.90; p = 0.004; ES = 0.31).
Figure 1a. Mean progress test scores of UMCG cohorts compared to those from two other medical schools: Mean scores (Y-axis) of the UMCG (dark grey line) cohorts from the AL curriculum (2001/2002 and 2002/2003) compared to the combined mean scores of the cohorts from two other medical schools (light grey line) on 24 progress tests (X-axis). A downwards arrow (↓) or an upwards (↑) marks the UMCG scoring significantly lower or higher than the other two schools, respectively.

Figure 1b. Mean progress test scores of UMCG cohorts compared to those from two other medical schools: Mean scores (Y-axis) of the UMCG (dark grey line) cohorts from the CBAL curriculum (2003/2004 and 2004/2005) compared to the combined mean scores of the cohorts from two other medical schools (light grey line) on 24 progress tests (X-axis). A downwards arrow (↓) or an upwards (↑) marks the UMCG scoring significantly lower or higher than the other two schools, respectively.
DISCUSSION

The aim of our study was to analyse the effects of the implementation of a competency-based active learning curriculum (CBAL) as compared to the previous active learning curriculum (AL). Using progress test results, we found relatively less knowledge acquisition in the first years of the CBAL curriculum than in the first years of the AL curriculum. However, we did not find such difference in the final year. Graduates who had been trained in a CBAL curriculum did not score higher on clinical performance nor did they feel better prepared for medical practice.

Implementing competency-based education requires that curriculum time is reserved for activities that facilitate competency development. As more time is allocated to the development of competencies, less time will be devoted to other curricular activities. In undergraduate curricula these activities usually involve knowledge acquisition. As a consequence, implementing a CBAL curriculum bears the risk of knowledge loss. We analysed students’ knowledge acquisition by comparing the scores of CBAL and AL cohorts on 48 progress tests to those of parallel cohorts from two other medical schools, which had not changed their curriculum during the time of our study. Our assumption was that if our students’ relative position remained unchanged, there would have been no knowledge loss. In comparison to the cohorts of the other medical schools, our AL cohorts scored significantly higher on 50% of the progress tests (24 out of 48), whereas our CBAL cohorts scored significantly higher on only 4% of the tests (2 out of 48). However, at the end of undergraduate education the CBAL and the AL cohorts demonstrated similar knowledge acquisition. The effect sizes of the differences were small to medium. As we interpret the outcomes concerning the progress tests as trends per cohort rather than results per test, we feel the effect sizes are large enough to conclude that students in the AL curriculum show higher knowledge acquisition than the students in the CBAL curriculum in the first years of their undergraduate education. Reserving time for competency development at the expense of time reserved for knowledge acquisition, seems to lead to lower knowledge acquisition in the short term, but not in the long term.

Throughout the medical curriculum, knowledge plays an important part in expertise development. As the CBAL cohorts seldom scored lower than the comparison cohorts and no long-term differences were found, we consider a permanent negative impact of implementing competency-based education on student learning and expertise development unlikely. An explanation for this finding might be that the clinical environment encourages students to regulate their own learning. During clerkships students are repeatedly stimulated to remedy deficiencies in medical knowledge. Undergraduate students’ prior knowledge deficiencies appear to be overcome during their clerkships.
We expected CBAL students to perform better in clinical practice than AL students. However, we did not find a significant difference, which may indicate that implementation of competency-based education has no effect on clinical performance. A possible explanation for this finding may be that all students must be competent to work with real patients at the start of their clerkships, which restricts differentiation among students. This homogeneity among clerks may explain why our clerks were mainly scored at the high end of the scale by their supervisors. Thus, we may have found no difference between the CBAL and the AL curriculum due to a restriction of range, caused by the requirements for entering the clinical phase.

We expected the CBAL students to feel better prepared for medical practice. To analyze students’ perceived preparedness we used survey data collected at graduation. The only difference we found between the two curricula is related to one of the core aims of competency-based medical education. Students from the CBAL curriculum felt better prepared to put a patient problem in a broad context of political, sociological, cultural and economic factors, which is in line with the aim to educate medical professionals who are sufficiently responsive to societal needs. It is also in line with the focus of competency-based medical education on the development of professionals in a societal context. However, we were unable to demonstrate any other effects of the implementation of competency-based education on students’ perceived preparedness.

The fact that we did not find a general increase in student’s perceived preparedness for medical practice may be related to the educational tools we implemented to facilitate competency development: portfolio use and explicit communication of competencies and their underlying framework. A recent study by Sargeant et al. revealed that explicit communication of competencies and the use of portfolios helps students to achieve informed self-assessment. Students in the CBAL curriculum are frequently informed of what is expected of them and they are explicitly stimulated to reflect on their performance, to remedy their deficiencies and to formulate points of improvement. The awareness that follows from these activities may help students to become increasingly conscious of their deficiencies. Possibly, CBAL students were more aware of their competencies and incompetencies than AL students, which is an important step in the development of competence. Consequently, the CBAL students may have underestimated their preparedness for practice as compared to AL students. Further research is needed to analyse the influence of implementing a CBAL curriculum on students’ reflectiveness and, subsequently, on their self-assessment.

A possible limitation of our study is that it is a single-site study, which affects the generalizability of our results. However, comparing curricula from the same institution has the advantage that most variables can be controlled. When the CBAL curriculum was introduced, teaching staff and learning methods remained largely unchanged. Consequently, our data have been gathered in the same context which increases the likelihood that possible effects can be attributed to the implementation of the CBAL curriculum. However, more studies are needed before generalizable conclusions can be drawn. Furthermore, our measurement of perceived preparedness had a limited response of
47%. However, the respondents and non-respondents were similar in gender distribution and clinical performance, which suggests that the sample was representative of the overall population.

Another limitation of our study might be that the measures we used – knowledge acquisition, clinical performance and perceived preparedness – are not specific to competency-based education. One could argue that for studying the effectiveness of competency-based education, measures are needed that fit conceptually. In our curricula, clinical competence was mainly assessed using global judgements. For research purposes, specific judgements may do more justice to the complexity of competencies. However, in this study such information was not available.

Finally, our study was limited to measurements during the course of undergraduate medical training and at graduation. Possibly, effects of competency-based education will become more apparent after graduation, in actual practice. Further research is needed to determine the long-term effects of implementing competency-based education at the undergraduate level. Despite the limitations of our study, we consider our outcome measures relevant because of their relation to performance in actual medical practice.\textsuperscript{16,21} Irrespective of the curriculum, medical graduates are expected to have sufficient knowledge and skills to practice professionally. Therefore, our study yields valuable information on the effect of implementing undergraduate competency-based education.

\textbf{CONCLUSION}

Implementing competency-based education in our undergraduate medical curriculum neither resulted in clerks who scored higher on clinical performance nor in graduates who felt better prepared for practice at the end of their training. Our study shows that there is some knowledge loss in the first study years of a CBAL curriculum as compared to the previous curriculum. Our study does not support the assumption that competency-based curricula result in graduates who are better prepared for medical practice. However, since this is one of the first studies in the field, it is too early to draw generalizable conclusions. More research is needed before we can conclude whether or not competency-based education meets the high expectations associated with its widespread implementation.
REFERENCES

Identifying essential clinical rotations for a competency-based curriculum

Wouter Kerdijk
Jos W Snoek
Martha J Dekker
Janke Cohen-Schotanus

Submitted
ABSTRACT

Introduction The current shift towards competency-based medical education implies that the focus of clerkships should be on competency development. Competency development has been suggested to require longer clerkship rotations. As a consequence of longer clerkships students may only rotate through a selection of disciplines in a competency-based curriculum. To safeguard the development all necessary competencies while only rotating through a selection of disciplines, we investigated the suitability of different disciplines for students to master 177 different competencies.

Methods Eight stakeholders in our medical curriculum, two medical students and six medical doctors from varying disciplines, were asked to judge the suitability of 13 different disciplines for students to master necessary competencies. They judged for each competency whether the discipline was ‘very suitable’ (+), ‘suitable’ (0) or ‘unsuitable’ (-) as a learning context. The scores off the judges were combined to determine a final suitability score per competency per discipline.

Results All disciplines were perceived as suitable for attaining most of the competencies (80.8%–100% per discipline). All disciplines were 100% suitable to develop competencies concerning ‘using basic knowledge and science’ and reflection. Community and occupational medicine and family practice were very suitable for developing a substantial number of competencies.

Discussion Almost every competency can be learned at almost any discipline, which implies that rotating through only a selection of clerkships is feasible in regard to competency development. The outcomes concerning family practice and community and occupational medicine suggest that these disciplines can play a central role in students’ competency development during clerkships.
INTRODUCTION

COMPETENCY-BASED MEDICAL education is increasingly being implemented in medical schools at the undergraduate level. So far, literature on competency-based medical education has mostly conveyed how to formulate competencies and create a comprehensive competency framework. Medical educators are confronted with the challenge of creating a curriculum that optimally facilitates the development and integration of competencies. In curriculum design, alignment between the educational goals and the programme is vital for students to learn effectively. When implementing undergraduate competency-based medical education this implies that the design of a clerkship programme has to revolve around competency development. Traditionally, clerkship programmes are discipline-based, focusing on rotating through as many disciplines as possible, while in competency-based curricula the focus is on the development of competencies, which requires longer rotations. This raises the question of how a shift from discipline-based to competency-based clerkships will affect the way a clerkship programme should be designed.

The clinical phase is vital for undergraduate medical education, probably even more so for competency-based medical education. During this phase, students develop and practise relevant competencies in a clinical context. Furthermore, they learn to integrate competencies by applying them simultaneously and complementarily in practice.

A shift in the purpose of clerkships towards competency development has important consequences for the clinical phase. Longer clerkship rotations are suggested to benefit students’ competency development. Each competency in a framework represents a complex interplay of knowledge, skills and professional behaviour. Mastering complex abilities requires more time than acquiring separate skills or knowledge. Furthermore, the development of complex abilities requires considerable cognitive resources. Each transition to another department and environment can lead to increased cognitive load. Too much cognitive load from transitions may interfere with students’ competency development, especially during their early weeks. Therefore, it is questionable whether short, one to two-week clerkships which are often part of discipline-based clerkship programmes are suitable for competency development. Another argument for longer rotations is that one of the aims of competency-based medical education is to teach students to function as professionals in a societal context. Theories on professional identity suggest that professional development in students requires considerable time spent in the same team or department.

At the basis of a competency-based curriculum stands a comprehensive framework of well-defined competencies which describes the curriculum’s end-goals. Therefore, shifting towards competency-based clerkships means that the purpose of clerkships shifts from mastering discipline-specific knowledge and skills to mastering a predefined set of
competencies. Consequently, students should have the opportunity to practise most or all competencies during clerkships. However, given the limited time available in the clinical phase, introducing longer rotations will limit the number of disciplines that students can rotate through. In addition, the number of students who can rotate through a discipline at a time is also limited, making it even harder for all students to rotate through a small selection of disciplines. Therefore, longer clerkships can only be implemented if most or all competencies can be learned in many different disciplines. This leads to the question how to safeguard the opportunities for students to master all relevant competencies in practice while also offering students rotations sufficiently long for the development of these competencies. In the current study, we investigated how a group of stakeholders in medical education perceive the suitability of different disciplines for students to master a predefined set of 177 competencies.

METHODS

Context
The current study took place as part of the development of a competency-based medical undergraduate curriculum at the University Medical Center Groningen (UMCG). Its competency framework was inspired by an early version of the CanMEDS framework and the Dutch general objectives for undergraduate medical education at the time. The framework consists of seven areas of competence: communication, clinical problem-solving, using basic knowledge and science, patient investigation, patient management, social and community contexts of healthcare, and reflection. Each area of competence was subdivided into applicable competencies, of which a total of 177 were specifically meant to be developed during clerkships (Table 1). Only four competencies in the area of ‘using basic knowledge and science’ were to be developed during the clinical phase because most competencies concerning this area were developed during a separate half-year scientific elective during which students write a thesis.

Procedure and analysis
Eight stakeholders in medical education were invited to participate in the study. The group consisted of two medical students near the end of their clerkships and six medical doctors with extensive experience in education. They were two gynaecologists, an internist, a neurologist,
Table 1. Definitions of the seven areas of competence in the competency-based active learning curriculum at the UMCG in Groningen.

<table>
<thead>
<tr>
<th>Area of competence</th>
<th>Definition</th>
<th>Number of competencies</th>
<th>Example competency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>The medical practitioner is competent to communicate effectively and efficiently in matters of patient management and professional attitude within the patient-care context and cooperation with other care providers.</td>
<td>29</td>
<td>The student is able to communicate relevant patient information orally and in writing.</td>
</tr>
<tr>
<td>Clinical problem-solving</td>
<td>The practitioner is competent to identify a problem and undertake adequate steps to solve it. He integrates this competency with any other competency and applies it in medical decision-making processes, which he can justify.</td>
<td>26</td>
<td>The student is able to revaluate and integrate information from patient problems, anamnesis, physical examination and further medical tests.</td>
</tr>
<tr>
<td>Using basic knowledge and science</td>
<td>The medical practitioner is competent to approach scientific information critically and form his opinion independently. Consequently, he adequately translates this information into medical policy.</td>
<td>4</td>
<td>The student is able to keep up with and select the relevant literature.</td>
</tr>
<tr>
<td>Patient investigation</td>
<td>The practitioner is competent to independently record the patient-oriented history, physical examination and additional diagnostics as well as interventions that are part of a treatment.</td>
<td>28</td>
<td>The student is able to assess a patient’s functioning in different aspects of his or her life.</td>
</tr>
<tr>
<td>Patient management</td>
<td>The medical practitioner is competent to manage the most common health problems of individual patients in relation to their living conditions and society. He is competent to name the goal of a treatment, list the potential treatments, reason the choice of treatment, monitor its effect and, if necessary, adjust it after consultation with third parties.</td>
<td>24</td>
<td>The student is able to refer a patient to the right person or service in primary or secondary care and determine the urgency.</td>
</tr>
<tr>
<td>Social and community contexts of healthcare</td>
<td>The medical practitioner is competent to place the patient’s individual problem (the complaint) in its social context and adjust to this. He is aware of the factors that affect health and illness and their consequences, both individually and collectively.</td>
<td>24</td>
<td>The student is able to explain the nature of a disease and its consequences for different aspects of the patients life.</td>
</tr>
<tr>
<td>Reflection</td>
<td>The medical practitioner is competent to reflect on his personal and professional contacts with patients and colleagues, on his work and on medical principles in a cultural context. On the basis of these reflections he will develop both as a professional and as an individual in the course of his career.</td>
<td>42</td>
<td>The student is able to identify his/her own blind spots and professional deficiencies in order to remove them.</td>
</tr>
</tbody>
</table>
a specialist in community and occupational medicine, and a general practitioner. The participants were asked to fill out a table with the 177 competencies listed against twelve disciplines on the other axis – 2124 cells in total. The disciplines were internal medicine, surgery, emergency medicine, paediatrics, obstetrics and gynaecology, psychiatry, neurology, revalidation, ‘otolaryngology, ophthalmology and dermatology’, family practice, nursing home medicine and community and occupational medicine. Otolaryngology, ophthalmology and dermatology were combined in the current study because of their similarities in the experience they offer (mostly outpatient care). However, each of these disciplines offers its own clerkship rotation.

Participants were asked to judge for each competency whether the discipline was ‘very suitable’(+), ‘suitable’(0) or ‘unsuitable’(-) as a learning context. Afterwards, the positive and negative scores from all participants were combined, resulting in a score for each cell. When at least half the participants had judged a clerkship very suitable or unsuitable for a specific competency, the cell was scored as such. When at least two positives and two negatives were present in one cell it was scored as conflicting.

RESULTS

All disciplines were judged to be suitable for attaining most of the competencies, with percentages varying between 80.79% and 100%. All the disciplines were deemed suitable for developing 100% of the competencies concerning ‘using basic knowledge and science’ and reflection. The percentage of competencies each discipline was judged very suitable for varied between 0% and 18.46% (Table 2).

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Very Suitable</th>
<th>Suitable</th>
<th>Not Suitable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Medicine</td>
<td>5%</td>
<td>94%</td>
<td>1%</td>
</tr>
<tr>
<td>Surgery</td>
<td>5%</td>
<td>95%</td>
<td>0%</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>5%</td>
<td>94%</td>
<td>1%</td>
</tr>
<tr>
<td>Paediatrics</td>
<td>1%</td>
<td>98%</td>
<td>1%</td>
</tr>
<tr>
<td>Obstetrics and Gynaecology</td>
<td>0%</td>
<td>99%</td>
<td>1%</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>3%</td>
<td>93%</td>
<td>4%</td>
</tr>
<tr>
<td>Neurology</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Revalidation medicine</td>
<td>1%</td>
<td>95%</td>
<td>5%</td>
</tr>
<tr>
<td>Otolaryngology, Ophthalmology and Dermatology</td>
<td>0%</td>
<td>98%</td>
<td>2%</td>
</tr>
<tr>
<td>Family practice</td>
<td>19%</td>
<td>81%</td>
<td>1%</td>
</tr>
<tr>
<td>Nursing home medicine</td>
<td>3%</td>
<td>92%</td>
<td>5%</td>
</tr>
<tr>
<td>Community and occupational medicine</td>
<td>10%</td>
<td>84%</td>
<td>6%</td>
</tr>
</tbody>
</table>
Community and occupational medicine was judged to be very suitable for the development of a substantial number of competencies (10.17%). This was mostly because community and occupational medicine was considered to be very suitable for mastering 50% of the competencies in 'social and community contexts of healthcare' (Table 3). The high suitability of family practice (18.46%) was because family practice was regarded as very suitable for 45% of the competencies concerning 'social and community contexts of healthcare' and 41% concerning 'social and community contexts of healthcare and occupational medicine was mainly judged to be unsuitable for developing 10% of the competencies concerning patient investigation (Table 3). Opinions about the suitability of a discipline for learning a specific competency were conflicting in 8 of 2124 cells (0.37%).

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Communication (% competencies)</th>
<th>Clinical problem-solving (% competencies)</th>
<th>Using basic knowledge and science (% competencies)</th>
<th>Patient investigation (% competencies)</th>
<th>Patient management (% competencies)</th>
<th>Social and community contexts of healthcare (% competencies)</th>
<th>Reflection (% competencies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Medicine</td>
<td>3</td>
<td>93</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Surgery</td>
<td>3</td>
<td>97</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Emergency Medicine</td>
<td>3</td>
<td>97</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Pediatrics</td>
<td>3</td>
<td>97</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Obstetrics and Gynaecology</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>14</td>
<td>83</td>
<td>3</td>
<td>96</td>
<td>4</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Neurology</td>
<td>14</td>
<td>83</td>
<td>3</td>
<td>96</td>
<td>4</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Revalidation medicine</td>
<td>0</td>
<td>93</td>
<td>7</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Otolaryngology, Ophthalmology and Dermatology</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Family practice</td>
<td>24</td>
<td>72</td>
<td>3</td>
<td>96</td>
<td>0</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>Nursing home medicine</td>
<td>24</td>
<td>72</td>
<td>3</td>
<td>96</td>
<td>0</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>Community and occupational medicine</td>
<td>10</td>
<td>83</td>
<td>7</td>
<td>96</td>
<td>4</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>
DISCUSSION

The key finding of the current study is that almost every competency can be learned at almost any discipline. This finding is important for those who intend to implement competency-based clerkships. The current shift in medical education towards the competency paradigm is an important impetus to increase the duration of clerkship rotations. Literature is clear on why clerkships should be longer, but as yet unclear on what selection of disciplines these longer rotations should be in. Curriculum designers are confronted with the difficulty of implementing longer clerkship rotations, while at the same time safeguarding the opportunity for students to develop all the competencies they need – often in circumstances where the number of medical students necessitate the use of as many departments of teaching hospitals as possible. The current study shows that most competencies can be practised and mastered during clerkships in almost any discipline.

An important implication of our findings is that – as far as competency development is concerned – the clinical phase can and probably should be designed in a less discipline-specific manner. Most competencies can be developed in almost any context. This outcome supports the current trend towards longitudinal attachments to one department or discipline. Longitudinal attachments would still offer the opportunity for students to develop all the necessary competencies while eliminating the current objections to the short rotational approach. Students would thus have time to engage with patients and the medical team, which should help them to develop as professionals.\textsuperscript{18,24}

Family practice and community and occupational medicine seem to be the disciplines which are most suitable for students to develop competencies concerning social and community contexts of healthcare. These concern competencies such as ‘the student is able to explain the nature of a disease and its consequences for different aspects of the patients life’ and ‘the doctor has knowledge of and insight into the importance of prevention of chronic incapacity for work and maintenance of ability to work’. Indeed, it seems intuitive that family practice and community and occupational medicine are disciplines that offer unique opportunities for students to practice such competencies. This finding is especially salient because an important reason for the present focus on competency-based medical education is the call for societal accountability and healthcare that matches society’s needs.\textsuperscript{19,28-31} One of the aims of competency-based education is that students understand and are able to work with social and community contexts in healthcare. Therefore, we feel that family practice and community and occupational medicine play a key part in implementing a competency-based clerkship programme.

We found that competencies concerning the use of basic knowledge and science and reflection can be learned equally well in any of the disciplines. All disciplines were judged suitable for learning all the competencies in these areas of competence, which might stem from their nature. Currently, medical education and medicine are both firmly
grounded in evidence-based practice. Therefore, any department should offer opportunities to practice competencies related to the use of basic knowledge and science. Reflection mostly concerns the students analysing and managing their experiences in practice. We feel that any department can offer ample opportunities for such reflection.

We acknowledge that the outcomes of this procedure are not necessarily generalizable to the context of other medical schools or other countries. The data we present concern 177 specific competencies based on the Dutch blueprint and are based on the judgments of stakeholders from a single medical school. Replication of the current study in other contexts is required to determine its generalizability. We feel the procedure presented here serves as an example of how to select clerkships based on a competency framework. It utilises the expertise present in the organisation and a well-picked group of stakeholders can help curriculum developers to ascertain broad support throughout a medical school. Furthermore, we feel that the general findings discussed above add to a basic understanding of the relationship between learning outcomes and curriculum design.

Our findings emphasize one of the advantages of formulating curricular goals as competencies. Current literature on competency-based medical education suggests that competencies should not be formulated too elementarily, but should rather reflect actual practice. Such generality allows for learning outcomes to be adaptable to different contexts in practice.

An important limitation of our study is that it focuses solely on competencies. Though it may be its most important purpose, offering students sufficient opportunities for competency development is not the only purpose of a clerkship programme. It should also offer a broad variety of patient groups and clinical problems to prepare students for a variety of postgraduate curricula and the full scope of the profession of medical doctor. We did not use the results from our procedure blindly. After the current study, its results were discussed in the project group responsible for curriculum design. Disciplines were selected if they were judged very suitable for developing at least six competencies (3.39%). Because no discipline was found to be particularly unsuitable (<5.65%), this score was not considered relevant for selecting disciplines. Consequently, family practice, community and occupational medicine, internal medicine, surgery, emergency medicine, psychiatry and nursing home medicine were initially selected as mandatory clerkship rotations. Nursing home medicine was not made mandatory because of logistical limitations. After discussion, paediatrics and neurology were added as mandatory disciplines because of their specific patient group and specific clinical picture of diseases, respectively. We provide this example to emphasize the importance of weighing factors in a group discussion even where the results may seem clear.

A narrow focus on competencies in clerkship design also carries with it another risk. Even though students are able to develop broadly-defined competencies, critical knowledge and skills concerning specific disciplines may not be developed when rotating through a limited number of disciplines. Furthermore, clerkships also serve the purpose of
career orientation. As they rotate through different disciplines, students familiarize themselves with the different paths they can choose after graduation. As such, clerkships can also serve to help students make deliberate decisions about their future careers.\textsuperscript{36,37} This is especially important since not working in the right field can lead to lower job satisfaction.\textsuperscript{38} One can easily imagine how less diversity in disciplines can reduce the chances for students to obtain a detailed picture of the field of medicine. Consequently, we feel that even though most competencies can be mastered in any discipline, diversity in disciplines remains important.

In conclusion, in this study we demonstrated how a competency framework can be used to classify clerkships according to their suitability to fit in a competency framework. We found that most disciplines are suitable for developing most competencies. Consequently, students do not need to rotate through specific disciplines to be able to develop all the competencies they require. This means that curriculum designers can implement longer clerkship rotations without including all disciplines and, thus, without lengthening the curriculum or making any discipline compulsory. Current literature presents arguments to favour a competency based view on clerkship programmes over a discipline-based view. The current study shows that most disciplines are fit to serve as building blocks for a competency-based clerkship programme.
REFERENCES


Acquisition and retention of discipline-specific knowledge during clerkships

Wouter Kerdijk
Jos W Snoek
Arno M M Muijtjens
René A Tio
Elisabeth A van Hell
Roy E Stewart
Janke Cohen-Schotanus

Submitted
ABSTRACT

Introduction Longer clerkship rotations are increasingly advocated in medical education literature. Due to limited time in the clinical phase, implementing longer rotations means that students can rotate through fewer disciplines. Little is known about the added value of visiting specific disciplines. We investigated the effect of clerkship rotations on students’ discipline-specific knowledge and whether it changes over time.

Methods To measure students’ (n=189) discipline-specific knowledge throughout their clinical phase, we gathered their item scores on twelve progress tests in five disciplines: gynaecology & obstetrics, paediatrics, neurology, psychiatry and family practice. For each discipline, we determined which progress tests individual students sat in each of the following conditions: before, during and after their clerkship rotation and, if applicable, during and after their second rotation in the same discipline. We compared students’ knowledge level in these conditions using a multilevel logistic regression for the probability that an item would be answered correctly, with items nested in progress tests which were nested in students. Additionally, we analysed the effect of the time passed since the clerkship on students’ knowledge, using the same multilevel logistic regression. Gender, time passed since the start of the clinical phase, item difficulty, entry level, and personal interest were added as covariates in all analyses.

Results We found a significant increase in discipline-specific knowledge during and after a corresponding clerkship rotation in all disciplines except family practice (odds ratio during and after vs. before= 1.33-2.77; p<.01). After completion of a clerkship, students showed a significant decrease in discipline-specific knowledge of 1-3% per month (p<.01) in three out of five disciplines.

Discussion Our study confirms that a clerkship rotation through a discipline generally increases students’ discipline-specific knowledge, however, it will decrease considerably after completion of the clerkship. Our findings suggest that longer rotations may negatively affect students’ knowledge about disciplines they do not rotate through as a consequence. However, given the significant decrease in discipline-specific knowledge after a clerkship, the impact of not rotating through a clerkship may be ameliorated.
INTRODUCTION

In current medical education literature, the traditional rotational approach to clerkships has been receiving criticism. There is an increasing emphasis on longitudinal attachments with patients, supervisors and departments to facilitate undergraduate students’ learning.1-3 As a consequence, Longitudinal Integrated Clerkships and hybrid models receive increasing attention and support.3-8 The underlying rationale is that longitudinal attachments facilitate learning. From a competency-based perspective, longitudinal attachments help students develop general competencies and attain high-quality formative feedback from their supervisors.1,9 From a socio-cultural perspective, learning could be more meaningful when students are immersed in a specific environment, which requires considerable time spent in that environment.1,2,9 Considering the limited duration of the clinical phase, implementing longer attachments will limit the number of disciplines students can rotate through during clerkships. This may impair students’ knowledge in disciplines they do not encounter. Insight into the relation between rotating through a discipline and students’ discipline-specific knowledge may help us take more informed decisions regarding clerkship design. Therefore, we examined the effect of following a clerkship on students’ discipline-specific knowledge development and retention during their clinical phase, for various disciplines.

The clinical phase is important for students’ knowledge development and retention. Students engage in workplace-based learning in which they do not only learn through modelling, but also through actively participating in situations that occur in the clinical learning environment.10-12 As students participate in medical practice, they apply their previously attained knowledge and, by doing so, develop new knowledge structures and gain a more complex understanding of the clinical problems they encounter.13 Furthermore, they are regularly challenged to actively retrieve discipline-specific knowledge, which has been shown to be one of the most powerful strategies for reinforcing knowledge retention.14,15

During clerkships, students’ discipline-specific knowledge will increase due to application and active retrieval of knowledge. However, this knowledge will probably not be fully retained. The accessibility of knowledge will generally decrease after a goal has been reached – for example, finishing a clinical rotation successfully.16-19 For knowledge to be retained, it needs to be reactivated regularly.14,20 When students rotate through a specific discipline, knowledge concerning other disciplines probably will be less activated and, therefore, it will decline. Knowledge is known to decline according to a forgetting curve in which a rapid initial decline is followed by a period of more gradual decline.17,21-24 Hence, when investigating how a clerkship affects discipline-specific knowledge, it is important to study the rate of knowledge decline after the clerkship rotation as well.
The call for more longitudinal attachments raises the question how not rotating through a specific discipline during clerkships affects students’ knowledge about that discipline. The effect may depend on the domain specificity of a discipline, because general medical disciplines offer a greater diversity of cases than specialized departments do. A general discipline will offer cases which are likely to be encountered at other departments. Not rotating through a general discipline may therefore be compensated by encountering relevant cases during rotations through other disciplines. As a corollary, one would expect students’ discipline-specific knowledge to increase, even when they have not rotated through a general discipline, and the decline of discipline-specific knowledge to be less pronounced due to regular retrieval. Cases concerning a more specialized discipline are less likely to be encountered in other departments. Therefore, one might expect that not rotating through a specialized discipline leads to hiatus in discipline-specific knowledge.

In the current study, we explored students’ knowledge development and retention during their clerkships. We investigated whether rotating through a discipline affects discipline-specific knowledge and whether and to what extent discipline-specific knowledge changes over time after the rotation.

**METHODS**

**Participants and context**

In the Netherlands, undergraduate medical education entails a 6-year competency-based curriculum that consists of a 3-year preclinical Bachelor’s programme and a 3-year clinical Master’s programme. In this study we gathered data about all 189 medical students who started their Master’s programme at the University of Groningen in September and October 2007 (Figure 1).

Figure 1. Schematic overview of the 3-year Master’s programme in medicine of the University of Groningen with clinical training center periods (CTC) clerkship rotations (grey blocks), progress tests (triangles), the scientific thesis and the clinical elective.

During their first master year students rotated through four ten-week modules. In each module, a five-week period in the clinical training centre is followed by a five week clinical rotation at the University Medical Center Groningen (UMCG) (Figure 1). During the second year they did 10 four-week clerkship rotations in one of six affiliated...
hospitals, sometimes a second rotation in a discipline. During the third year students spent 20 weeks on writing a scientific thesis and 20 weeks on a clinical elective in a discipline of their choice.

Throughout the Master’s programme students sat the Dutch interuniversity progress test four times per year. Each progress test is administered to all students of participating medical schools and consists of 200 multiple choice questions assessing knowledge in a variety of disciplines. Each discipline is assessed using a fixed number of questions per test (Table 1).

### Data Gathering
We measured students’ discipline-specific knowledge using their results on the 12 progress tests after the start of their Master’s programme. Additionally, we gathered data on several covariates. From the university administration we extracted information on students’ gender and entry level. Entry level was defined as obtaining the medical Bachelor’s degree nominally (within 3 years), or not nominally (>3 years), or with a biomedical Bachelor’s degree after a transitional year, as described by Cohen-Schotanus et al. We also collected data on the disciplines of the students’ clinical electives, as an indicator for personal interest.

### Analysis
From the disciplines assessed by the progress test (Table 1), we first selected disciplines that were covered by sufficient items in each progress test (≥7) and that students could rotate through during clerkships. Subsequently, we excluded general internal medicine and general surgery, because departments offering rotations in these disciplines highly differed in their extent of specialization. For example, for an internal medicine clerkship students could rotate through a department of general internal medicine, but also solely through a gastroenterology or pulmonology department. Ultimately, we investigated four specialized disciplines: gynaecology & obstetrics, paediatrics, neurology and psychiatry. One general discipline was selected, family practice.

| Table 1. Number of questions per discipline of the Dutch interuniversity progress test. |
| Discipline | Questions per progress test |
| Surgery | 17 |
| Dermatology | 5 |
| Gynaecology & Obstetrics | 10 |
| Family practice | 20 |
| Internal Medicine | 29 |
| Paediatrics | 11 |
| Otorhinolaryngology | 4 |
| Neurology | 7 |
| Ophthalmology | 4 |
| Psychiatry | 16 |
| Social Medicine | 4 |

We performed the following procedure for each of the five selected disciplines. For each progress test we determined whether a student sat the test before, during or after the clerkship rotation. If students had done a second rotation in the same discipline, we additionally determined when they sat progress tests during or after their second rotation. Accordingly, each progress test was sat in one out of 5 conditions: before any rotation, during a first rotation, after a first rotation, during a second rotation and after a second rotation. Too few students sat a test during a
second clerkship rotation to achieve enough power for our analysis and, therefore, this condition was excluded from analysis.

We analysed the data using a multi-level model with answers to items nested in progress tests which, in turn, were nested in students. To analyse the effect of rotating through a clerkship on discipline-specific knowledge, we used multi-level logistic regression to calculate the relative odds that students answered a discipline-specific question correctly in each of the four conditions. To analyse knowledge retention after rotating through a clerkship we selected items answered after the first rotation. In addition, we used multi-level logistic regression to analyse how the odds of answering a discipline-specific question correctly were affected by the amount of time passed since the end of the rotation. Gender, item difficulty, entry level, and personal interest were added as covariates in both analyses. Item difficulty was derived from the results of all master students of all medical schools participating in the Dutch interuniversity progress test. Personal interest was a binary variable, indicating whether a student had chosen the specific discipline as a clinical elective in the third master year. Furthermore, time passed since the start of the clinical phase was also added as a covariate to correct for advancement in the curriculum, since knowledge as measured by the progress test generally increases over time, independent of a specific intervention.²⁷, ²⁹-³¹

**RESULTS**

For each discipline except family practice, the odds that students would answer a discipline-specific question correctly was 1.33 to 2.70 times higher during and after a first and after a second clerkship, than before rotating through the specific discipline (p<0.01; Table 2). For family practice the odds during a clerkship did not differ significantly from the odds before the clerkship, while the odds after a clerkship were significantly lower than the odds before (odds ratio after vs. before =0.84; p<0.001). After rotating through a first clerkship the odds were lower than during the clerkship, for all disciplines except paediatrics (odds ratio after vs. during between 0.60 and 0.75). After a second clerkship odds for a correct answer were significantly higher than after a first clerkship for neurology (odds ratio after 2nd vs. after 1st =1.43; p<0.01) and paediatrics (odds ratio after 2nd vs. after 1st =1.38; p<0.01).

Discipline-specific knowledge was not affected by time passed since the clerkship for psychiatry and family practice. For the other disciplines, the odds for answering a question correctly decreased significantly with 1-3% per month (p<0.01).
Table 2. Odds ratio of a correct answer to a progress test question in a specific discipline during and after a clerkship rotation and after a second clerkship rotation versus a reference condition.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Reference condition</th>
<th>During first rotation Odds ratio</th>
<th>After first rotation Odds ratio</th>
<th>After second rotation Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gynaecology &amp; Obstetrics</td>
<td>No rotation</td>
<td>2.70*</td>
<td>1.62*</td>
<td>1.92*</td>
</tr>
<tr>
<td></td>
<td>During rotation</td>
<td></td>
<td>0.60*</td>
<td>0.71*</td>
</tr>
<tr>
<td></td>
<td>After rotation</td>
<td></td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>Paediatrics</td>
<td>No rotation</td>
<td>1.39*</td>
<td>1.33*</td>
<td>1.82*</td>
</tr>
<tr>
<td></td>
<td>During rotation</td>
<td></td>
<td>0.96</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>After rotation</td>
<td></td>
<td>1.38*</td>
<td></td>
</tr>
<tr>
<td>Neurology</td>
<td>No rotation</td>
<td>2.43*</td>
<td>1.64*</td>
<td>2.35*</td>
</tr>
<tr>
<td></td>
<td>During rotation</td>
<td></td>
<td>0.68*</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>After rotation</td>
<td></td>
<td>1.43*</td>
<td></td>
</tr>
<tr>
<td>Family practice**</td>
<td>No rotation</td>
<td>1.12</td>
<td>0.84*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>During rotation</td>
<td></td>
<td>0.75*</td>
<td></td>
</tr>
<tr>
<td>Psychiatry</td>
<td>No rotation</td>
<td>1.87*</td>
<td>1.36*</td>
<td>1.53*</td>
</tr>
<tr>
<td></td>
<td>During rotation</td>
<td></td>
<td>0.73*</td>
<td>0.82</td>
</tr>
</tbody>
</table>

* = Significant at the α = 0.01 level
** = Comparisons with ‘during a second rotation’ are missing because students can only rotate once through family practice

DISCUSSION

In this study we found a significant increase in discipline-specific knowledge after rotating through a corresponding clerkship in four out of five disciplines. This confirms the theory that retrieval of existing discipline-specific knowledge leads to better retention and that being engaged with discipline-specific cases stimulates knowledge development.13-15,18 We also found that after the rotation the odds to answer a discipline-specific question decreased in three out of five disciplines. These findings show that even though clerkships positively affect students’ discipline-specific knowledge, one should keep in mind that at the end of the curriculum a large part of the knowledge gained may not have been retained, with odds decreasing with 1-3% per month.

Our results suggest that most of the disciplines we investigated are indeed specific and, therefore, a corresponding clerkship rotation has a pronounced effect on students’ discipline-specific knowledge. This is further underlined by the fact that we did not find an effect of rotating through family practice, which is known as a broad discipline. Family practitioners work with a very diverse case load.32,33 Therefore, many cases involving knowledge specific to family practice can be encountered in other disciplines as well. This may explain why rotating through general disciplines did not cause an improvement in discipline-specific knowledge. That the odds to answer a question correctly were
even significantly lower after than before the rotation is probably caused by the correction for general knowledge growth. Given the broadness of family practice, it may well be that declarative knowledge concerning family practice is broader than the questions on the topic in the progress test.

We did not find a significant decrease after a rotation in paediatrics. This may be caused by the relatively small odds increase we found during the clerkship. It implies that rotating through paediatrics may not have improved students discipline-specific knowledge as much as rotating through the other disciplines, which means that students have less knowledge to be retained. Alternatively, this outcome may indicate that paediatrics is not as specialized as the other disciplines but less general than family practice.

Our analysis shows that the decrease was 1-3% per month for the disciplines where decreasing odds were observed. This implies that after one year students have knowledge retention of approximately 70-90% and after three years 30-70%. These figures suggest that a large part of the body of knowledge developed during a clerkship rotation at the beginning of a three-year curriculum may be lost at graduation. However, this finding should be interpreted with caution. On the one hand, we used progress tests to measure knowledge in order to achieve some kind of standardization and comparability between measurement points. However, these tests are taken only four times per year, which means that we have few measurements at or just after the end of a clerkship rotation. Since knowledge most rapidly declines just after a task is finished, we may have missed the most rapid decline and, consequently, may underestimate the decrease in knowledge. On the other hand, relearning information that is forgotten is more effective than learning completely new information. Therefore, we do not presume a direct relationship between the benefit of clerkship rotations and knowledge retention.

Concerning the added value of a second clerkship rotation our results are ambiguous. The results from paediatrics and neurology do suggest that the odds to answer a question correctly are significantly higher after a second than after a first rotation. However, this was not found for gynaecology & obstetrics and psychiatry. This possibly stems from differences between the first two and the latter two disciplines, and may be attributable to students’ motives for a second clerkship. It may be that second rotations in paediatrics and neurology are more often chosen to mend deficiencies, while gynaecology & obstetrics and psychiatry are more often chosen out of interest. Alternatively, the difference may stem from the extent of specialization. Possibly, more cases related to paediatrics and neurology can be encountered outside these disciplines than cases related to the other two. Which factor makes the difference, however, is unknown. Further research is needed into what determines the added value of a second clerkship rotation in the same discipline and, if so, whether this positively affects discipline-specific knowledge.

An important limitation of this study is that we solely focused on factual declarative knowledge. However, research on the development of medical expertise suggests that during medical training a profound
change occurs in the way knowledge is structured.\textsuperscript{13, 36} Elaborate causal networks, mainly accumulated during the preclinical years and consisting of declarative knowledge about causes and consequences of diseases, gradually become encapsulated into diagnostic models that are used to explain symptoms and signs. As a result of experiences in clinical practice, this encapsulated knowledge is then reorganized into highly personalized, narrative structures, called ‘illness scripts’, which focus less on biomedical facts and more on clinical information.\textsuperscript{13, 36} Furthermore, in particular during their clerkships medical students learn to integrate knowledge, skills and professional behaviour.\textsuperscript{37} Knowledge is only part of the competencies students develop during their training.\textsuperscript{38, 39} However, we are just starting to assess competencies and competency development.\textsuperscript{40-42} Our medical school does not have periodical standardized tests for the assessment of competencies available yet. Therefore, we limited our study to declarative knowledge.

This study was also limited to a single medical curriculum. All first-year rotations took place in the UMCG and students rotated through the second-year clerkships in one of six affiliated hospitals. Therefore, we expect that the diversity – in for example teachers or departmental cultures – was well sampled.

How we plan our clerkship rotations affects students’ knowledge development. With a limited time in the clinical phase and increasing duration of clerkship rotations, rotating through one discipline means not rotating through another. The current study shows that not rotating through a discipline during clerkships will negatively affect students’ discipline-specific knowledge, because it will be significantly less. However, the relevance of this outcome should be discussed in the light of the findings that this knowledge will also decrease significantly over time on the one hand and that relearning knowledge is easier than initial learning on the other hand. Furthermore, our outcomes do not lead us to discard the idea of longer rotations. In our study we did not investigate the potential benefits of longer rotations nor did we investigate how clerkship rotations affect other educational outcomes, like procedural knowledge and competency development. A much broader perspective is needed before we can decide how to plan our clerkships. The current study, offers a piece of the puzzle.
REFERENCES


Cumulative assessment: strategic choices to influence students’ study effort

Wouter Kerdijk
René A Tio
Florentine (B F) Mulder
Janke Cohen-Schotanus

Published in: BMC Medical Education 2013;13:172
ABSTRACT

Background It has been asserted that assessment can and should be used to drive students’ learning. In the current study, we present a cumulative assessment program in which test planning, repeated testing and compensation are combined in order to influence study effort. The program is aimed at helping initially low-scoring students improve their performance during a module, without impairing initially high-scoring students’ performance. We used performance as a proxy for study effort and investigated whether the program worked as intended.

Methods We analysed students’ test scores in two second-year (n=494 and n=436) and two third-year modules (n=383 and n=345) in which cumulative assessment was applied. We used t-tests to compare the change in test scores of initially low-scoring students with that of initially high-scoring students between the first and second subtest and again between the combined first and second subtest and the third subtest. During the interpretation of the outcomes we took regression to the mean and test difficulty into account.

Results Between the first and the second subtest in all four modules, the scores of initially low-scoring students increased more than the scores of initially high-scoring students decreased. Between subtests two and three, we found a similar effect in one module, no significant effect in two modules and the opposite effect in another module.

Conclusion The results between the first two subtests suggest that cumulative assessment may positively influence students’ study effort. The inconsistent outcomes between subtests two and three may be caused by differences in perceived imminence, impact and workload between the third subtest and the first two. Cumulative assessment may serve as an example of how several evidence-based assessment principles can be integrated into a program for the benefit of student learning.
BACKGROUND

IN MEDICAL EDUCATION, the assertion that assessment drives learning evokes positive and negative reactions. Critics state that assessment stimulates learning for assessment rather than learning per se, or that assessment drives surface rather than deep learning. Others are more pragmatic and reason that if assessment drives learning, why not use it to stimulate learning? The common end-of-course test may negatively affect study effort, because students start preparing for a test three to four weeks in advance. Consequently, if a course lasts longer than three to four weeks, students will be less engaged with the content during the first part of the course, which may impair their learning. In this exploratory study, we present a cumulative assessment program which combines frequent testing, repetition of content and compensation among tests in order to stimulate students’ study effort.

In the preclinical phase, medical knowledge is often assessed by written tests. Students’ performance on written tests can be influenced by their study effort, which, in turn, can be influenced by characteristics of the assessment program. Test dates and deadlines determine when students spend time on test preparation and other academic tasks. Instead of studying from the beginning of a course, students tend to start studying when the test date comes closer, which is called academic procrastination. It is estimated that 95% of students procrastinate to some extent and up to 30% procrastinate to such an extent that they delay many of their tasks until just before or even beyond the deadline. Students, on average, start preparing for a test three to four weeks in advance. Consequently, regular tests every three to four weeks should support students to put continuous effort into their learning.

Repeated testing also encourages students to put effort into studying the same content repeatedly. Repetition of content has been demonstrated to improve retention. People learn and retain information better through repeated exposure. Actively retrieving content during a test strengthens retention even more. Consequently, for an assessment program to be effective, the same content should be repeatedly tested and assessment within a course should be organized in such a way that each test includes the study material from preceding tests.

When using multiple tests to assess the same content, it is advisable to combine test scores and allow for compensation between the tests within the course. Compensatory assessment enables students to compensate poor performance on one test with good performance on others. A major advantage of compensatory assessment is that students are not discouraged too much by initial poor test results, since there is still a possibility for repair, which encourages increased study effort. A possible disadvantage of compensatory assessment is that initially high scoring students might refrain from studying intensively for the next test. However, if each subsequent test has an increasing number of items, initial good test results will not guarantee a successful
final grade. This way, all students will have to keep studying to pass the entire assessment program. For a compensatory assessment program to be effective, a condition is that students receive information about their performance between the tests. This information should help students correct their errors and reinforce correct responses.\footnote{16-18} It should not be provided during a test or when other activities require students’ attention, but rather when students are in a position to actively process it.\footnote{18,19}

The cumulative assessment program under study is designed to encourage students to continuously study throughout a course. We expect students with an initial low test score to benefit from the program, because it offers them the opportunity to identify knowledge deficits and compensate initial poor performance with higher performance on subsequent tests. Frequent and repeated testing offers students the opportunity to repeatedly recall the course content and remedy their knowledge deficits. The cumulative assessment program can be expected to be less beneficial for students who scored high on the first test, since there is less room for improvement. However, frequent testing with an increasing number of questions and weight per test should stimulate high-performing students to keep putting effort into studying. Repetition of content should increase their retention as well and help them maintain their high scores. In summary, we expect the cumulative assessment program to benefit the performance of initially low-scoring students, without impairing that of initially high-scoring students. Therefore, we expected initially low-scoring students to improve their scores on subsequent tests and initially high-scoring students to retain relatively high scores.

**METHODS**

**Context**
The undergraduate medical curriculum of the University of Groningen comprises a three-year pre-clinical bachelor’s program and a three-year clinical master’s program. Cumulative assessment is implemented throughout the bachelor’s program.

The cumulative assessment program is applied to ten-week modules in which different content areas are integrated. All content of a module is assessed by one multiple choice test. The test is divided into three separate mandatory subtests scheduled at the end of weeks four, eight and ten of the module (frequent testing). Each subtest contains questions covering the content of all preceding weeks (repetition). The final grade is based on the total number of questions from the three subtests, and is calculated at the end of a module (compensation). Shortly after each subtest, information about students’ performance is provided through the digital learning environment by publishing the correct answers and the number of questions each student answered correctly.

The distribution of the content of a module over three subtests is based on a conceptual model, in which the content of each week is assessed using the same number of multiple choice questions. Each subtest contains an increasing number of questions, covering the content
of all preceding weeks. In Table 1 this model is specified for a test of 200 questions, covering each week with 20 questions. The first subtest contains 50% of the questions regarding the content of the first four weeks. The second subtest contains 25% of the questions about the content of the first four weeks and 50% of the questions about the content of weeks five through eight. The final subtest contains the remaining questions: 25% of the questions about the content of the first four weeks and 50% of the questions about the content of weeks five through eight, and all questions about the content of the last two weeks. This distribution of questions over subtests results in an assessment program in which students can compensate for low initial scores, without making one of the subtests superfluous for initially high-scoring students.

Table 1. Conceptual model of a 10-week cumulative assessment program: Conceptual distribution of questions over three subtests in a 10-week cumulative assessment program, each week being assessed with 20 questions.

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtest 1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>Subtest 2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>Subtest 3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>Total row</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>200</td>
</tr>
</tbody>
</table>

Participants

We used students’ test results from two second-year modules: modules 1 and 2 (n=494 and n=436, respectively) and two third-year modules: modules 3 and 4 (n=383 and n=345, respectively). The data were gathered during the time that, under Dutch law, educational studies were exempt from institutional board review. In accordance with the university privacy policy and Dutch law, data were derived from the student records and anonymized before analysis.

Analysis

To test our expectations we compared the score change between tests of initially high and low scoring students as a proxy for an increase or decrease in study effort. During the analysis we faced two challenges. First, we had to take into account regression to the mean. Regression to the mean is caused by random measurement error when the same participants are repeatedly measured. Based on this statistical phenomenon, one would expect the high-scoring group to have a lower score and the low-scoring group to have a higher score on a subsequent test, purely due to personal variation. To ensure that the results of our study were not caused by regression to the mean, we judged cumulative testing beneficial when the mean difference in test scores between two tests was larger for low-scoring than for high-scoring students (Figure 1a). When the direction of the mean difference of one group was positive and that of the other group negative, we compared the absolute mean differences.

Our second challenge was that, when comparing students’ performance on two different tests, differences in test difficulty might systematically bias the results. In our medical school, knowledge test items are teacher-made and checked in-house on face validity by a peer and an educationalist. Therefore, there was no a priori knowledge about the difficulty of the subtests available. Consequently, subtest difficulty could not be controlled and could vary substantially. All students in a
module took the same tests, so low and high-scoring students’ test scores should have been affected by test difficulty in the same way. However, during the interpretation of the comparisons between high and low-scoring students’ score change, we needed to take test difficulty into account because it may change the direction of the mean score change between two tests for one of the groups. If the second subtest is more difficult than the first one, we would expect both groups to decrease in score. If cumulative assessment has an effect, we would expect high-scoring students’ scores to decrease more than those of low-scoring students (Figure 1b). Similarly, if the second subtest is less difficult than the first one, we would expect an increase in scores of both groups and the low-scoring students to improve more, due to cumulative assessment (Figure 1c). We operationalized test difficulty as the average facility index of the items of the test – the proportion of students that sat the test that answered the question correctly.

We expected students to revaluate their performance and adjust their study behaviour after they received new information about subtest 2. Therefore, we identified new quartiles of low and high-performing students after subtest 2, based on the combined score on the first two subtests. Again, we used independent sample t-tests to compare the mean differences of the low and high-performing groups between subtests 1 and 2.

Figure 1. Expected score change for different test difficulties: Directions of expected score change for initially low scoring students (solid line) and initially high scoring students (dashed line) when the second test is equally difficult (a), more difficult (b) or less difficult (c) compared to the previous test used independent sample t-tests to compare the mean differences of the low and high-performing groups between subtests 1 and 2.
RESULTS

For each of the four modules, the difficulty level of each subtest is reported in Table 2.

Table 2. Difficulty per subtest per module expressed as the average proportion of questions answered correctly by all students who sat the subtests.

<table>
<thead>
<tr>
<th>Module</th>
<th>Subtest 1 (p)</th>
<th>Subtest 2 (p)</th>
<th>Subtest 3 (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.65</td>
<td>0.68</td>
<td>0.66</td>
</tr>
<tr>
<td>2</td>
<td>0.68</td>
<td>0.62</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>0.68</td>
<td>0.68</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>0.68</td>
<td>0.73</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Comparing the mean differences between subtests 1 and 2 of initially low and high-scoring students, we found significant differences in score change for all four modules. In modules 1, 3 and 4 the difficulty of the second subtest was only slightly higher than that of the first one. In these modules, we found the average improvement of low-scoring students to be significantly higher than the average decrease in high-scoring students' scores, which is in line with our expectations (Table 3). In module 2, both groups decreased in scores as expected based on the higher difficulty of subtest 2. On average, high-scoring students' scores decreased significantly more than low-scoring students' scores.

When we compared the mean difference between the combined subtests 1 and 2, and subtest 3, we found significant differences in modules 1 and 3 (Table 4). In module 1, where test difficulty was similar between tests, the scores of low-scoring students increased whereas those of high-scoring students' decreased. Contrary to our expectations, the decrease in scores was significantly higher in the high-scoring group than the small increase in scores in the low-scoring group. In module 3, the third subtest was less difficult than subtests 1 and 2. Therefore, both groups showed improvement between the first two and the third subtests. In line with our expectations, the scores of the low-scoring students increased significantly more than those of high-scoring students. Against expectation, we found no significant differences in score change between subtests 2 and 3 in modules 2 and 4.

Table 3. For initial low and high scorers in four modules: mean test scores and absolute difference and t-tests comparing their absolute mean difference in test scores between subtests 1 ($T_1$) and 2 ($T_2$).

<table>
<thead>
<tr>
<th>Module</th>
<th>Group</th>
<th>Students</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>Absolute difference</th>
<th>T-test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Initial low scorers</td>
<td>111</td>
<td>49.82</td>
<td>59.36</td>
<td>[9.54]</td>
<td>8.19</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>139</td>
<td>73.29</td>
<td>72.27</td>
<td>[-1.02]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Initial low scorers</td>
<td>100</td>
<td>53.17</td>
<td>52.12</td>
<td>[-1.05]</td>
<td>6.49</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>122</td>
<td>77.21</td>
<td>68.29</td>
<td>[-8.92]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Initial low scorers</td>
<td>82</td>
<td>50.78</td>
<td>59.01</td>
<td>[8.24]</td>
<td>2.52</td>
<td>.01*</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>109</td>
<td>71.58</td>
<td>65.71</td>
<td>[-5.87]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Initial low scorers</td>
<td>96</td>
<td>43.28</td>
<td>52.10</td>
<td>[8.82]</td>
<td>9.66</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>76</td>
<td>58.90</td>
<td>57.77</td>
<td>[-1.13]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*= significant at the $\alpha$ = 0.05 level
Table 4. For initial low and high scorers in four modules: mean test scores and absolute difference and t-tests comparing their absolute mean difference in test scores between the combined subtests 1 and 2 ($T_{1+2}$) and subtest 3 ($T_3$).

<table>
<thead>
<tr>
<th>Module</th>
<th>Group</th>
<th>Students</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>Absolute difference</th>
<th>T-test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial low scorers</td>
<td>133</td>
<td>54.30</td>
<td>55.19</td>
<td>0.89</td>
<td>-3.24</td>
<td>.00*</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>124</td>
<td>75.04</td>
<td>71.16</td>
<td>3.88</td>
<td>.97</td>
<td>.33</td>
</tr>
<tr>
<td>2</td>
<td>Initial low scorers</td>
<td>107</td>
<td>51.05</td>
<td>53.50</td>
<td>2.45</td>
<td>.97</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>110</td>
<td>74.02</td>
<td>75.57</td>
<td>1.54</td>
<td>7.00</td>
<td>.00*</td>
</tr>
<tr>
<td>3</td>
<td>Initial low scorers</td>
<td>106</td>
<td>55.08</td>
<td>62.23</td>
<td>7.16</td>
<td>.02</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>101</td>
<td>69.56</td>
<td>71.15</td>
<td>1.58</td>
<td>.02</td>
<td>.98</td>
</tr>
<tr>
<td>4</td>
<td>Initial low scorers</td>
<td>87</td>
<td>46.71</td>
<td>69.04</td>
<td>22.32</td>
<td>-.02</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td>Initial high scorers</td>
<td>83</td>
<td>58.93</td>
<td>81.27</td>
<td>22.35</td>
<td>.02</td>
<td>.98</td>
</tr>
</tbody>
</table>

* = significant at the $\alpha = 0.05$ level

DISCUSSION

In this study, we presented a cumulative assessment program that is strategically designed to influence student learning. We found evidence for our expectation that initially low-scoring students will improve their scores on subsequent tests while high-scoring students will retain a relatively high score. The effect was most obvious between the first and the second subtests. Between subtests 1 and 2, the scores of initially low-scoring students increased significantly more or decreased significantly less than the scores of initially high-scoring students decreased. Taking into account the difficulty of each subtest, we found support for our expectation in each module. Our finding suggests that our cumulative assessment program encourages low-scoring students to increase their study effort, while it stimulates high-scoring students to keep up their study effort.

The underlying assumption of our study is that students’ changes in test scores reflect their study effort. In the literature, test performance has also been linked to other factors such as learning strategies and deep learning.\textsuperscript{21-24} However, effective deep learning is associated with study effort and applying different learning strategies requires students to put in effort as well.\textsuperscript{21} Furthermore, a recent study has shown that the positive effect of factors such as deep learning and resource management on student performance is mediated by student participation, which is a
form of study effort as well. Further research should establish whether our results can indeed be attributed to an increase in study effort and whether cumulative assessment leads to more participation or other changes in study strategies.

The results between subtests 2 and 3 were less clear. We only found a significant difference in two out of four modules. The results for module 3 confirmed our expectation that initially low-scoring students would improve more than initially high-scoring students. The results for module 1 revealed that the scores of initially high-scoring students decreased more than the scores of low-scoring students increased. We did not find a significant difference in the other two modules. These varying findings may have been caused by general effects of assessment on learning behaviour. Recently, Cilliers et al. found that the imminence of assessment, the perceived impact of the test and the amount of workload associated with the test generally affect the way students learn for their exams. In our cumulative assessment program, compared to the first two subtests, the third subtest determines 50% of the final grade and covers the content of the entire module. Besides, there are only two weeks between subtests 2 and 3. One could imagine how students may perceive the third subtest differently than the first two, when it comes to imminence, impact and workload of assessment. Furthermore, with only two weeks left before the next test, students may not have been able to adjust their study effort after evaluating their deficits. We argue that these factors may have affected students’ learning behaviours more during their preparation for the third subtest than for the other two subtests. Perhaps, an increase in imminence, impact and workload of subtests may influence students’ performance and study behaviour more than the cumulative assessment program.

Our cumulative assessment program is well-grounded in theory and combines frequent testing, repetition of content and compensation among tests. Several studies report positive effects of repeated testing of content in isolated courses. In these studies, tests were added to the regular program of a single course and were not part of a formal assessment program. The beneficial effects of the other two aspects of our cumulative assessment program have mostly been established in laboratory studies and simulated classroom experiments. This study adds to the literature by investigating these principles in a naturalistic setting. Furthermore, our study was embedded in a formal assessment program, which raises the stakes for students and causes an increased ecological validity of our findings. However, our findings are limited to the extent that we cannot attribute them to any separate aspect of the program. Further research is necessary to understand the interplay and separate roles of these aspects in the cumulative assessment program.

The use of naturalistic data has other possible limitations. Both the student sample and the characteristics of modules and tests can be seen as potential sources of bias. To minimize the influence of such bias, we investigated four modules to see whether the results were the same for different modules. Furthermore, during the interpretation of our results we took regression to the mean and test difficulty into account. Indeed, any difference in test difficulty between two tests or between modules was the same for all students, which increased the validity of our outcomes.
The findings in this exploratory study about the effects of a cumulative assessment program seem promising and add to the evidence that assessment can be used to support student learning. We cannot be sure whether cumulative assessment stimulates deep learning or other beneficial learning behaviours. However, in over half of the tests, initially low-scoring students increased their performance, while initially high-scoring students did not equally decrease in their performance. This suggests that implementing a cumulative assessment program may benefit students’ study effort and test performance. To support this evidence, an experimental design in a high stakes setting could help to further establish the value of cumulative assessment for educational practice.

CONCLUSION

The cumulative assessment program under study seems to influence study effort positively. How its influence may be mediated or moderated by the perceived imminence, impact and workload of the test requires further investigation. Based on our findings, we argue that implementing a cumulative assessment program may benefit students’ study progress. Furthermore, we feel that cumulative assessment serves as a good example of how several evidence-based principles of assessment can be integrated into a program that benefits students’ learning.
REFERENCES

18. Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Mem Cognit 2003;31:560–70.


Cumulative assessment versus assessment at the end of a course: effects on self-study time and test performance
ABSTRACT

Introduction Academic procrastination is one of the factors associated with study delay. Cumulative assessment uses distributed learning in order to stimulate students to study more frequently throughout a course and to prevent procrastination. In this random controlled study, we investigated how cumulative assessment may affect time spent on self-study and test performance as compared to end-of-course assessment. Furthermore, we investigated how students perceive both assessment methods to influence their learning.

Methods 78 students in a second-year course were randomly divided over two conditions: cumulative assessment or end-of-course assessment. At the start of the experiment, students completed a questionnaire about the extent to which either assessment method influences certain aspects of their learning. Students in the cumulative assessment condition were assessed in weeks 4, 8 and 10, and students in the end-of-course condition were only assessed in week 10. Each week, students reported the number of hours they spent on self-study. Performance was measured with 48 questions in week 10 in both conditions.

Results Students in the cumulative assessment condition spent significantly more time on self-study in all weeks (p<.05), except weeks 5, 9 and 10. Their self-study time peaked around exam weeks. Overall, they spent 69 hours more on self-study during the course than their peers in the end-of-course assessment condition (p<.05). Students in the cumulative assessment condition scored significantly higher on questions concerning the content of the last two weeks of the course. No differences were found in overall performance. Cumulative assessment was generally perceived to have a more positive influence on learning than end-of-course assessment.

Discussion Our study shows that cumulative assessment benefits student learning by encouraging students to distribute their learning activities over the course and preventing procrastination. Our findings also suggest that cumulative assessment leaves students more room to study content of the last part of the course. Cumulative assessment can be a useful tool to guide students’ study behaviour.
STUDY DELAY is a common problem in higher education.\textsuperscript{1,2} A factor that is associated with study delay is students’ tendency to procrastinate, i.e. postponing test preparation.\textsuperscript{3-5} Up to 30\% of students procrastinate, delaying their tasks until just before or even beyond the deadline, while even more students procrastinate slightly. Consequently, students’ self-study time decreases.\textsuperscript{6-9} Since assessment influences students’ study behaviour, frequent testing may help prevent procrastination.\textsuperscript{10-12} However, empirical evidence for this assumption is lacking. Cumulative assessment combines principles of distributed testing, repetition of content and compensation among tests in order to stimulate students to study regularly and benefit students’ test performance.\textsuperscript{13} In this randomized controlled study, we investigated how cumulative assessment affects students’ self-study time and test performance as compared to end-of-course assessment. Furthermore, we investigated students’ perceptions of the influence of both assessment methods on their study behaviour.

One of the main reasons why students postpone test preparation is that they experience the test date to be too far away.\textsuperscript{14} If there is a lot of time before the test date, students are likely to prioritize other activities such as work, friends or hobbies. This line of reasoning is supported by the finding that most students start test preparation three to four weeks before a test.\textsuperscript{10,14} As a corollary, students who procrastinate will run short on time to prepare for tests appropriately. The availability of sufficient self-study time in curricula is central to students’ study progress.\textsuperscript{2} However, students who procrastinate, try to master all knowledge of the course using only part of the available self-study time. Therefore, it is not surprising that procrastination is associated with lower grades and increased time needed to graduate a course.\textsuperscript{4,5}

A possible solution to the problem of procrastination may be to increase the number of tests in a course, in order to consistently decrease students’ perceived temporal distance to the tests. With cumulative assessment, students are frequently tested which should encourage them to study more regularly and prevent procrastination.\textsuperscript{10,12,14} Furthermore, each subsequent test has an increasing number of questions and students can compensate among tests, which should encourage them to further increase the time they spend on self-study.\textsuperscript{10,13,15} Therefore, we hypothesize that students who participate in cumulative assessment will study more and more frequently during a course than students who are only assessed at the end of the course.

Intuitively, one would expect that spending more time on self-study results in better test performance. Additionally, cumulative assessment may improve students’ test performance because each subtest assesses the content of all preceding weeks and, consequently, the same content will be repeatedly tested and studied.\textsuperscript{10} Earlier studies have shown that repeated testing of the same content can positively influence knowledge retention.\textsuperscript{16-20} Distributed testing and studying can also positively
influence knowledge retention.\textsuperscript{21-25} However, these benefits mostly concern long-term knowledge retention. On the short term, massed learning and distributed learning can yield similar test results.\textsuperscript{21-23} As a corollary, it is difficult to predict whether cumulative assessment will benefit students’ test performance. On the one hand, test performance may be higher in cumulative assessment, because of repeated testing and studying. On the other hand, students who participate in cumulative assessment and students who are only assessed at the end of the course may perform equally well, because distributed and massed learning can yield similar short-term results. Therefore, we investigate whether students who participate in cumulative assessment outperform students who are assessed only at the end of a course.

The implementation of educational measures, such as cumulative assessment aimed at regulating students’ study behaviour, can lead to resistance among staff and students.\textsuperscript{26,27} When we initially implemented cumulative assessment, we experienced such resistance. One purpose of this study was to find further support for the benefits of cumulative assessment. Several years after the implementation, we were interested in how students experience cumulative assessment as compared to end-of-course assessment. Therefore, we investigated how they perceive both assessment methods to influence their study behaviour and which method they prefer.

In summary, cumulative assessment combines several theories about assessment in order to stimulate students to study regularly, to prevent procrastination and to increase students’ test performance. In a randomized-controlled experiment we endeavoured to answer three research questions. First, what is the effect of cumulative assessment on self-study time? Second, what is the effect of cumulative assessment on students’ test performance? Third, how do students perceive cumulative assessment and end-of-course assessment to influence their study behaviour? Concerning the first research question, we hypothesized that students who participate in cumulative assessment will spend more time on self-study and study more frequently during the course. For the second and third research questions, we did not formulate hypotheses.
METHODS

Participants and Setting
At the University of Groningen each year of the preclinical phase is divided into four integrated 10-week courses, addressing a combination of topics. Cumulative assessment has already been implemented in two out of four courses of the second year of the undergraduate medical curriculum. In the other two courses, students are encouraged to keep studying through PBL tutorials. The current study was conducted among students in the fourth course of the second year (n=395). In this course, lectures were combined with PBL tutorials and students were assessed at the end of the course. All students had experienced cumulative assessment as well as end-of-course assessment in earlier courses. The topic of the course was chronic illnesses (internal medicine). Before the start of the course, all students were invited by e-mail to participate in the experiment. Initially, 105 students responded and after information was provided 78 of them signed an informed consent form. They were randomly assigned to either the cumulative or the end-of-course assessment condition. When the experiment was finished, students received a 20 Euro gift certificate and got the opportunity to observe an open heart surgery.

Ethical approval for the experiment was obtained from the Ethical Review Board of the Netherlands Association for Medical Education (NVMO-ERB) and the institutional Board of Examiners. In order to minimize the influence of the experiment on students’ pass rate, we conducted the study in a course where students were already stimulated to keep studying using PBL tutorials.

Procedure
After informed consent was obtained, the participants were invited to fill out an online questionnaire. Based on their experience, we asked all students to respond to six items and indicate on a 10-point scale to what extent cumulative assessment and end-of-course assessment stimulated them to influence their study behaviour. Furthermore, at the end of each week they were asked to report the number of hours they spent on self-study. Students in the cumulative assessment condition sat three multiple choice tests at the end of weeks four, eight and ten. These exams consisted of 19, 28 and 48 questions, respectively. In each test the study content of all previous weeks was assessed. After each test, the right answers to the test and information about students’ performance were made available through the digital learning environment of the university. Compensation among tests was possible. Students in the end-of-course assessment condition sat one multiple choice test at the end of week 10. This test consisted of 95 questions, including the 48 questions of the third test in the cumulative assessment condition. These 48 questions were used as a measure of performance and covered the content of weeks 1-4, 5-8 and 9-10, with 10, 21 and 17 questions, respectively.
Analysis

We used t-tests to investigate differences between the two conditions regarding time spent on self-study per week and during the entire course. To assess the effect of cumulative assessment on students’ test performance, we compared both conditions in terms of percentages of correctly answered questions of weeks 1-4, 5-8 and 9-10 separately and of the total test, i.e. 48 questions. We analysed test performance with t-tests. For all comparisons between conditions, we calculated Cohen’s d as a measure of effect size. Effect sizes are considered small when they are lower than 0.3 and large when they are higher than 0.8 (Cohen). Differences in perception of cumulative assessment versus end-of-course assessment were investigated using paired t-tests.

RESULTS

Over the course of our study seven students dropped out of the cumulative assessment condition because they had missed one or more subtests. Furthermore, exploratory analysis revealed that students, who were retaking the course, were distributed asymmetrically across conditions. Because we expected retaking the course to influence performance and study behaviour, we excluded 7 and 2 students from the cumulative assessment and the control condition, respectively. The final test analysis was performed with 25 students in the cumulative and 37 in the end-of-course assessment condition. During the first four weeks, students in the cumulative assessment condition spent twice the amount of time on self-study than their peers in the end-of-course assessment condition. Differences between the groups were significant in each week (Table 1; Figure 1a). In weeks six through eight students in the cumulative assessment condition also spent significantly more time on self-study. In weeks five, nine and ten, we found no significant differences in self-study time. At the end of the ten-week course students in the cumulative assessment condition had, on average, spent 182 hours on self-study and students in the end-of-course condition 113, which is a significant difference of 69 hour with an effect size of 1.01 (Table 1; Figure 1b).
Chapter 6

Table 1. Means, standard deviations, T values and Effect sizes for differences in self-study time between students in the cumulative and end-of-course assessment conditions in a 10-week course.

<table>
<thead>
<tr>
<th>Week</th>
<th>Cumulative assessment condition</th>
<th>End-of-course assessment condition</th>
<th>T-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (hrs)</td>
<td>SD</td>
<td>Mean (hrs)</td>
</tr>
<tr>
<td>1</td>
<td>7.8</td>
<td>8.6</td>
<td>3.8</td>
</tr>
<tr>
<td>2</td>
<td>8.2</td>
<td>7.4</td>
<td>4.5</td>
</tr>
<tr>
<td>3</td>
<td>13.8</td>
<td>9.9</td>
<td>6.2</td>
</tr>
<tr>
<td>4</td>
<td>29.5</td>
<td>13.3</td>
<td>5.8</td>
</tr>
<tr>
<td>5</td>
<td>10.5</td>
<td>11.5</td>
<td>7.6</td>
</tr>
<tr>
<td>6</td>
<td>11.6</td>
<td>9.6</td>
<td>7.3</td>
</tr>
<tr>
<td>7</td>
<td>17.8</td>
<td>7.7</td>
<td>10.3</td>
</tr>
<tr>
<td>8</td>
<td>30.8</td>
<td>13.8</td>
<td>11.8</td>
</tr>
<tr>
<td>9</td>
<td>20.5</td>
<td>11.5</td>
<td>18.4</td>
</tr>
<tr>
<td>10</td>
<td>30.4</td>
<td>12.4</td>
<td>37.0</td>
</tr>
<tr>
<td>Course total</td>
<td>182.4</td>
<td>76.8</td>
<td>113.2</td>
</tr>
</tbody>
</table>

* = significant at the α = .05 level ** = significant at the α = .001 level

Concerning the 48 items assessing performance, we found that students in the cumulative assessment condition performed significantly better on the 17 test items about the content of the last two weeks of the course \( t(59.6) = -2.01; p < .05; ES = 0.52 \). No significant differences between the two conditions were found for the 10 and 21 items regarding the content of weeks 1-4 and 5-8, respectively (Table 2). No difference...
in overall performance was found (Table 2).

Students generally perceived cumulative assessment to stimulate them more in regulating their own learning than end-of-course assessment (Table 3). They also regarded cumulative assessment as more stimulating to plan self-study time, prepare for a test, repeat content, and study the content in detail. Furthermore, they perceived cumulative assessment to prevent procrastination more than end-of-course assessment. The students did not regard cumulative assessment as more stimulating critical thinking. Finally, 52 out of the 62 students preferred cumulative assessment over end-of-course assessment.

Table 3. Means, standard deviations and T values for differences in students' perceptions of cumulative assessment versus end-of-course assessment.

<table>
<thead>
<tr>
<th>Question (10-point scale)</th>
<th>Cumulative Assessment</th>
<th>End-of-course assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>the extent to which it...</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>stimulates planning self-study time</td>
<td>7.8</td>
<td>1.0</td>
</tr>
<tr>
<td>helps to prevent procrastination</td>
<td>7.8</td>
<td>1.1</td>
</tr>
<tr>
<td>stimulates repeated studying of the same content</td>
<td>7.3</td>
<td>1.3</td>
</tr>
<tr>
<td>stimulates to prepare for tests</td>
<td>7.6</td>
<td>0.8</td>
</tr>
<tr>
<td>stimulates to study the content in detail</td>
<td>6.8</td>
<td>1.2</td>
</tr>
<tr>
<td>stimulates critical thinking</td>
<td>6.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>

*= significant at the α = .05 level

DISCUSSION

The results of our study confirmed the hypothesis that students in the cumulative assessment condition distribute their self-study time more and spend more hours on self-study throughout the course than their peers in the end-of-course assessment condition. We also found a positive effect of participating in cumulative assessment on students’ test performance concerning the content of the last part of the course. Furthermore, 84% of all participating students preferred cumulative assessment over end-of-course assessment. In general, cumulative assessment was perceived to influence students’ study behaviour more positively than end-of-course assessment.

Our study revealed significant differences in self-study time between the two conditions in weeks 1-4 and 5-8. However, we did not find significant differences in week 5 and weeks 9-10. A possible explanation for not finding a significant difference in week 5 may be that students who participated in cumulative assessment took some time to unwind after preparing and taking the first test. This is even more likely because week 5 was furthest away from the subsequent test. Not finding a significant difference in weeks 9-10 can be explained by the fact that, in both conditions, students were preparing intensively for the final test during these weeks. Our findings provide evidence that cumulative assessment is an effective method to stimulate students to study regularly.
throughout a course and to prevent procrastination. Furthermore, our results confirm earlier findings that students’ test preparation increases drastically as test dates get closer.\textsuperscript{10,14,28} In the cumulative assessment condition, self-study time evidently peaked in the weeks the tests were taken – weeks four, eight and ten – suggesting that test dates influence students’ study efforts and students are just-in-time managers (Figure 1).

We found a significant difference in performance across conditions regarding items about the content of the last two weeks of the course. A possible explanation may be that students who participated in end-of-course assessment may not have had enough time to study all course material, because they were cramming. Students in the cumulative assessment condition had already studied the course material of the first eight weeks and probably had more time left to study the content of the last two weeks of the course.

We did not find differences in overall test performance between both conditions. Considering the large differences in self-study time across conditions, this finding is rather counterintuitive. However, it is in line with the notion that massed learning can be as effective as distributed learning in the short term.\textsuperscript{21-23} Since test preparation peaked just prior to the final test in both conditions, students in both conditions may have reached similar knowledge levels. It has been argued that the superiority of distributed learning over massed learning is that information is forgotten more slowly.\textsuperscript{21,23} Therefore, future research investigating the effect of cumulative assessment on long-term knowledge retention may still yield results in favour of cumulative assessment.

The PBL tutorials throughout the course may have interfered with our results concerning students’ test performance. PBL tutorials engage students, support self-directed learning and stimulate deep rather than surface learning.\textsuperscript{29,30} Furthermore, similar to cumulative assessment, PBL tutorials stimulate students to retrieve and apply acquired knowledge regularly throughout a course, which improves retention.\textsuperscript{31,32} Therefore, an effect of cumulative assessment on overall test performance may have been tempered by the PBL tutorials. A similar experiment in a more traditional, lecture-based course may determine whether this was the case.

Most students (84\%) preferred cumulative assessment over end-of-course assessment. They also perceived cumulative assessment to stimulate regulating their own learning. These findings suggest that initial resistance students may have experienced during the implementation of cumulative assessment is no longer commonly held. The reason that students did not regard cumulative assessment as more stimulating critical thinking, may be that this intervention is mostly aimed at time management and continuous engagement with the course content rather than the level of engagement and how students process the content.

A limitation of this study concerns the dropout we experienced in the cumulative assessment condition. During the experiment seven students dropped out during the study because they missed one or more
tests. This dropout rate is inherent to an experiment in which assessment programmes with single versus multiple tests are being compared. It is unclear how this influenced our findings. However, we feel that the medium to large effect sizes we found consistently across the course are too substantial to be explained away by bias stemming from dropout.

Another limitation is that we cannot attribute the effects in self-study time solely to distributed testing. Given the peaks in self-study time just prior to tests, it stands to reason that the increased self-study time in the cumulative assessment condition is related to the distribution of tests across the course. However, it remains unclear to what extent the other aspects of cumulative assessment influenced students’ study behaviour. For example, if the tests would not have been compensatory, students with initial bad results might have been discouraged and might not have studied as much during the rest of the course. The increasing number of questions in each subsequent test and the fact that each test assesses content of all preceding weeks may have been of influence as well. Further research is needed to unravel how each aspect of cumulative assessment influences students’ study behaviour.

The current study shows that cumulative assessment can prevent procrastination. Distributed testing in cumulative assessment has a significant effect on the time students spend studying during a course. Students take more time for self-study and distribute their learning more throughout a course. Students also perceived cumulative assessment to help them study effectively. Our findings suggest that cumulative assessment benefits students’ study progress, since procrastination has been negatively associated with study progress. Furthermore, the availability and use of sufficient self-study time in a curriculum is positively associated with study progress. Concerning test performance, we only found a positive effect of cumulative assessment with respect to the last part of the course. However, an earlier study yielded evidence that cumulative assessment helps initially low-scoring students improve their test performance, while it stimulates initially high-scoring students to keep up good performance. Additional evidence for an effect of cumulative assessment on test performance and long-term retention is needed to further determine its effectiveness and usefulness. For now, we conclude that cumulative assessment can be a useful tool to positively influence students’ study behaviour.
REFERENCES


General Discussion
IN THIS THESIS we investigated how educational change can be used to facilitate students’ knowledge and competency development. Comparing a competency-based curriculum and its predecessor, we found that students in both curricula performed equally well in clinical practice and felt equally well-prepared for practice at graduation. Students in a competency-based curriculum demonstrated a slight knowledge loss compared to students in a primarily knowledge-based curriculum. However, this difference was not found at the end of the curriculum. In the clinical phase we found that having students rotate through a selection of clinical rotations to promote longer rotations is feasible with respect to knowledge and competency development. Not rotating through a clerkship will lead to students missing out on discipline-specific knowledge. Furthermore, after finishing a clerkship, this knowledge decreases significantly over time. We also found that a group of stakeholders considered all disciplines suitable for the development of almost any competency. In the preclinical phase, we found cumulative assessment to benefit student study behaviour. Students who initially scored low or high in a cumulative assessment programme were shown to improve and maintain their scores, respectively. Furthermore, students in a cumulative assessment programme were shown to spend significantly more time on self study than students in an end-of-course assessment programme.

ON KNOWLEDGE DEVELOPMENT

Knowledge development runs like a thread through this thesis. It is addressed in four out of our five studies. The implications of the findings of each individual study are presented in their respective chapters. However, several findings and themes emerge when we look at the findings overall.

Our findings on cumulative assessment add significant evidence to the discussion of whether assessment can drive learning. Many have said that assessment should be used to drive learning.1–5 Opponents claim that evidence for this statement is lacking.4,6 In Chapters 5 and 6 we showed that cumulative assessment can be used to encourage students to study more regularly and study substantially more in general. Furthermore, we found evidence that, given enough time between tests, initially low-scoring students will improve their scores, while initially high-scoring students will maintain their high scores. These findings offer considerable support for the statement that assessment drives learning and show that cumulative assessment can be used to benefit at least initially low-scoring students.

The finding that cumulative assessment can drive student learning raises two important questions which should be addressed in further research. First, how does cumulative assessment drive student
learning? Cumulative assessment consists of a combination of different assessment principles, such as frequent testing, repetition of content and compensation across tests, and feedback between tests. It is uncertain which of these principles or what combination of these principles is most effective. The results from our experiment suggest that at least the planning of tests plays an important role in student study behaviour, as we observed a considerable increase in self-study time as test dates came closer. However, the other assessment principles might have contributed to this phenomenon. Alternatively, it could very well be that some principles influence when students study, while others, such as repetition of content and feedback, influence what students study. Further research should aim to unravel how the separate parts of cumulative assessment influence student study behaviour.

A second, related question concerns whether cumulative assessment, as we investigated it, is designed optimally to benefit student learning or whether there remains room for improvement. For example, research into knowledge retention has shown that retention is optimal when retesting knowledge is done at increasing time interval. Cumulative assessment in its current form features decreasing time intervals between tests: a four-week interval between the first and second test, and a two-week interval between the second and third test. Perhaps, starting with a shorter interval and then increasing the duration of the time interval as the course continues could benefit students’ knowledge retention more than the current system. Future research should determine what intervals lead to optimal results. Another example concerns the content and modus of the feedback students receive in cumulative assessment. In the current programme, students receive information on which questions they have answered correctly, through a digital learning environment. Its strength is that students can seek feedback at a time when they can process the feedback optimally. However, it is uncertain whether all students make use of this feedback. Literature suggests that structural and information-rich feedback is beneficial for student learning. Whether cumulative assessment would benefit from more structured modes of delivering feedback or more information-rich feedback could also be a topic for further research.

From the results in chapters 2 and 4 we learned that knowledge continues to develop during clerkships. This is in line with earlier research. It emphasises that students keep accumulating new declarative knowledge, even while simultaneously encapsulating existing knowledge and developing illness scripts. As students progress through the curriculum their knowledge is transformed into causal networks, and then illness scripts, which facilitate efficient and effective diagnostic reasoning and decision making. This process of transformation mostly occurs in practice where preclinical knowledge is applied to actual cases. Such complex knowledge structures are not measured by declarative knowledge tests and do not represent the increasing scores we found. Therefore, our findings suggest that the development of basic knowledge is an ongoing process, even at the end of undergraduate education. It raises the question of how knowledge develops throughout the medical education continuum.
In Chapter 2 we found that reserving time for competency development at the expense of time specifically reserved for knowledge development leads to a slight loss in knowledge development. Apparently, students in the competency-based curriculum developed their declarative knowledge to some extent more slowly than students in the previous curriculum. Intuitively, this finding could be thought to be related to the changes that were made in the clerkship programme. After all, one change was that students stopped rotating through every discipline during clerkships. Furthermore, in Chapter 4 we found evidence that students develop less discipline-specific knowledge when they do not rotate through a discipline. Consequently, part of the loss in knowledge development can be explained by the students acquiring less discipline-specific knowledge related to disciplines that they did not rotate through. However, we feel this is unlikely for two reasons. First, there were no differences in knowledge scores near the end of the curriculum in any of the cohort comparisons in Chapter 2. Second, when a student does not rotate through one discipline he is rotating through another. This means that students just develop knowledge of different topics rather than less knowledge.

ON COMPETENCY DEVELOPMENT

Concerning competency development, this thesis adds to the literature in two ways. First, by addressing knowledge development as an important condition for competency development. Each competency represents the integrated application of certain knowledge, skills and professional behaviour. Without appropriate knowledge, competency development would be impossible for students. Therefore, the findings described earlier concerning knowledge development also offer useful insights for those aiming to facilitate competency development. After all, practicing competencies, such as taking someone’s medical history, would lose all authenticity without the presence of the necessary knowledge structures to understand the patients’ complaints and illnesses.

Second, this thesis focuses on the suggestion that competency development can be facilitated by more longitudinal attachments during clerkships. Longitudinal attachments potentially allow students to build more meaningful relationships with patients, supervisors and other health care professionals.

Our studies in chapters 3 and 4 offer evidence that the increase in clinical rotation duration could be feasible. Under the rationale that longer rotations mean fewer rotations, we found support for the
feasibility of longer clerkships with respect to competency development. All disciplines were judged as suitable for learning most competencies. Furthermore, we found that the negative effects of not rotating through disciplines on students’ knowledge development may be mitigated. After all, even though students develop less discipline-specific knowledge, this knowledge also decreases significantly after a clinical rotation is finished. However, we also know that relearning information is a far easier process than learning it.\textsuperscript{24,25} This raises the question of whether and to what extent not rotating through certain disciplines as well as the forgetting that occurs after a rotation negatively affects the availability of such knowledge at later stages of medical students’ careers. One way to find out would be to investigate how much discipline-specific knowledge is accessible at several points after graduation.

Perhaps our most important finding regarding competency development is that most competencies can probably be learned in many different contexts. This is an encouraging finding as we want students to become broadly developed doctors, capable of working in an increasingly specialized, fragmented medical field.\textsuperscript{26} Its implication is that students can learn a comprehensive set of competencies in very different combinations of contexts. This offers possibilities with regard to creating the call for continuity in the clinical phase.\textsuperscript{27,28} It also potentially offers us ways to influence students’ career choices by offering them meaningful thematic combinations of rotations. Exposure to specific combinations of rotations and patient groups, for instance, could be used to increase the odds that certain students will choose careers in alignment with societal needs.

The question still remains whether longitudinal attachments are really beneficial for student competency development. As yet, empirical evidence for this widely supported statement is mostly lacking. Some evidence is emerging for a special form of longitudinal attachment in the clinical phase – longitudinal integrated clerkships. Longitudinal integrated clerkships are a very specific format where students are longitudinally attached to a general or family practice, from where they then longitudinally follow one or more patients at a time.\textsuperscript{29} This format for the clinical phase has been widely advocated recently and has been adopted in an increasing number of medical schools.\textsuperscript{30,31} Students in such clerkships perform equally well or better than students in rotational formats on knowledge and skills tests.\textsuperscript{29,32–35} Furthermore, students gain greater awareness of the patient experience and the healthcare system and they seem to be able to assess themselves more accurately.\textsuperscript{29,34,36} All these factors could contribute to better competency development, but whether they are actually successful in doing so is still uncertain. Furthermore, it has been correctly noted that most of the evidence for longitudinal integrated clerkships is preliminary and whether this format offers any long term benefits have yet to be established.\textsuperscript{37}

The recent interest in longitudinal integrated clerkships resonates with one of the key findings in Chapter 3 – that family practice is perceived as being very suitable for learning a considerable part of the competencies we want our students to master during the clinical phase. This is probably because primary care offers the most diverse caseload. Family practice may offer the benefits of longer attachment without the risk of a lower diversity caseload. Therefore, we feel that if longer
rotations are implemented, family practice should play a key role. Of course, the feasibility of such wide involvement of family practice is necessarily limited by local healthcare infrastructure and logistic limitations, such as the ratio between available family practitioners and the number of students.

A theme for future research concerns professional behaviour. As mentioned in the introduction of this thesis, what competency-based medical education brings new to the table is the notion that knowledge and skills should be meaningfully integrated with professional behaviour. As such, research into medical professionalism has received important impetus. However, in our curriculum study (Chapter 2), the focus on professional behaviour was new in competency-based education and there was no available baseline to measure any effect of its implementation on professionalism. As a corollary, it is uncertain how and whether competency-based curricula contribute differently to students’ professional development than other curricula. Future research could address this issue in two ways. First, medical schools who have not yet adopted competency-based education could carry out a pre and post measurement of professional development. Second, whether the standards that have been formulated in competency frameworks are truly met in practice and after graduation should be explored. Such research should help us ascertain whether our competency-based curricula achieve their goals.

**METHODOLOGICAL CONSIDERATIONS**

The studies in this thesis were mainly carried out in naturalistic settings, except the study in Chapter 3. Consequently, we mostly measured student outcomes in an actual high-stake educational setting. This increases the ecological validity of our findings, meaning they are more easily generalizable to real life educational settings and daily practice. However, naturalistic studies also have a downside. Given the specific context of each medical school and each healthcare system, it remains difficult to generate results which can be applied in every country or medical school. Furthermore, it can be very challenging to deduce whether an observed effect can be attributed to a planned intervention or whether other factors in the curriculum might have also contributed. It has been suggested that the latter is especially true of research at a curriculum level, such as the study in Chapter 2. We found few differences between the two curricula we studied and the effects that were found were hard to attribute to the curriculum change. However, we feel that even though it may be challenging to interpret the outcomes of research at the curriculum level, findings at this level can still be informative and guide us to meaningful research questions. For example, they raise the question of why explicitly paying attention to competencies did not increase students’ preparedness for practice or how students’ knowledge development can be maintained.
To some extent, the current thesis presents us with a streetlights and shadows problem. The streetlights and shadows effect refers to the fact that when we lose something, we often look for it where there is light, even though we may have lost it somewhere else. The studies in this thesis focus on very relevant measures, mostly regarding declarative knowledge, because such measures were available in the existing curriculum. However, other important pieces of the puzzle might have been left in the shadows because no tools were available to us to measure them on a large scale or to measure them at all. The progress test and the tests used in cumulative assessment address factual knowledge and basic relationships between concepts. However, the relationship between students' declarative knowledge and future performance in practice is indirect since students mainly develop expertise by applying their knowledge in practice. Therefore, acquiring sufficient declarative knowledge is no guarantee for becoming a good doctor, but is merely a prerequisite to construct more meaningful knowledge structures. Furthermore, in addition to declarative knowledge and expertise, students should also develop procedural knowledge, especially during clerkships. Procedural knowledge concerns knowledge about how certain tasks should be performed in order for them to be executed successfully. Both procedural knowledge and the development of more meaningful knowledge structures are important for students to become competent practitioners. Tools for measuring procedural knowledge and more complex knowledge structures could enrich further research into both competency-based education and knowledge development.

An important methodological lesson is that measuring knowledge using tests remains difficult without a priori information about characteristics of test items – item difficulty especially. Lacking knowledge about what a students' score on an item should mean can have a detrimental effect on the inferential power of any research design. Without knowledge about how difficult a test is, it is hard to say whether students performed well or not. Furthermore, lacking information on test difficulty can make the comparability of tests, indeed, daunting.

In our exploratory study on cumulative assessment (Chapter 5) we found a way of working around this. We described different hypotheses for different scenarios, correcting for both regression to the mean and fluctuating test difficulty. This methodology is an example of how we can deal with missing information about the difficulty of test items. Nevertheless, when researching student test performance, using test items of which the difficulty is known through previous measurements or for example an Angoff procedure is preferable. Unfortunately, naturalistic studies are often restricted to available outcomes measures. As a corollary, studies concerning students' test performance often have to rely on teacher-made tests.

In that respect and in others, the Dutch interuniversity progress test has been invaluable to our research. Our research demonstrates its utility for benchmarking. In our study, comparing two curricula, we showed how we used a benchmarking method adapted from Muijtjens et al. to compare our medical school to others, in order to say something about the effectiveness of our new curriculum. Our research into the effect of clinical rotations on discipline-specific knowledge (Chapter 4),
shows two other advantages of the progress test. First, the progress test is made according to a specific blueprint assessing a range of disciplines and topics with a set number of questions. Consequently, the progress test can be used to measure very specific areas of declarative knowledge over time and to investigate how these areas are affected by educational interventions. Second, the progress test offers information about the item difficulty from a much larger group than the students in a study sample, a student cohort or even a medical school. This offers researchers the opportunity to determine item difficulty relatively independently from the sample.

**PRACTICAL IMPLICATIONS**

Our findings have several implications for practice. First, cumulative assessment is a promising method for stimulating student test preparation in preclinical education. As shown in Chapter 6, cumulative assessment increases students’ self study time dramatically and consequently, their engagement with the study content, in comparison to end-of-block-assessment. Furthermore, the method stimulates students who perform badly initially to improve their scores later on. As such, combining several proven assessment principles has been shown to genuinely benefit student learning.

Our research also offers several recommendations concerning longer clinical rotations. First, as far as knowledge and competency development are concerned, longer clinical rotations can be implemented. Second, if we implement more longitudinal attachments we must consider which disciplines are vital for students to have some experiential knowledge of. Even though students may forget a large part of their discipline-specific knowledge, recalling forgotten knowledge is always easier than accumulating new knowledge and rotating through a specific discipline still yields a considerable increase in declarative knowledge. This does not necessarily imply that students should go everywhere, but this finding invites us to consider which topics students should really have some experience with. Third, our results showed that family practice
and public health and occupational medicine appear to be especially suitable for student competency development in patient management and social and community contexts of healthcare during the clinical phase. Therefore, we recommend making these disciplines mandatory when designing a competency-based clerkship programme that values these areas of competence.

A final implication of our findings is that competency-based education can be implemented without major educational losses. The flip side to this is that no gains were found either. However, we only studied traditional educational outcomes also available in a non-competency-based curriculum. Therefore, implementing competency-based education could yet be found to yield the benefits it promises.

**FINALLY...**

In the 1970s William Spady, in a reflection on contemporary primary and secondary education, coined competency-based education a ‘bandwagon without a definition’, indicating that everybody was getting on board and nobody really knew what they were doing.51 The field of medical education is more advanced. Since the start of this century, quite a few clear and influential articles have been published with meticulous definitions of what competency-based medical education should entail and how we should formulate competencies as learning outcomes.20,32-35 However, it has also been acknowledged that the competency movement is still in its infancy. Much remains uncertain about how competencies should be learned and how educators can facilitate this process.

Nevertheless, competency-based education is gaining ground every day and perhaps this is not surprising. Whatever its educational benefits, competency-based education has in many instances forced educators and curriculum developers to explicate and question what they want to teach their students and what they are actually teaching their students. Competency-based education offers a unique perspective on a curriculum in the sense that it makes its outcomes, which might have been there already, explicit.

This also implies that the specific effects of competency-based
education may be hard to detect. Much of what is now well defined may have already been part of the curriculum, implicitly or explicitly. Therefore, a future research agenda aiming to unravel what parts of our competency frameworks really are new and what parts were already taught implicitly before implementation may increase our insight into the consequences of implementing competency-based education. Subsequently, it may become easier to uncover the potential benefit of an educational paradigm that has already been widely adopted.
REFERENCES

26. Cassel CK, Reuben DB. Specialization, subspecialization, and
Chapter 7

44. Albanese M. Problem-based learning: why curricula are likely to show little effect on knowledge and clinical skills. Med Educ 2000;34:729–38.
51. Spady WG. Competency based education: a bandwagon in search of a


Summary
One of the main challenges for medical schools is to achieve structural alignment between educational goals, teaching formats and the assessment programme, because the quality of alignment ultimately determines the effectiveness of a curriculum. To achieve this alignment strategic choices are required. The studies in this thesis involve the strategic choices associated with the implementation of undergraduate competency-based medical education at the University of Groningen. The studies are set in a context of educational change, curriculum development and curriculum implementation. They address strategic choices in curriculum design that should help reach two important educational goals: knowledge development and competency development.

Chapter 1 illustrates the rise of competency-based education in the field of medical education. In brief, an overview is provided of the current consensus on competency-based medical education as well as the challenges that are still lying ahead regarding this topic. Much has been written about how to design competency frameworks and how to formulate competencies. However, very little is known about the cost and benefits associated with the implementation of competency-based education and which strategic choices are necessary for its effectiveness. Subsequently, an outline of the thesis is presented revolving around three questions. The first question concerns the design and effectiveness of an existing competency-based undergraduate curriculum. The second question regards the educational risks and benefits of longer attachments during clerkships. The third question addresses how strategic choices in assessment can be used to help students develop medical knowledge, when curriculum time is reallocated from knowledge development to competency development.

Chapter 2 addresses the design choices in competency-based undergraduate education and how its implementation influence educational outcomes. We investigated the impact of this curriculum change on students' knowledge development, clinical performance and perceived preparedness for practice at the end of their undergraduate programme. The biggest change that is required when implementing competency-based education is that time needs to be reserved specifically for competency development. Logically, this time needs to be taken from other curricular activities. During the implementation of a competency-based curriculum at the University of Groningen, the emphasis on active learning principles to stimulate students’ knowledge development was maintained. However, 15% of the curriculum time for education aimed at knowledge development was now allocated specifically for small group sessions aimed at competency development.

The first two cohorts of the competency-based curriculum (n=372) and the last two cohorts of the preceding curriculum (n=453) were compared on knowledge development, clinical performance and perceived preparedness for practice. During the first study years, the knowledge development of students in the competency-based curriculum lagged somewhat behind as compared to that of students in the previous curriculum. However, at the end of the curriculum no such differences were found. No significant difference in clinical performance was found. With regard to perceived preparedness for practice we
found that students from the competency-based curriculum out of 33 competencies, only felt better prepared for putting a patient problem in a broad context of political, sociological, cultural and economic factors. This finding is in line with the notion that competency-based education emphasises societal responsiveness of medical doctors.

In Chapter 3 a study is presented in which we explored the suitability of various medical disciplines for students to master a set of prescribed competencies. Literature suggests that longer clerkship rotations benefits students’ competency development. However, given the limited duration of the clinical phase, longer rotations will lead to less variety in disciplines students rotate through. This raises the question whether students can learn all prescribed competencies in a limited number of different rotations. We asked a group of eight stakeholders – six medical doctors from various disciplines with extensive experience in medical education and two students near the end of their clerkships – to judge the suitability of 13 different disciplines for students to master 177 competencies that have to be developed during the clinical phase. Based on their scores, we estimated the overall suitability of each discipline for students’ competency development. The results showed that all disciplines were considered suitable for students to develop most of the necessary competencies. All disciplines were considered suitable for developing all competencies concerning reflection and the application of science. Community and occupational medicine was a discipline that was considered particularly suitable for developing competencies concerning the social and community context of healthcare. Family practice was also judged very suitable for students to develop competencies concerning the social and community context of healthcare as well as competencies concerning patient management. Consequently, both disciplines can play a key part in clerkship programmes. The results of this study imply that a clerkship programme with longer, and consequently, a smaller number of rotations is feasible with regard to students’ competency development.

Chapter 4 addresses how rotating to fewer disciplines may affect knowledge about the disciplines that are not rotated through during clerkships. To assess the benefit of rotating through a discipline for students’ knowledge about that discipline we investigated the knowledge of 189 students concerning five disciplines over three years. The disciplines under study were gynaecology & obstetrics, paediatrics, neurology, psychiatry and family practice. Students’ knowledge was measured using their results on the Dutch interuniversitary progress test. For each discipline, we compared students’ discipline-specific knowledge before, during and after a clerkship and, if applicable, after a second clerkship in that discipline. We also investigated the effect of time on knowledge loss after a rotation was finished. These analyses were performed using a multilevel structure with discipline-specific test items nested in progress tests which were, in turn, nested in students. Several covariates, such as item difficulty, students’ aptitude and personal interest were taken into account. For all disciplines except family practice we found that rotating through the discipline generally increased students’ discipline-specific knowledge. For three out of five disciplines we found that this knowledge decreased significantly with time after the rotation, 1–3% per month. We concluded that longer clerkship rotations and, consequently,
rotating through a smaller number of disciplines may negatively affect students’ knowledge about certain disciplines. However, given the significant decrease in discipline-specific knowledge after completion of a clerkship, the impact of not rotating through a clerkship may be smaller in the long term.

When implementing a competency-based curriculum, part of the former curriculum time aimed at knowledge development will be reserved for competency development in the new curriculum. As a corollary, it may be necessary for students to develop the same amount of knowledge in less curriculum time. When less time is available, assessment may offer an alternative for helping students’ knowledge development. Chapter 5 describes a study investigating whether cumulative assessment supports students’ knowledge development. The purpose of cumulative assessment is to keep students studying by applying a combination of principles, such as frequent testing, repetition of content and compensation among tests. On top of that, because of its repeated testing cumulative assessment offers initially low-scoring students the possibility for repair in subsequent tests, which encourages increased study effort. In our study, we analysed undergraduate students’ test scores in 4 modules with cumulative assessment, two in the second and two in the third year of medical training (n = 1658). The students were assessed in weeks 4, 8 and 10 of the module. For each module, we compared the change in test scores of initially low-scoring with that of initially high-scoring students between the first and second subtest. Based on the combined scores of subtests 1 and 2, a new division of lowest and highest scoring students was established. Subsequently we compared the change in scores between the combined score on the first two subtests and the third subtest for initially low-scoring versus initially high-scoring students. For the first part of the modules (i.e. between the first and the second subtest), we found that cumulative assessment worked as intended. Initially low-scoring students improved their scores on subsequent tests while initially high-scoring students retained a relatively high score. The results between the second and third subtests were less clear. This may have been caused by the temporal distance between the second and third subtests, which was only two weeks. The temporal distance between the first and second subtest was four weeks. This may have resulted in differences in perceived imminence, impact and workload that students associate with the test. In general, the results suggest that cumulative assessment can be a useful tool in supporting initially low-scoring students without disadvantaging initially high-scoring students.

In Chapter 6 a randomized controlled experiment is described, comparing cumulative assessment with end-of-course-assessment in a 10-week module. We investigated two main purposes of cumulative assessment: (1) to prevent procrastination and have students spend sufficient time on their studies and (2) to support students in developing their knowledge. We divided 78 students randomly over the two assessment programmes. In the beginning of the experiment, students indicated which assessment programme they usually preferred and to what extent each programme influenced their study behaviour. Students in the cumulative assessment condition were assessed in weeks 4, 8 and 10, and students in the end-of-course condition were only assessed in
week 10. At the end of each week, students reported the number of hours they had spent on self-study. Performance was measured by 48 questions in week 10 which were exactly the same in both conditions. We found that students in the cumulative assessment condition spent significantly more time on self-study in most weeks. Their self-study time peaked in the weeks the tests were sat. On average, students in the cumulative assessment condition spent 50% more time on self-study (i.e. 69 hours) during the course than students in the end-of-course assessment condition. They scored significantly higher on questions regarding the content of the last two course weeks. However, the effect was too small to affect overall performance. Students perceived cumulative assessment to have a more positive influence on their study behaviour than end-of-course assessment. This study showed that cumulative assessment can help prevent procrastination among students. Our findings also suggest that cumulative assessment leaves students more room to study the content of the last part of the course. As such, cumulative assessment seems to be an effective tool to support students’ knowledge development.

Chapter 7 provides a brief summary of the main findings of this thesis and includes a general discussion in which our findings are considered in the light of knowledge and competency development in undergraduate medical education. Implementing competency-based medical education may slightly hamper students’ knowledge development. On the other hand, cumulative assessment is found to be a promising tool for its support. This implies, among other things, that our research on cumulative assessment supports the thesis that assessment drives learning. A distinction is proposed for mechanisms that affect when students learn and those that affect what students learn. Concerning competency development it is discussed among other things that longer clerkship rotations seem feasible with regard to our studies. Furthermore, most medical disciplines are suitable for the development of most competencies. To investigate the benefit of competency-based education its desired outcomes may need further operationalization. There are several methodological considerations and implications for medical education practice in relation to this thesis. An important point to consider is that our studies have high ecological validity because of a naturalistic setting, but this same setting may have allowed other factors than those under study to influence our outcomes. Also, most educational outcomes we investigated were already important before competency-based education and additional tools for measuring outcomes are needed. Yet another methodological consideration regards the difficulty of measuring knowledge in a reliable and valid way. In this regard we found the benefits of using an interuniversitary progress test for measuring declarative knowledge considerable. Finally, we conclude that competency-based medical education can be implemented at the undergraduate level without major educational losses. Evidence for gains associated with such an implementation is still lacking. We stress that competency-based education as it is now mainly focuses on the educational goals. As a corollary, many of the outcomes explicitly defined in competency-based curricula may have been already present, but undefined in preceding curricula. Investigating which parts of competency-based education are truly novel may help unravel the effectiveness of the competency approach to medical education.
Samenvatting
Eén van de belangrijkste uitdagingen voor elke opleiding is de afstemming tussen leeruitkomsten, werkvormen en het toetsprogramma. De kwaliteit van deze afstemming bepaalt immers de effectiviteit van een curriculum. Strategische keuzes zijn nodig om een goede afstemming te bewerkstelligen. De onderzoeken in dit proefschrift zijn gericht op de strategische keuzes bij de implementatie van competentiegericht onderwijs in de basisopleiding geneeskunde aan de Rijksuniversiteit Groningen. De onderzoeken vonden plaats in een context van onderwijsvernieuwing, curriculumontwikkeling en -implementatie. Specifiek betreffen de onderzoeken strategische keuzes in curriculumontwerp ter bevordering van kennis- en competentieontwikkeling – twee belangrijke doelen in het onderwijs.

Hoofdstuk 1 illustreert de opkomst van competentiegericht medisch onderwijs. Dit hoofdstuk biedt een kort overzicht van de huidige consensus over dit onderwerp en de uitdagingen die het competentiegericht onderwijs in de toekomst met zich mee zal brengen. Er is veel geschreven over hoe men competentiemodellen zou moeten ontwikkelen en hoe competenties geformuleerd zouden moeten worden. Er is echter nog maar weinig bekend over de kosten en baten van competentiegericht onderwijs en de strategische keuzes die nodig zijn om dit onderwijs effectief te maken. Vervolgens biedt het hoofdstuk een overzicht van de structuur van het proefschrift, aan de hand van drie vragen. De eerste vraag betreft het ontwerp en de effectiviteit van een bestaand competentiegericht curriculum. De tweede vraag betreft de onderwijskundige risico’s en voordelen van langere coschappen in de basisopleiding geneeskunde. De derde vraag richt zich op hoe strategische keuzes in toetsing de ontwikkeling van medische kennis van studenten kan bevorderen, terwijl er tijd voor kennisontwikkeling verdwijnt ten behoeve van competentieontwikkelingsonderwijs.

Hoofdstuk 2 gaat in op de ontwerpkeuzes die horen bij een competentiegericht geneeskundecurriculum en de invloed van deze keuzes op onderwijskundige uitkomsten. We onderzochten de impact van de implementatie van dit curriculum op de kennisontwikkeling en de klinische prestaties van studenten en hoe goed zij zich voorbereid voelden op de praktijk aan het eind van hun opleiding. Bij het implementeren van competentiegericht onderwijs is één van de grootste curriculumveranderingen dat er tijd in het curriculum gereserveerd moet worden voor competentieontwikkeling. Logischerwijs zal deze tijd worden afgetrokken van andere activiteiten in het curriculum. Tijdens de implementatie van het competentiegerichte geneeskundecurriculum aan de Rijksuniversiteit Groningen, bleef de nadruk liggen op actief leren om kennis ontwikkeling van studenten te stimuleren. Echter werd 15% van de curriculumtijd van dit onderwijs afgehaald ten bate van kleine groepsonderwijs gericht op competentieontwikkeling. De eerste twee cohorten van het competentiegerichte curriculum (n=372) en de laatste twee van het voorgaande curriculum werden met elkaar vergeleken wat betreft kennisontwikkeling, klinische prestaties en hoe goed men zich voorbereid voelt op de praktijk. Gedurende de eerste studiejaren, bleven studenten uit het competentiegericht curriculum achter in hun kennisontwikkeling in vergelijking met studenten uit het voorgaande curriculum. Aan het einde van het curriculum werden deze verschillen echter niet meer gevonden. Er werden ook geen
significante verschillen gevonden in klinische prestaties. We vonden dat, van de 33 competenties waar we naar vroegen, studenten uit het competentiegerichte curriculum zich alleen beter voorbereid voelden op het plaatsen van patiëntproblemen in een brede context van politieke, sociologische, culturele en economische factoren. Deze laatste bevinding is in lijn met één van de achterliggende ideeën van competentiegericht medisch onderwijs: dat dokters rekening houden met de verwachtingen van patiënt en maatschappij.

In Hoofdstuk 3 wordt een onderzoek beschreven naar de geschiktheid van verschillende coschappen voor de ontwikkeling van noodzakelijke competenties. De literatuur suggereert dat langere coschappen voordelig zijn voor competentieontwikkeling van studenten. De beschikbare tijd voor coschappen is echter beperkt. Langere coschappen zullen dus leiden tot minder verschillende coschappen, waardoor studenten kennis maken met een kleiner aantal disciplines. Dit roept de vraag op of studenten alle noodzakelijke competenties kunnen leren in een beperkt aantal verschillende coschappen. We vroegen een groep van acht stakeholders – zes dokters uit verschillende disciplines met veel ervaring in het medisch onderwijs en twee studenten die bijna aan het einde van hun coschappen waren – om 13 verschillende disciplines te beoordelen op geschiktheid voor de ontwikkeling van 117 verschillende competenties. Gebaseerd op de scores van de stakeholders werd de algemeen geschiktheid van de disciplines bepaald. Uit de resultaten bleek dat alle disciplines geschikt waren voor het ontwikkelen van het grootste deel van de 177 competenties. Alle disciplines werden geschikt bevonden voor het ontwikkelen van alle competenties met betrekking tot reflectie en omgaan met wetenschap. Sociale geneeskunde werd vooral geschikt geacht voor de ontwikkeling van competenties met betrekking tot omgaan met maatschappelijke context. Huisartseneeneeskunde werd ook geschikt geacht voor het ontwikkelen van competenties die te maken hadden met het omgaan met maatschappelijke context maar ook de competenties die te maken hebben met patiëntbehandeling. Beide disciplines kunnen dus een belangrijke rol spelen in coschapprogramma’s. De resultaten van deze studie impliceren dat een coschapprogramma, met langere en daardoor minder verschillende coschappen, haalbaar is wat betreft de ontwikkeling van competenties.

Het onderzoek in Hoofdstuk 4 betreft de invloed van het doorlopen van een coschap op kennis over dat coschap. Een vermindering van het aantal coschappen heeft mogelijk een negatief effect op de kennis van disciplines die de student misloopt. Om de invloed van het doorlopen van een coschap op discipline-specifieke kennis te onderzoeken, gebruikten we de gegevens van 189 studenten in 5 disciplines over 3 jaar tijd. Het onderzoek betrof de disciplines gynaecologie & obstetrie, kindergeneeskunde, neurologie, psychiatrie en huisartseneeneeskunde. Kennis van studenten werd gemeten met discipline-specifieke toetsvragen uit de Nederlandse interuniversitaire voortgangstoets Geneeskunde. Per discipline vergeleken we de discipline specifieke kennis van studenten voor, tijdens en na het coschap. Daarnaast onderzochten we het effect van tijd op het verlies van kennis nadat het coschap geëindigd was. De analyses werden uitgevoerd met een multilevelanalyse waarin discipline-specifieke toetsvragen genest waren binnen voortgangstoetsen, die op hun beurt weer genest waren binnen studenten. In de analyse werden
verschillende covariaten meegenomen, zoals moeilijkheidsgraad van de items, entreeniveau van de student, en persoonlijke voorkeuren. We vonden dat discipline-specifieke kennis toenam, wanneer studenten daadwerkelijk een coschap hadden gelopen in een discipline. Deze bevinding gold voor alle disciplines behalve huisartsengeneeskunde. We vonden dat kennis in drie van de vijf disciplines significant afnam na een coschap, met 1-3% per maand. We concludeerden dat een vermindering van het aantal coschappen negatieve effecten kan hebben op de kennis van sommige disciplines. De vraag is echter hoe ernstig deze effecten zijn, gegeven het feit dat discipline-specifieke kennis ook weer afneemt zodra een coschap doorlopen is.

Bij de overgang naar een competentiegericht curriculum een deel van de onderwijstijd voor kennisontwikkeling gebruikt moeten gaan worden voor competentieontwikkeling. Het gevolg is dat studenten evenveel kennis moeten verwerven in minder tijd. Wanneer er minder tijd beschikbaar is voor kennisonderwijs, zou toetsing een geschikt alternatief kunnen bieden om de kennisontwikkeling van studenten te ondersteunen. **Hoofdstuk 5** beschrijft een onderzoek waarin we onderzochten of cumulatief toetsen de kennisontwikkeling van studenten ondersteunt. Het doel van cumulatief toetsen is te zorgen dat studenten blijven studeren. Dit wordt gestimuleerd door de toepassing van verschillende toetsprincipes zoals reguliere toetsmomenten, herhaling van leerstof en compensatie tussen toetsen. Cumulatief toetsen biedt door het herhaalde toetsen ook de mogelijkheid voor initiële laag-scorende studenten om hun score te repareren met latere toetsen. Dit bevordert hun studieinzet. In ons onderzoek analyseerden we de toetsscores van studenten, in 4 blokken met cumulatief toetsen, twee blokken in het tweede en twee blokken in het derde jaar van de basisopleiding geneeskunde (n=1658). De studenten werden getoetst in week 4,8 en 10 van de module. We vergeleken de verandering in toetsscores tussen de eerste en tweede subtests van initiële laag-scorende met die van initiële hoog-scorende studenten. Op basis van de eerste twee toetsen werden de laag- en hoog-scorenders opnieuw gedefinieerd. Vervolgens vergeleken we de verandering in scores tussen de gecombineerde score van de eerste twee subtests en de derde subtest voor laag-scorende versus hoog-scorende studenten. In het eerste deel van de blokken (tussen de eerste en de tweede subtest), vonden we dat cumulatief toetsen werkte zoals verwacht. Studenten die initiële laag scores verbeterden hun scores op de tweede subtest. Studenten die initiële hoog scoorden behielden een hoge score. De resultaten tussen de tweede en derde subtest waren minder eenduidig. Een mogelijke oorzaak ligt in de korte tijd van twee weken tussen de tweede en derde subtest. De tijd tussen de eerste en de tweede subtest was vier weken. Dit verschil heeft mogelijk geleid tot verschillen in hoe studenten de nabijheid, impact en werklast van de twee toetsen ervaarden. Over het algemeen wekken onze uitkomsten de suggestie dat cumulatief toetsen ervoor zorgt dat initiële laag-scorenders hun scores verbeteren door cumulatief toetsen zonder dat dit de scores van initiële hoog-scorenders negatief beïnvloedt. Daarmee lijkt het een geschikte methode voor de ondersteuning van kennisontwikkeling.

In **hoofdstuk 6**, beschrijven we een gerandomiseerd experiment, waarin cumulatief toetsen wordt vergeleken met toetsing aan het eind
Samenvatting

Wij onderzochten of cumulatief toetsen het studeren stimuleert en uitstellen voorkomt. Daarnaast werd gekeken naar het effect op de kennisontwikkeling van studenten. We verdeelden 78 studenten willekeurig over een cumulatief toetsprogramma en een programma met alleen een toets aan het eind van het blok. Aan het begin van het experiment werd studenten gevraagd aan te geven welk toetsingsprogramma hun voorkeur had en hoe elk programma hun studie gedrag beïnvloedt. Studenten die cumulatief werden getoetst kregen een toets in week 4, 8 en 10. Studenten in de andere conditie kregen één toets in week 10. Aan het einde van elke week, gaven de studenten aan hoeveel tijd ze hadden besteed aan zelfstudie. Toetsprestatie werd in beide condities gemeten met dezelfde 48 vragen in week 10. We vonden dat studenten in het cumulatieve toetsprogramma significant meer tijd besteedden aan zelfstudie. De zelfstudietijd was het hoogst in de weken waarin werd getoetst. Studenten in het cumulatieve toetsprogramma werden gemiddeld 50% meer tijd aan zelfstudie dan de studenten die alleen getoetst werden aan het einde van het blok. Ze scoorden ook significant hoger op vragen over de stof van de laatste twee weken van het blok. Dit verschil was echter te klein om de algehele toetsprestatie te beïnvloeden. Studenten vonden dat cumulatief toetsen een grotere positieve invloed had op hun studiegerecht dan één toets aan het einde van een blok. Uit dit onderzoek blijkt dat cumulatief toetsen helpt met het voorkomen van uitstelgedrag onder studenten. Onze bevindingen impliceren ook dat cumulatief toetsen de studenten meer ruimte geeft om de inhoud van het laatste deel van een blok te bestuderen. Het lijkt er dus op dat cumulatief toetsen een efficiënt instrument is om kennisontwikkeling van studenten te ondersteunen.

Hoofdstuk 7 Geeft een korte samenvatting van de belangrijkste bevindingen in dit proefschrift. Deze bevindingen worden vervolgens beschouwd in het licht van kennis- en competentieontwikkeling in medisch onderwijs. Ons onderzoek suggereert dat de kennisontwikkeling enigszins belemmerd kan worden door de implementatie van competentiegericht onderwijs. Cumulatief toetsen daarentegen stimuleert juist de kennisontwikkeling. Ons onderzoek naar cumulatief toetsen ondersteunt het idee dat toetsing het leren stuurt. Voor verder onderzoek naar dit onderwerp stellen we voor om onderscheid te maken tussen mechanismen die beïnvloeden wanneer studenten leren en mechanismen die beïnvloeden wat studenten leren. Wat betreft competentieontwikkeling wordt onder andere bediscussieerd dat, op basis van onze uitkomsten, langere coschappen haalbaar lijken. Verder bleken de meeste disciplines coschappen te kunnen bieden die geschikt zijn voor het ontwikkelen van de meeste competenties. Om de potentiele voordelen van competentiegericht onderwijs verder te onderzoeken, is het nodig om gewenste uitkomsten van zulk onderwijs verder te operationaliseren. Verschillende methodologisch onderwerpen in dit proefschrift worden ook besproken. Een belangrijke punt ter overweging is dat onze onderzoeken aan de ene kant een hoge ecologische validiteit hebben door de naturalistische setting waarin ze plaatsvonden. Aan de andere kant kan een naturalistische setting ertoe hebben geleid dat externe factoren onze resultaten beïnvloedt hebben. Ook waren de meeste uitkomstmaten die wij hebben onderzocht al belangrijk voor de invoering van competentiegericht onderwijs. Nog een
andere methodologische overweging betreft de moeilijkheid om kennis te meten op een betrouwbare en valide wijze. De interuniversitaire voortgangstoets bleek wat dat betreft een zeer nuttig meetinstrument tijdens onze onderzoeken. Het hoofdstuk eindigt met de conclusie dat competentiegericht onderwijs geïmplementeerd kan worden in het medisch onderwijs, zonder grote onderwijskundige nadelen. Bewijs voor de onderwijskundige voordelen van competentiegericht onderwijs ontbreekt tot op heden. Echter willen we benadrukken dat het huidige competentiegericht onderwijs zich vooral richt op onderwijskundige doelen. Het gevolg is dat veel uitkomsten, die expliciet gedefinieerd zijn in competentiegerichte curricula, mogelijk al impliciet aanwezig waren in voorgaande curricula. Daarom stellen we onderzoek voor naar welke onderdelen van competentiegericht onderwijs werkelijk nieuw zijn. Mogelijk wordt het zo makkelijker de daadwerkelijke effectiviteit van competentiegericht medisch onderwijs te bestuderen.
Dankwoord
Het zit erop. Wat een avontuur. Eerlijk gezegd ben ik blij dat mijn promotietraject voorbij is, maar evenzo blij dat ik het gedaan heb. Wanneer ik het gehele traject zo op een weegschaal leg dan is het zeker de moeite waard geweest. Ik heb ontzettend veel geleerd de afgelopen jaren en heel veel leuke momenten meegemaakt met heel veel leuke mensen. Helaas heb ik niet genoeg ruimte in dit dankwoord om al die lol- en leermomenten op een rijtje te zetten. Gelukkig heb ik wel de ruimte om iedereen te bedanken iedereen te bedanken waarmee ik de afgelopen vier jaar dit avontuur heb beleefd. Maar ik wil ook nog een paar personen eruit lichten, te beginnen met mijn promotoren.

Janke Cohen-Schotanus, ik wil je bedanken voor de begeleiding en samenwerking tijdens mijn promotietraject. Voor een alleseter als ik, is het zeer inspirerend om met jou samen te werken, omdat je zoveel en zulke brede kennis hebt over het onderwijs, geïntegreerd in zowel de theorie als de praktijk. Ik heb ontzettend veel van je geleerd over beknopt en helder schrijven (natuurlijk maak ik in dit dankwoord graag van de gelegenheid gebruik om dit niet te doen). Ik denk met plezier terug aan onze sparsessies achter de computer.

Jos Snoek, als tweede promotor was je vanaf het begin erg betrokken bij mijn traject. Je hield de nieuwste ontwikkelingen in de literatuur in de gaten maar ook mijn sociaal-emotionele staat. Je hebt wel eens gegrapt dat jullie als twee ouders over mij hadden gesproken. Zo voelde dat soms ook wel een beetje. Ik wil je in het bijzonder bedanken voor je bevlogenheid over het opleiden van goede dokters. Deze bevlogenheid is zeer aanstekelijk en heeft er deels voor gezorgd dat ik het medisch onderwijs als veld altijd belangrijk en boeiend ben blijven vinden.

René Tio, ik wil je bedanken voor je betrokkenheid en scherpe commentaren. Je hebt je in alles verdiept waar ik mee bezig was, ook al ben je pas later bij mijn promotietraject betrokken geraakt. Ik vond het erg leuk dat je regelmatig even langs liep om te vragen hoe het ervoor stond en dat je ondanks je drukke schema in het ziekenhuis, altijd tijd had om even snel naar stukken of vragen te kijken.

Tineke Bouwkamp-Timmer, met je scherpe en analytische blik heb je mij geholpen met het afronden van nagenoeg elk hoofdstuk in dit boek, zowel inhoudelijk als taalkundig. Verder wil ik je bedanken voor je rollen als mijn vraagbaak, klaagmuur, sparpartner, kletsmaatje, coach, koffiemaatje, bibliothecaris en adviseur over eigenlijk alles. We hebben samen heel wat werk verzet en dat was altijd erg gezellig.

Naast mijn promotiecommissie wil ik graag mijn medeauteurs bedanken. Floor Mulder, Ally van Hell, Arno Muijtjens, Roy Stewart, Martha Dekker en Friso Muntinghe, allen bedankt! Het was erg fijn samenwerken. Jullie hebben elk op je eigen manier een unieke bijdrage geleverd aan de stukken in dit proefschrift. Twee personen wil ik in het bijzonder bedanken. Arno, toen ik bij je kwam met de vraag of je met ons wilde samenwerken leek dat je erg leuk, maar vergeet je dat je eigenlijk geen tijd had. Toch heb je die tijd er alsnog voor genomen. Een aantal keer heb je zeer kritisch gekeken naar de data en
naar het stuk. Beiden zijn hiervan zeker beter geworden. Floor, jou wil ik in het bijzonder bedanken omdat de onderzoeken naar cummulatief toetsen in dit proefschrift oorspronkelijk jouw project waren. Ik mocht daarbij aanschuiven en toen je naar Amsterdam vertrok liet je mij het onderzoek overnemen. Na je vertrek naar Amsterdam, heb je me ook steeds geholpen als dit nodig was. Ik kon je altijd mailen met vragen en kreeg vaak direct een helder antwoord. Zonder jou was dit proefschrift letterlijk half zo goed niet.

Dan mijn collega’s bij het CIOMO. Leo Aukes, Mirjam van Lohuizen, Anja Karg, Nienke Boendemaker, Tessa Waslander. Jullie zitten er allemaal niet meer, maar ik wil jullie bedanken voor een gezellige tijd en een fijne samenwerking. Karen Stegers-Jager, jou wil ik hier ook bedanken. De keren dat je ons kwam versterken vond ik altijd erg leuk en constructief. Johanna Schönrock-Adema, ik wil jou bedanken voor je hulp met lastige beslissingen over mijn onderzoek. Janet Raat, bedankt voor de oppépers op de slechtste momenten in mijn traject. Dan mijn roomies; Nienke en Nienke, respectievelijk Schripsema en Renting of vice versa, ik zie jullie nog steeds wekelijks wat ik ontzettend leuk vind. Ik vind het erg leuk dat jullie het af en toe leuk vinden om met mij over jullie onderzoek te praten. Also, I would like to thank Yoyo Suhoyo for a great time. I enjoy working with you and our meetings outside of work very much. The same goes for Dario Fernandes. I am glad we share an interest in research on assessment, which has led to many useful insights on the topic near the end of my trajectory.

Ook de collega’s bij het DOK wil ik bedanken. Jullie hebben het mij mogelijk gemaakt om regelmatig een kijkje te nemen in de praktijk van het medisch onderwijs en daardoor heb ik velen malen meer geleerd dan wanneer ik alleen in de boeken had gezeten. Ik heb erg genoten van onze discusses over onderwijs en jullie uitleg over hoe het er in de praktijk aan toe gaat (niet op de laatste plaats om mijn eigen uitkomsten te begrijpen). Ik heb met jullie allemaal een goede tijd gehad. Van het (toenmalige) DOK wil ik Marjolein Heijne-Penninga in het bijzonder bedanken. Door jou suggesties over mijn eerste artikel, toen het nog in de kinderschoenen stond, is het artikel vele malen beter en gewichtiger geworden. Daarnaast wil ik jou, maar ook Ally van Hell en Hanke Dekker hartelijk bedanken voor jullie peer support. Het is altijd fijn om promotielotgenoten om de hoek te hebben voor een praatje of advies.

Veel van het onderzoek in dit proefschrift is gebaseerd op studentgegevens uit onze eigen administratie. Ik heb hulp gehad van heel veel mensen om deze gegevens te verkrijgen. Ik ben hiervoor praktisch iedereen die werkte bij het curriculumbureau en het onderwijsbureau gedurende mijn promotie veel dank verschuldigd. In deze categorie horen ook Hilde Wiersma en Mirella Zuidema van het KTC. Bedankt voor jullie behulpzaamheid, snelle antwoorden en harde werk. Jullie hebben me niet alleen geholpen met het realiseren van mijn onderzoek, maar ook met het oplossen van allerhande mysteries in mijn datasets, waardoor onze bevindingen zuiverder zijn geworden. Mijn dank gaat in deze ook uit naar de verschillende affiliated ziekenhuizen die voor het onderzoek naar het effect van coschappen op kennisontwikkeling een belangrijk deel van de coschappaden hebben aangeleverd. Een bijzonder bedankje voor het Deventer ziekenhuis waar ik aan het begin van mijn
traject een aantal dagen heb kunnen meelopen om het onderwijs in de coschappen van dichtbij te zien.

Zoals de meeste mensen wel weten houd ik van afwisseling in mijn werk en kan ik niet uren achtereen achter mijn computer zitten om onderzoek te doen. Ik denk dat ik zonder leuke en leerzame nevenactiviteiten dit traject nooit had volgehouden. Een belangrijk deel van mijn afleiding vond ik in het geven en organiseren van onderwijs. Joke Fleer, Frank van Es, Harry Stuger en Hans Burgerhof, jullie wil ik bedanken voor de kansen die jullie me hebben gegeven om mee te denken en te helpen in het onderwijs. In het bijzonder wil ik Ineke Molenaar bedanken. Je gaf mij de kans mee te helpen aan de uitvoering en verbetering van de lijn wetenschappelijke scholing. Dankzij het behalen van mijn BKO bij jouw afdeling en onze samenwerking, heb ik inmiddels een ontzettend inspirerende nieuwe baan bij de opleiding tandheelkunde.

Wat betreft mijn andere nevenactiviteiten wil ik vooral de mensen van SHARE, de GSMS en het promovendinetwerk van het NVMO bedanken. Met de wisselende leden van de PhD councils heb ik erg leuke en vruchtbare bijeenkomsten gehad en de promovendidag organiseren was erg leuk. Daarnaast wil ik Truus van Ittersum en Renate Kroese bedanken voor hun prettige samenwerking. Natuurlijk ook een speciaal bedankje aan Marijke Sterman, van het NVMO. Zonder jou zou een gemiddelde promovendidag zeker in het water vallen.

Mijn vader zegt regelmatig ‘zonder ontspanning, geen inspanning’ en daarin heeft hij gelijk. Voor de nodige ontspanning tijdens mijn promotietraject wil ook een heleboel mensen bedanken. Sterker nog, de meeste mensen binnen het OWI hebben mijn gedachtenbrij regelmatig doorbroken, stopgezet of opgefleurd met een praatje. Mijn roomies, bedankt voor de vele ontspannende en leuke moment op en buiten het werk (Thank you for the distractions and fun, for you non-Dutch roomies). Harry Stuger, Anne Elgersma, Ger Albers en Roelof Kammenga bedankt voor de afleiding met uitdagende discussies die over iets heel anders gingen dan mijn onderzoek. Tenslotte natuurlijk mijn rookmaatjes, Lottie Datema, Jettie Veendijk, Ilona van Ardennen (ook al rook je zelf niet), Irene Bulk en Bas en de andere IT-ers/ICT-ers, bedankt voor de welkome afwisseling. Of het nou ging over iets serieus zoals voortgangstoetsen en urenadministratie of over Pinkpop, Prince, kippen, vakantie en telefoons in de vriezer, het hielp altijd om even mijn hoofd leeg te maken en op te laden. (Overigens bedankt dat ik nog steeds mee mag, ook al ben ik gestopt met roken).

Ook de vormgeving van dit proefschrift is natuurlijk niet vanzelf ontstaan. Ik wil mijn vader bedanken voor het beschikbaar stellen van zijn verzameling tekeningen en schilderijen, waardoor ik nu een mooie cover heb. Rob de Kort en Erik Noppert hartelijk bedankt dat jullie met zo’n snelheid en tegen slechts een amicale beloning – een paar biertjes, een fles whiskey en gezelligheid – zo’n mooi boekje hebben ontworpen. Jullie ideeën en oplossingen zijn vele malen beter dan die ik zelf had kunnen bedenken. Door jullie werk komt mijn proefschrift voor mij echt tot leven.
Dankwoord

Dan komen we nu toch wel echt bij het ererijtje. Ik wil mijn ouders, Peter en Irene Kerdijk, bedanken voor 30 jaar support en steun in alles wat ik heb nagestreefd. Jullie hebben me altijd gesteund, ook als ik het mezelf moeilijker maakte dan nodig. De reden dat ik zover heb kunnen komen zijn jullie. Ook tijdens mijn promotie heb ik met jullie de mooiste en naarste momenten kunnen delen en ik weet eerlijk gezegd niet wat een kind nog meer kan wensen. Kortom, bedankt voor alles.

Tenslotte een woord van dank voor mijn paranimfen, Joep Hofhuis en Martha van der Wal. Het is niet zomaar dat ik jullie gevraagd heb voor deze rol. Joep, hoewel het niet een bewuste keus is, loop ik toch een beetje in je voetsporen. Eerst met onderzoek naar diversiteit in mijn bachelor- en masterscriptie. Daarna een promotietraject, op een plek die ik via jouw gevonden heb. Na mijn promotie denk ik dat ik maar docent en onderwijscoördinator wordt en daarnaast onderzoek ga doen wat ik vervolgens misschien weer eens kan uitbouwen. Oh wacht, daar ben ik mee bezig. Net als jij dus! Naast deze parallel in ons carrièrepad ben ik je vooral heel dankbaar voor alle wijze raad met betrekking tot het managen van mijn eigenwijze zelf, mijn promotietraject en het leven in het algemeen. Het is mij een eer en genoegen dat je op de dag van mijn promotie naast me staat.

Martha, voor jou een ereplek; in mijn hart, in mijn leven en in dit dankwoord. De echte reden waarom dit promotietraject een succes is, is simpelweg dat ik jou heb leren kennen. De ultieme reden waarom ik geen spijt kan hebben van promoveren. Ik vond promoveren op sommige momenten echt zwaar, maar ons fijne leventje samen maakte het lichter en leuker. Ik kan niemand bedenken die ik liever op deze dag naast me heb staan dan jij. Dankjewel voor alles.
This thesis is published within the Research Institute SHARE (Science in Healthy Ageing and Health caRE) of the University Medical Center Groningen / University of Groningen. Further information regarding the institute and its research can be obtained from our internetsite: www.share.umcg.nl

More recent theses can be found in the list below.
((co-) supervisors are between brackets)

2014

- **Visser L** Early detection and prevention of adolescent alcohol use; parenting and psychosocial factors  
  (prof SA Reijneveld, dr AF de Winter)
- **Tovote KA** Acceptance or challenge? Psychological treatments for depressive symptoms in patients with diabetes  
  (prof R Sanderman, prof PMG Emmelkamp, prof TP Links)
- **Trippolini M** Evaluation of functioning in workers with whiplash-associated disorders and back pain  
  (prof MF Reneman, prof PU Dijkstra, prof JHB Geertzen)
- **Eriks-Hoogland IE** Shoulder impairment in persons with a spinal cord injury & associations with activities and participation  
  (prof LHV van der Woude, prof G Stucki, prof MWM Post, dr S de Groot)
- **Suwantika AA** Economic evaluations of non-traditional vaccinations in middle-income countries: Indonesia as a reference case  
  (prof MJ Postma, dr K Lestari)
- **Behanova M** Area- and individual-level socioeconomic differences in health and health-risk behaviours; a comparison of Slovak and Dutch cities  
  (prof SA Reijneveld, dr JP van Dijk, dr I Rajnicova-Nagyova, dr Z Katreniakova)
- **Decker H** Teaching and learning professionalism in medical education  
  (prof J Cohen-Schotanus, prof T van der Molen, prof JW Snoek)
- **Dontje ML** Daily physical activity in patients with a chronic disease  
  (prof CP van der Schans, prof RP Stolk)
- **Gefenaite G** Newly introduced vaccines; effectiveness and determinants of acceptanc  
  (prof E Hak, prof RP Stolk)
- **Dagan M** The role of spousal supportive behaviors in couples’ adaptation to colorectal cancer  
  (prof M Hagedoorn, prof R Sanderman)
- **Monteiro SP** Driving-impairing medicines and traffic safety; patients’ perspectives  
  (prof JJ de Gier, dr L van Dijk)
- **Bredeweg S** Running related injuries  
  (prof JHB Geertzen, dr J Zwerver)
- **Mabood SI** Selection of medical students and their specialty choices  
  (prof JCC Borleffs, dr RA Tio)
- **Krieke JAJ van der** Patients’ in the driver’s seat; a role for e-mental health?  
  (prof P de Jonge, prof M Aielllo, dr S Sytema, dr A Wunderink)
- **Jong LD de** Contractures and hypertonia of the arm after stroke; development, assessment and treatment  
  (prof K Postema, prof PU Dijkstra)
- **Tiessen AH** Cardiovascular risk management in general practice  
  (prof K van der Meer, prof AJ Smit, dr J Broer)
- **Bedde MI** Complex Regional Pain Syndrome type 1 & amputation  
  (prof JHB Geertzen, prof PU Dijkstra, dr WFA van der Dunnen)
- **Lakke AE** Work capacity of patients with chronic musculoskeletal pain  
  (prof JHB Geertzen, prof MJ Reneman, prof CP van der Schans)