

Evaluating the Veterinary Clinical Teacher

Het evalueren van de veterinair klinisch docent

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht

op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,

ingevolge het besluit van het college voor promoties

in het openbaar te verdedigen op dinsdag 25 oktober 2011

des ochtends te 10.30 uur

door

Tobias Bertus Bernard Boerboom

geboren op 20 januari 1981 te Zevenaar

PROMOTOREN:

Prof.dr. P. van Beukelen

Prof.dr. A.J.J.A. Scherpbier

CO-PROMOTOREN:

Dr. D.H.J.M. Dolmans

Dr. A.D.C. Jaarsma

**FINANCIAL SUPPORT FOR THE PRINTING OF THIS THESIS
HAS BEEN KINDLY PROVIDED BY:**

Vets-net Beheer bv.

Boerboom Loodgieters bv.

F.W. Boerboom Beheer bv.

To (opa) Bernard Verhoeven

Laboris gloria ludi

Work hard, play hard

ISBN: 978-90-393-5612-8

Cover design: Anjolieke Dertien

Printing: Ipskamp Drukkers Enschede

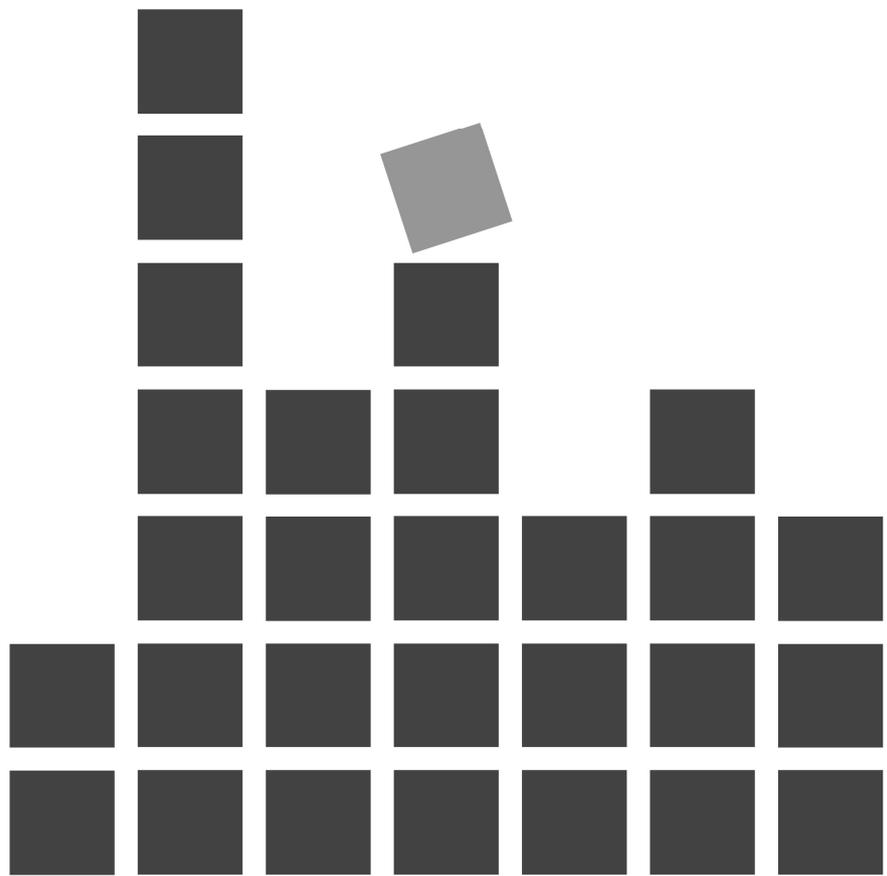
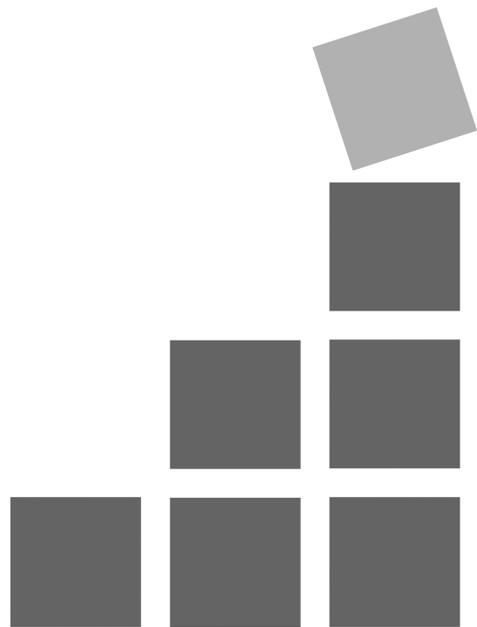


Table of contents

1. General Introduction	1
2. Does a faculty development programme improve teachers' perceived competence in different teacher roles? Med Teach 2009;31:1030-1	15
3. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations Med Teach 2011;33:e84-91	21
4. Evaluating clinical teachers with the MCTQ: how much 'teacher' is in student ratings? 2011 Under review	43
5. Peer group reflection helps clinical teachers to critically reflect on their teaching 2011 Under review	61
6. The effect of facilitated student feedback on ratings of clinical teaching effectiveness 2011 Under review	83
7. General Discussion	101
Summary	121
Samenvatting	131
Dankwoord	141
Curriculum Vitae	147

1.

General Introduction



INTRODUCTION

Clinical teaching is defined as 'teaching and learning focused on, and usually directly involving, patients and their problems'.¹ As early as 1932, the American Veterinary Medical Association (AVMA) stipulated that veterinary colleges could only be accredited if they included a clinical (teaching) hospital.² It is partly due to this decision that today's veterinary students spend one or more years of their training in the clinical workplace.³ Like medical students, all veterinary students are introduced to the clinical workplace at some point in the undergraduate curriculum. This is not surprising, because in both the medical^{1,4-6} and the veterinary context^{3,7,8} workplace learning is regarded as probably the most important part of the undergraduate curriculum.

In most veterinary curricula learning in the clinical workplace is preceded by one to several years that are largely devoted to acquiring a biomedical knowledge base and basic skills, including diagnostic and surgical competencies, but also non-technical skills like communication and practice management skills.^{7,9,10} The objectives pursued during workplace learning are different. During clerkships students develop a fundamental core of clinical skills, such as history taking, physical examination and clinical reasoning, and integrate knowledge with other complex competencies while having a role in patient care. In the daily routine of clinical work, decision-making and discussions take place in real time. Working side by side with clinicians in an authentic learning environment, students learn how to deal with professional ethics and conflicting information, communicate effectively, work in a team, ensure patient safety, take account of public health aspects and work within economic constraints.^{1,5,7,11,12} Cognitive and constructivist theories suggest that learning is strongly influenced by context, culture and interactions with others and is often guided by experts in an apprenticeship model.¹³⁻¹⁵ Workplace learning fits these contemporary concepts of learning. Generally, veterinary and medical students are exposed to learning opportunities in the clinical workplace during clinical rotations in a teaching hospital.⁷

These clinical clerkships are basically the same in medical and veterinary education. Students are confronted with real patient cases under supervision of clinicians in different clinical disciplines, such as surgery, anaesthesiology, internal medicine and gynaecology.^{1,7,11} The length of rotations, however, tends to be shorter in veterinary education, because rotations have to be completed for different animal species. The gynaecology rotation, for example, takes place in the Companion Animal Department as well as in the Equine Health Department.

GENERAL INTRODUCTION

Because workplace learning depends on patients or clients and their animals, who are opportunistically available, teaching tends to be rather haphazard.^{1,12} Several authors have tried to identify factors that contribute to the learning of students in the clinical setting.^{1,4,6,11,16-20} Their work has revealed several key factors, such as feedback, assessment, students' motivation and the learning climate, but above all, the teaching effectiveness of the clinical teacher.

Clinical teachers have a key role in the transformation of students from novice to professional.^{4,12,21-23} Kilminster and Jolly defined a clinical teacher's supervision as 'the provision of monitoring, guidance and feedback on matters of personal, professional and educational development in the context of the doctor's care of patients'.²⁴ Clinical teachers help to build the crucial bridge connecting students' preclinical knowledge and their new clinical experiences.⁷ Students and teachers engage in shared activities in the course of daily routine,²⁵ with clinical teachers acting as role models, supervising and coaching their future colleagues and helping them to reflect on their actions.^{4,21,26} In their 2001 paper, Irby and Papadakis review the characteristics that make a clinical teacher an excellent teacher.²³ Outstanding clinical teachers show enthusiasm for teaching, are well organised, accessible, supportive and compassionate towards students. When supervising students during clinical tasks, clinical teachers provide direction and feedback and demonstrate professionalism.

Even though these characteristics may seem obvious, demonstrating them is frequently undermined by the constrictions of the context.¹ As in human medicine, the rising costs of care, competition between hospitals, increased (sub)specialisation, increased service expectations of the public and reduced government support for teaching hospitals, pose huge challenges to clinical teaching in the veterinary setting.^{1,3,8} One challenge that is unique to veterinary clinical education is that patient care is often limited by an animal owner's financial means. Only a small minority of pets are insured against veterinary medical expenses.³ All these factors impede the funding of the teaching hospital. In these circumstances clinical teachers are under pressure to spend a great deal of time on client service and thus face competing demands from patient care and teaching.^{3,21,27} This has led to a less than optimal use of learning opportunities in the clinical setting, because students' learning is not the prime objective of the teaching hospital organization.²⁸⁻³⁰

Clinical teaching is not only affected by the above mentioned changes, the teaching effectiveness of clinical staff is also subject to influences from the modern academic environment. Clinical teachers are primarily doctors.³¹ Traditionally, the main requirement to be met by teachers in higher education was professional expertise,

but educational approaches have been changing.³² And although most clinical teachers are well trained for their tasks in patient care, their training in teaching is usually limited.³³ In the wake of the 1993 report of the General Medical Council, 'Tomorrow's Doctors', clinical teachers had to change their traditional approach to teaching.³⁴ Today it is no longer considered sufficient for clinical teachers to just be able to transmit knowledge.³⁵ The change from traditional didactic teaching to a new teaching role where the emphasis has shifted to supporting the learning of students, means that teachers need to develop new teaching competencies, such as providing concrete and useful feedback and stimulating students to reflect.³² Many institutions, both medical and veterinary, however, appear to place more value on research activities than on teaching.^{8,36} The quality of teaching is under threat because relatively easily quantifiable research performance and clinical service have become the principal measures by which clinical teachers' professional productivity is judged.³⁶

At the same time faculty development for medical teachers is gaining increasing prominence.^{21,37} It comprises a range of activities which institutions can offer to assist staff in the performance of their teaching roles.^{33,38} Faculty development can provide clinical teachers with educational insights and skills and alter attitudes and beliefs concerning teaching tasks.³³ Yvonne Steinert, an important ambassador of faculty development for medical teachers, has described and studied a broad range of activities to stimulate professional (educational) growth in medical teachers.^{33,37,39-42} In a systematic review of faculty development Steinert et al. describe a number of important implications for practice and future research with regard to faculty development aimed at improving teaching effectiveness:³⁷

- More deliberate use of theories and educational principles in the development of faculty development programmes.
- Faculty development programmes should extend over time and stimulate reflection to allow for cumulative learning practice and growth.
- Instruments which can measure teaching performance should be developed. These instruments should be assessed on validity and reliability.
- More rigorous research studies, which also utilize qualitative data and control groups, should be conducted to get insight into the effects of a faculty development programme.
- Different faculty development methods should be compared to get insight into which features contribute to changes in teacher behaviour.

GENERAL INTRODUCTION

- Changes in teacher behaviour after a faculty development intervention should be assessed over time.

An important part of faculty development is the evaluation of clinical teaching.⁴³ Since supervision is the key to successful learning in the workplace, instruments to evaluate clinical teachers' supervisory skill are receiving a great deal of attention in the literature.⁴⁴ Such evaluation instruments can not only identify where clinical teaching can be improved, they also provide feedback to stimulate teachers to reflect on their teaching practice and come up with concrete actions for improvement.^{44,45}

GENERAL PROBLEM DEFINITION

In many institutions student ratings are the main source of information about the quality of the teaching of individual clinical teachers.⁴⁶⁻⁴⁸ There is good evidence that students' evaluations of teaching are very useful for this purpose⁴⁹ and in recent years many instruments for student rating of individual teachers' clinical teaching effectiveness have been developed for the medical context.^{47,50} Most of these instruments are developed in cooperative efforts involving different stakeholders (students, staff, and experts). Good examples are the Cleveland Clinic's Clinical Teaching Effectiveness Instrument (CTEI)⁵¹ and the Stanford Faculty Development Program's Clinical Teaching Framework.⁵² Although the validity and reliability of these instruments are appropriate,^{47,53} the theoretical dimensions underlying the items are not sufficiently clearly specified.^{29,54}

A fairly new rating instrument, the Maastricht Clinical Teaching Questionnaire (MCTQ) measures students perceptions of individual clinical teachers' supervisory skills during undergraduate clinical rotations in the medical curriculum.^{26,54} Developed by Stalmeijer et al. in 2008, it has a strong theoretical basis in the cognitive apprenticeship model, describing a teaching method in which the generally tacit cognitive processes of experts are made explicit.⁵⁵ The MCTQ is composed along the six teaching strategies of the cognitive apprenticeship model: Modelling, Coaching, Scaffolding, Articulation, Reflection and Exploration. These strategies help to make explicit teachers' cognitive processes when supervising students.⁵⁴ Although the validity of the MCTQ has been demonstrated for the

medical context,²⁶ its validity and reliability has not been proven for the veterinary context.

According to several authors it is important to collect validity evidence from a broad range of sources when introducing an instrument into a new context.^{44,53} Although many studies have reported on the robustness and validity of instruments to obtain student ratings of clinical teachers,^{26,44,47,50-54,56-58} one aspect of validity is regularly overlooked: the potential impact of between-student differences and teacher and student characteristics not directly related to teaching performance.⁵⁹ Another source of validity evidence which, according to The American Psychological and Education Research Associations, often remains unexamined is the consequences and the effects of the evaluation and the feedback on the clinical teachers.⁶⁰ What is known so far is that written feedback alone is not sufficient.⁶¹⁻⁶³ Feedback is assumed to be more effective when teachers are stimulated to reflect on the evaluations of their performance, but much more research remains to be done in this area. Information about the reactions of clinical teachers after receiving feedback from student ratings and the effect of the feedback on teaching effectiveness can add substantially to the validity of an instrument.⁵³ This part of the impact of evaluation instruments, however, tends to remain underexposed.

The studies in this thesis were conducted to shed light on these areas in relation to the implementation of student rating instruments for the evaluation of individual clinical teachers in a veterinary context. We decided not to develop a new evaluation instrument, but to use an existing instrument from medical education and subject it to a thorough validation process for its applicability in the veterinary context. In 2004 and 2005 Beckman et al. investigated the validity evidence of a range of clinical teaching assessment instruments^{47,53} using the model of standards published by the American Psychological and Education Research Associations.⁶⁰ This model distinguishes five sources of validity evidence:

1. **Content:** The relationship between an instrument's content and the construct it is intended to measure.
2. **Response process:** The occurrence of unwanted variance when filling out the instrument.
3. **Internal structure:** The statistical validity and reliability of the instrument.
4. **Relations to other variables:** Relations with other possible predictive scores.
5. **Consequences:** The effect of the evaluation on the clinical teachers.

GENERAL INTRODUCTION

The instrument that is at the centre of the studies in this thesis is the MCTQ,^{26,54} because it already met the first source of validity evidence. The MCTQ's content validity is assured because it is based on theory about effective apprenticeship learning, the cognitive apprenticeship model¹³ and because of the involvement of different stakeholders in its development.⁵⁴

The wish of the Faculty of Veterinary Medicine, Utrecht University to implement evaluation of individual clinical teachers, the implications of the review of faculty development³⁷ and the model of sources of validity evidence⁶⁰ together provided the motivation to examine the general research question of this thesis:

What is the validity evidence of the MCTQ in a veterinary context?

We formulated four specific research questions to address the general research question:

- Does a faculty development programme improve teachers' perceived competence in different teaching roles? (Chapter 2)
- What is the validity evidence for the MCTQ in a veterinary curriculum? (Chapter 3 and 4)
- Does facilitated feedback based on MCTQ ratings stimulate clinical teachers to critically reflect on their teaching behaviour and formulate plans for improvement of their behaviour? (Chapter 5)
- What is the effect of facilitated feedback on teachers' clinical teaching performance? (Chapter 6)

The research was carried out in the clinical learning environment of the Faculty of Veterinary Medicine, Utrecht University, the Netherlands. Both quantitative and qualitative research methods, such as rating scales, questionnaires, pilot groups and analysis of reflection reports were used to gain insight into the implementation of an evaluation instrument as a tool for faculty development.

THESIS OUTLINE

We end this introduction with an outline of the chapters in this thesis. **Chapter 2** describes the seven roles of a medical teacher. The aim of this study was to investigate, using a questionnaire, whether participation in a faculty development programme has a positive effect on veterinary teachers' perceptions of their competence in these seven different teaching roles.

The studies in Chapters 3 and 4 focus on the first four sources of evidence of the validity of the MCTQ in the clinical context of the Utrecht Faculty of Veterinary Medicine. In **Chapter 3** we examined the validity and reliability of the MCTQ using pilot groups, confirmatory factor analysis, and a generalizability study of the MCTQ as an instrument to evaluate individual clinical teachers during short clinical rotations in veterinary education. This chapter provides an extensive description of the data collection methods that were used.

In **Chapter 4** we focus on the statistical validation of the MCTQ in the veterinary setting. In this study we used a dataset of 1277 student evaluations to explore to which degree instrument scores are due to differences between students versus differences between teachers. In this study we used multilevel analysis to examine how much of the variance in MCTQ scores resided within the teachers and which characteristics, such as gender or participation in a faculty development programme, influenced the student ratings.

Chapter 5 describes a more qualitative study. Other studies have shown that simply providing teachers with student evaluations does not automatically promote critical reflection and the formulation of alternative methods of action and improvement of teaching strategies. We investigated which types of reflection were shown by clinical teachers after receiving MCTQ generated feedback in a feedback strategy comprising self-assessment, a feedback report and a pre-structured reflection report. We also explored whether adding peer group reflection to this feedback strategy helped teachers to reflect more critically on feedback and formulate more concrete alternatives for action. We also asked the teachers to share their experiences with the feedback strategies.

In the final study (**Chapter 6**) we investigated if providing veterinary clinical teachers with MCTQ student feedback combined with the above mentioned facilitation strategies leads to higher MCTQ ratings. We compared three groups: (1) a group of veterinary clinical teachers who received facilitated MCTQ feedback without a peer reflection meeting and (2) a group that participated in

GENERAL INTRODUCTION

the same feedback facilitation strategy but with an additional peer reflection meeting. A third group (3) of clinical teachers served as control group and received no feedback.

Finally, the conclusions from the studies reported in Chapters 2-6, which represent the main body of this thesis, are summarised and discussed in **Chapter 7**. Since all the studies were written to be read on their own, there is inevitably some repetition and overlap across chapters.

REFERENCES

1. Spencer J. ABC of learning and teaching in medicine: Learning and teaching in the clinical environment. *Brit Med J* 2003;326:591-4.
2. West RL. Education in transition. *J Vet Med Educ* 1988;15:20-2.
3. Lane IF, Strand E. Clinical veterinary education: Insights from faculty and strategies for professional development in clinical teaching. *J Vet Med Educ* 2008;35:397-406.
4. Deketelaere A, Kelchtermans G, Struyf E. Disentangling clinical learning experiences: An exploratory study on the dynamic tensions in internship. *Med Educ* 2006;40:908-15.
5. Leinster S. Learning in the clinical environment. *Med Teach* 2009;31:79-81.
6. Pratt DD, Harris P, Collins JB. The power of one: Looking beyond the teacher in clinical instruction. *Med Teach* 2009;31:133-7.
7. Smith BP, Walsh DA. Teaching the art of clinical practice: The veterinary medical teaching hospital, private practice, and other externships. *J Vet Med Educ* 2003;30:203-6.
8. Bolt DM, Witte TH, Lygo-Baker S. The complex role of veterinary clinical teachers: How is their role perceived and what is expected of them? *J Vet Med Educ* 2010;37:388-94.
9. Brown JP, Silverman JD. The current and future market for veterinarians and veterinary medical services in the United States. *J Am Vet Med Assoc* 1999;215:161-83.
10. Jaarsma ADC. Developments in veterinary medical education: Intentions, perceptions, learning processes and outcomes [dissertation]. Utrecht: Utrecht University; 2008.
11. Van der Hem-Stokroos HH, Scherpbier AJJA, Van der Vleuten CPM, De Vries H, Haarman HJTM. How effective is a clerkship as a learning environment? *Med Teach* 2001;23:599-604.
12. Ramani S, Leinster S. AMEE guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30:347-64.
13. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *American Educator* 1991;15:6-11.
14. Schön DA. *Educating the reflective practitioner: Towards a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass; 1991.
15. Irby DM. Teaching and learning in ambulatory care settings: A thematic review of the literature. *Acad Med* 1995;70:898-933.
16. Remmen R, Denekens J, Scherpbier AJJA, Hermann I, Van Der Vleuten CPM, Royen PV, et al. An evaluation study of the didactic quality of clerkships. *Med Educ* 2000;34:460-4.
17. Van der Vleuten CPM, Scherpbier AJJA, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. *Med Teach* 2000;22:592-600.
18. Smith CA, Varkey AB, Evans AT, Reilly BM. Evaluating the performance of inpatient attending physicians: A new instrument for today's teaching hospitals. *J Gen Intern Med* 2004;19:766-71.
19. Boor K, Scheele F, van der Vleuten CPM, Teunissen PW, den Breejen EME, Scherpbier AJJA. How undergraduate clinical learning climates differ: A multi-method case study. *Med Educ* 2008;42:1029-36.

GENERAL INTRODUCTION

20. Yeates PJA, Stewart J, Barton JR. What can we expect of clinical teachers? Establishing consensus on applicable skills, attitudes and practices. *Med Educ* 2008;42:134-42.
21. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: Maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000;34:820-6.
22. Prideaux D, Spencer J. On theory in medical education. *Med Educ* 2000;34:888-9.
23. Irby DM, Papadakis M. Does good clinical teaching really make a difference? *Am J Med* 2001;110:231-2.
24. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: A literature review. *Med Educ* 2000;34:827-40.
25. Ross A, Stenfors-Hayes T. What do medical teachers do? *The Clinical Teacher* 2008;5:159-62.
26. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Muijtjens AMM, Scherpbier AJJA. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010;85:1732-8.
27. Hoffman KG, Donaldson JF. Contextual tensions of the clinical environment and their influence on teaching and learning. *Med Educ* 2004;38:448-54.
28. Dolmans DHJM, Wolfhagen HAP, Essed GGM, Scherpbier AJJA, van der Vleuten CPM. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. *Acad Med* 2002;77:332-5.
29. Dolmans DHJM, Wolfhagen HAP, Gerver WJ, de Grave W, Scherpbier AJJA. Providing physicians with feedback on how they supervise students during patient contacts. *Med Teach* 2004;26:409-14.
30. Irby DM, Bowen JL. Time-efficient strategies for learning and performance. *The Clinical Teacher* 2004;1:23-8.
31. Seabrook M. Medical teachers' concerns about the clinical teaching context. *Med Educ* 2003;37:213-22.
32. Postareff L, Lindblom-Ylänne S, Nevgi A. The effect of pedagogical training on teaching in higher education. *Teach Teach Educ* 2007;23:557-71.
33. Steinert Y. Staff development for clinical teachers. *The Clinical Teacher* 2005;2:104-10.
34. General Medical Council. *Tomorrow's doctors. recommendations on undergraduate medical education.* London: General Medical Council; 1993.
35. Harden RM, Crosby J. AMEE guide no 20: The good teacher is more than a lecturer - the twelve roles of the teacher. *Med Teach* 2000;22:34-47.
36. Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the flexner report. *N Engl J Med* 2006;355:1339.
37. Steinert Y, Mann K, Centeno A, Dolmans DHJM, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach* 2006;28:497-526.
38. Centra JA. Types of faculty development programs. *J High Educ* 1978;49:151-62.
39. Steinert Y. Faculty development in the new millennium: Key challenges and future directions. *Med Teach* 2000;22:44-50.

40. Steinert Y, Nasmith L, McLeod PJ, Conochie L. A teaching scholars program to develop leaders in medical education. *Acad Med* 2003;78:142-9.
41. Steinert Y, Cruess S, Cruess R, Snell L. Faculty development for teaching and evaluating professionalism: From programme design to curriculum change. *Med Educ* 2005;39:127-36.
42. Steinert Y, Mann KV. Faculty development: Principles and practices. *J Vet Med Educ* 2006;33:317-24.
43. Ramani S. Twelve tips to promote excellence in medical teaching. *Med Teach* 2006;28:19-23.
44. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
45. Morrison J. ABC of learning and teaching in medicine: Evaluation. *Brit Med J* 2003;326(7385):385-7.
46. Irby DM, Ramsey PG, Gillmore GM, Schaad D. Characteristics of effective clinical teachers of ambulatory care medicine. *Acad Med* 1991;66:54-5.
47. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004;9:971-7.
48. McOwen KS, Bellini LM, Guerra CE, Shea JA. Evaluation of clinical faculty: Gender and minority implications. *Acad Med* 2007;82:94-6.
49. Marsh HW, Roche LA. Making students' evaluations of teaching effectiveness effective. *Am Psychol* 1997;52:1187-97.
50. Fluit C, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers. *J Gen Intern Med* 2010;25:1337-45.
51. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000;75:161-6.
52. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998;73:688-95.
53. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005;20:1159-64.
54. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:272-7.
55. Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. *Knowing, learning and instruction: Essays in honor of Robert Glaser*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.; 1989. p. 453-94.
56. Williams BC, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med* 2002;77:177-80.
57. Van der Hem-Stokroos HH, van der Vleuten CPM, Daelmans HEM, Haarman HJTM, Scherpbier AJJA. Reliability of the clinical teaching effectiveness instrument. *Med Educ* 2005;39:904-10.
58. Marsh HW. Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *J Educ Psychol* 2007;99:775-90.

GENERAL INTRODUCTION

59. Downing SM. Threats to the validity of clinical teaching assessments: What about rater error? *Med Educ* 2005;39:353-5.
60. American Education Research Association and American Psychological Association. Standards for educational and psychological testing. Washington DC: American Education Research Association; 1999.
61. Atwater LE, Waldman DA, Brett JF. Understanding and optimizing multisource feedback. *Hum Resour Manage* 2002;41:193-208.
62. Seifert CF, Yukl G, McDonald RA. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *J Appl Psychol* 2003;88:561-9.
63. Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Person Psychol* 2005;58:33-66.

2.

Does a faculty development programme improve teachers' perceived competence in different teacher roles?

T.B.B. Boerboom

D.H.J.M. Dolmans

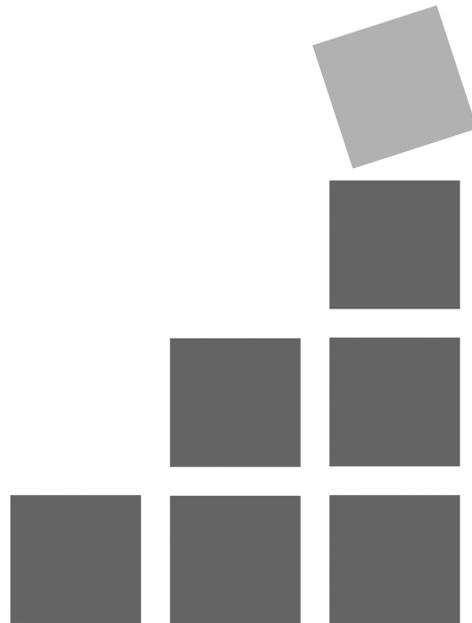
A.M.M. Muijtjens

A.D.C. Jaarsma

P. van Beukelen

A.J.J.A. Scherpbier

Medical Teacher 2009
31(11) 1030-1031



ABSTRACT

Context

Changing concepts of education have led many medical schools to adopt student-centred approaches to teaching, requiring different teaching competencies instead of more traditional approaches. The aim of this study was to investigate whether participation in a faculty development (FD) programme had a positive effect on veterinary medical teachers' perceptions of their competence in seven different teaching roles.

Methods

All teaching staff of a veterinary medical school (251) were invited to complete a questionnaire asking about their perceived competence in seven teaching roles. A stepwise multiple linear regression analysis examined the effect of the completion of a FD programme on perceived teaching competence.

Results

There was a significant positive effect of the FD programme on teachers' perceived competence in the majority of the teaching roles at issue.

Conclusions

FD appeared to have a positive effect on the perceived competence of veterinary medical teachers. Further research should investigate whether FD actually results in improved teaching performance.

INTRODUCTION

In recent decades, constructivist views have gained ground in education as is reflected in the introduction of student-centred approaches to teaching, where teachers are less transmitters of knowledge than facilitators of the learning of students.^{1,2} As a result, many teachers today are required to be able to guide students' thinking about subjects and challenge students' ideas by asking probing questions and engaging students in discussions.

Apart from specific, discipline-related knowledge and skills, medical education today also includes generic aspects, such as academic and communication skills and professionalism.³ With these changing notions concerning education, it is no longer sufficient for teachers to have good didactic skills but they are expected to be competent designers, developers, evaluators and organisers of education as well.²

The high demands made upon teachers by changing teacher roles have spurred the development of faculty development (FD) programmes in many countries.^{4,5} A review by Steinert et al. (2006)⁶ showed that participation in FD resulted in teachers reporting more positive attitudes towards teaching as well as gains in teaching skills. However, the majority of the interventions included in this review targeted didactic skills in the classroom and at the bedside with little attention to other teacher roles, such as organiser or developer of education. Unsurprisingly, the review reported little evidence of an impact on teacher roles other than didactic ones.⁴

The study we report in this article addresses a broad range of teacher roles. More specifically, we investigated the impact of the completion of a FD programme on teachers' perceptions of their competence in seven roles in education: Person as Teacher, Content Expert, Designer, Counsellor, Evaluator, Organiser and Scholar.

METHODS

This study was conducted at the Faculty of Veterinary Medicine, Utrecht University, the Netherlands (FVMU) in 2007, where curriculum revisions in 1995, 2001 and 2007 had promoted student-centred methods, such as small group work. At Utrecht University, all teaching faculty are offered a 2-year (125 h) FD programme.

All 251 faculty members of FVMU were invited to participate in this study. Of these, approximately two-thirds were level 1 teachers (postgraduate students,

FACULTY DEVELOPMENT

junior lecturers and assistant professors) and one-third were level 2 teachers (associate and full professors). 62% were clinical teachers.

We elicited teachers' perceptions of their competence by administering a questionnaire based on a framework developed by Tigelaar et al. (2004),² comprising the seven teaching roles named in the Section 'Introduction'. An expert panel of five experienced FVMU teachers agreed upon the wording and relevance of the items. The questionnaire is available from the author T.B.B. Boerboom.

The questionnaire was sent to all faculty members. For each teacher role they were asked to indicate on a five-point Likert scale, to what extent they agreed with the following statement: 'I feel competent to fulfil this role' (1=fully disagree; 5=fully agree). Two additional questions asked them to state their number of years of experience as a teacher and whether they had completed the FD programme.

Participation was voluntary and participants were assured of confidentiality. The questionnaire was coded in such a way as to ensure that the results could not be traced to individual respondents.

Stepwise multiple linear regression analysis was performed to examine the effects of the FD programme on teachers' perceived competence in the seven roles. The independent variable was the completion of the FD programme and the dependent variables were teachers' perceived competence in each of the seven roles. Three moderator variables, i.e. teacher rank (level 1 or level 2), department (clinical or non-clinical) and years of teaching experience were included in the model for all the analyses. In a stepwise procedure, FD and its interactions with the moderator variables were examined for their contribution to the variation in the dependent variables.

RESULTS

A total of 194 teachers (77%) returned the completed questionnaire. The mean number of years of teaching experience was 14.4 years (SD=10.4). Fifty-five percent of respondents had completed the FD programme. Perceived competence was highest for Person as Teacher ($M=4.14$, $SD=0.73$) and lowest for Teacher as Designer ($M=3.82$, $SD=0.76$).

The regression analysis revealed significant positive effects of the completion of the FD programme on perceived competence in the following teaching roles: Person as

Teacher, Counsellor, Content Expert and Scholar, with regression coefficients of 0.27, 0.25, 0.23 and 0.21, respectively.

Although the effects of FD on the roles of Designer and Organiser were significant, they were moderated by years of experience. There was no effect for teachers with average experience (14 years), a positive effect for teachers with limited experience (e.g. 4 years, regression coefficients 0.24 and 0.30, respectively), and a negative effect for highly experienced teachers (e.g. 24 years, regression coefficients -0.24 and -0.30 , respectively). No significant effect was related to the role of Evaluator.

DISCUSSION

Completion of the FD programme is associated with higher perceived competence in the roles of Person as Teacher, Counsellor, Content Expert and Scholar. This suggests that the FD programme is successful in enhancing perceived competence in most of the roles addressed by the programme. The absence of a significant effect in respect of the Evaluator role may well be attributable to its low prominence in the FD programme. The positive impact of the FD programme on the perceived competence in the Designer and Organiser roles of the less experienced teachers (<14 years) is in line with our expectations, but the concomitant negative effect for the highly experienced teachers (>14 year) is rather puzzling. We think that an explanation may be found in the strong focus on self-reflection during the FD programme. However, further in-depth research is needed to test this hypothesis.

A limitation of this study is that the results are based on self-assessment, which has been shown to be not entirely reliable. Further research will have to tap into additional data sources, such as interviews with teachers and students or observations of teaching sessions, in order to provide stronger evidence that the completion of a FD programme does indeed improve teacher's performance.

FACULTY DEVELOPMENT

REFERENCES

1. Harden RM, Crosby J. AMEE guide no 20: The good teacher is more than a lecturer - the twelve roles of the teacher. *Med Teach* 2000;22:34-47.
2. Tigelaar DEH, Dolmans DHJM, Wolfhagen IHAP, Van der Vleuten CPM. The development and validation of a framework for teaching competencies in higher education. *High Educ* 2004;48:253-68.
3. Robley W, Whittle S, Murdoch-Eaton D. Mapping generic skills curricula: A recommended methodology. *Journal of further and higher education* 2005;29:221-31.
4. Postareff L, Lindblom-Ylänne S, Nevgi A. The effect of pedagogical training on teaching in higher education. *Teach Teach Educ* 2007;23:557-71.
5. McLean M, Cilliers F, Van Wyk JM. Faculty development: Yesterday, today and tomorrow. *Med Teach* 2008;30:555-84.
6. Steinert Y, Mann K, Centeno A, Dolmans DHJM, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach* 2006;28:497-526.

3.

Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations

T.B.B. Boerboom

D.H.J.M. Dolmans

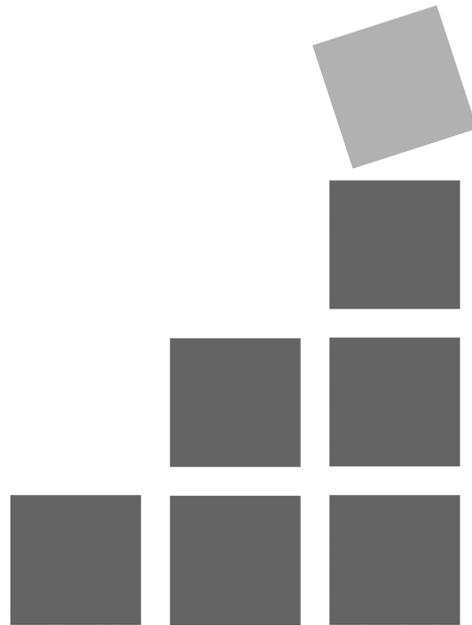
A.D.C. Jaarsma

A.M.M. Muijtjens

P. van Beukelen

A.J.J.A. Scherpbier

Medical Teacher 2011
33(2) e84-91



ABSTRACT

Context

Feedback to aid teachers in improving their teaching requires validated evaluation instruments. When implementing an evaluation instrument in a different context, it is important to collect validity evidence from multiple sources. We examined the validity and reliability of the Maastricht Clinical Teaching Questionnaire (MCTQ) as an instrument to evaluate individual clinical teachers during short clinical rotations in veterinary education.

Methods

We examined four sources of validity evidence: (1) Content was examined based on theory of effective learning. (2) Response process was explored in a pilot study. (3) Internal structure was assessed by confirmatory factor analysis using 1086 student evaluations and reliability was examined utilizing generalizability analysis. (4) Relations with other relevant variables were examined by comparing factor scores with other outcomes.

Results

Content validity was supported by theory underlying the cognitive apprenticeship model on which the instrument is based. The pilot study resulted in an additional question about supervision time. A five-factor model showed a good fit with the data. Acceptable reliability was achievable with 10–12 questionnaires per teacher. Correlations between the factors and overall teacher judgement were strong.

Conclusions

The MCTQ appears to be a valid and reliable instrument to evaluate clinical teachers' performance during short rotations.

INTRODUCTION

Clinical rotations in hospitals are an important part of the undergraduate curriculum in both medical and veterinary education. They provide an authentic learning environment in which students can participate in patient care and learn to integrate theory and clinical practice. Clinical rotations in a medical and in a veterinary curriculum are quite similar. However, the length of the rotations is often shorter in the veterinary situation because of the different animal species. For example, students not only have an anaesthesia rotation at the Companion Animal Department, they are also required to complete this rotation at the Equine Health Department.

Although routine clinical work provides students with opportunities to learn about history taking, physical examination, clinical reasoning and professionalism,¹⁻⁴ the clinical learning environment is not always optimally used for the learning of students.^{2,5,6}

Less than optimal use of learning opportunities in clinical settings is related to students' learning not being the prime objective of the hospital organization.^{2,7} Clinical staff in teaching hospitals face competing demands from patient care, patient safety, administration, research and teaching.^{8,9} As a result they do not always have sufficient time for teaching and observing students, which in turn can diminish the quality of clinical education.^{2,4}

Most clinical teachers are well prepared for and dedicated to their tasks in patient care, but their training in teaching skills is usually scant despite their general enthusiasm for teaching. The fact that this can be detrimental to their effectiveness as teachers^{2,10} has given rise to the growing prominence of faculty development in medical education.^{8,11}

Institutions that invest in faculty development need instruments to measure teaching effectiveness to identify areas of clinical teaching where training is needed and to measure the return on their investment in faculty development.¹² Even more importantly, measurement of teaching effectiveness can provide feedback to guide, support and motivate clinical teachers to improve their teaching.^{4,13,14}

Stalmeijer et al. (2008) designed the Maastricht Clinical Teaching Questionnaire (MCTQ) as an instrument to evaluate individual clinical teachers' supervisory skills during undergraduate clinical rotations in the medical curriculum.¹⁵ The MCTQ is based on the cognitive apprenticeship model and measures students' evaluations of

VALIDITY AND RELIABILITY

a teacher as a good role model and supervisor. The cognitive apprenticeship model focuses on the teacher's cognitive processes during performance of complex tasks and comprises six teaching strategies (Modelling, Coaching, Scaffolding, Articulation, Reflection and Exploration) which help to make explicit teachers' cognitive processes when supervising students.^{15,16} Stalmeijer et al. added Learning Climate as a seventh domain to this model, because a positive learning climate is considered essential for successful learning in the clinical workplace.^{15,17}

Stalmeijer et al. (2010) demonstrated the validity of the MCTQ for use in clinical rotations lasting between 4 and 6 weeks.¹⁸ The MCTQ can thus be regarded as a valuable instrument for evaluation, feedback, self-assessment, self-reflection and faculty development of clinical teachers. In this study, we investigated whether the MCTQ also provides reliable and valid information about clinical teachers' supervisory competence in a veterinary learning environment with relatively short rotations.

When assessing the validity of an instrument like the MCTQ in a different context than the one for which it was developed, it is important to collect validity evidence from a broad range of sources.^{14,19} Beckman et al. investigated the validity and reliability of a range of similar clinical teaching assessment instruments^{19,20} using the model of standards published by the American Psychological and Education Research Associations, which distinguishes five sources of validity evidence:^{19,21,22}

1. Content: the relationship between an instrument's content and the construct it is intended to measure.
2. Response process: analysing the response process affords insight into factors affecting the data collected with an instrument and can be elicited by asking respondents to articulate their thought processes during completion of an instrument. These factors (for example the wording of items) may be irrelevant to the construct being measured but can be a source of unwanted variance. Other aspects of response process are instrument delivery method, scoring and reporting. Evidence about response process can be used to control or eliminate all possible sources of error associated with the administration of an instrument.
3. Internal structure and reliability: this addresses the question whether the data generated by an instrument fits the underlying construct. It concerns the unidimensionality of the sub-scales, the reliability of the scores and the statistical and psychometric characteristics of an instrument.

4. Relations to other variables: the relations between instrument scores and other variables with relevance to the construct being measured also provide evidence of validity.
5. Consequences: the effect of the use of an instrument on those being evaluated is another source of validity evidence.

In this study, we examined the first four sources of validity evidence to determine the validity and reliability of the MCTQ in the veterinary education setting.

METHODS

Context

We conducted our study at the Faculty of Veterinary Medicine, Utrecht University, the Netherlands, (FVMU) between November 2008 and September 2009. FVMU offers a 6-year undergraduate curriculum consisting of four years of preclinical training and two years of clinical clerkships.

The first clerkship year follows the Uniform Clinical Rotation Programme, which involves 30 weeks of rotations in different clinical departments. In the second clerkship year, students undertake rotations in disciplines related to their chosen animal species track.

All clinical departments participated in the study: the Equine Health Department, the Companion Animal Health Department, the Farm Animal Health Department and the Pathology Department. These departments provide all the patient-based clinical training in both the first and the second clinical year. In each department, different disciplines (e.g. surgery or gynaecology) also contribute to the clerkship. Rotations last from 1 day to 6 weeks and approximately 190 staff members have roles in clinical teaching. Individual clinical teachers' supervisory skills during undergraduate clinical rotations are not evaluated systematically at FVMU.

Content

The MCTQ addresses seven domains (the six teaching strategies of the cognitive apprenticeship model and General Learning Climate, GLC) in 24 items. Content validity is ensured by the fact that the instrument is based on the cognitive apprenticeship model, which is underpinned by theory of effective apprenticeship

VALIDITY AND RELIABILITY

learning. The MCTQ's content validity is further supported by the involvement of different groups of stakeholders in its development^{15,16,18} and the fact, as stated in the Section 'Introduction', that the clinical rotations in the veterinary context are quite similar to the rotations in the medical context.

Response process

Response process was investigated in a pilot study among 28 students who had practical experience with clinical teaching. The students participated in the study when they had almost or completely finished one of the following rotations: first- and second-year Equine Animal Health rotations, first- and second-year Companion Animal Health rotations, first- and second-year Farm Animal Health rotations, and first-year Pathology rotation. The students were asked to fill out the instrument individually and discuss the relevance and wording of the MCTQ items and identify factors that affected their answers. These group discussions lasted 90 minutes. We asked them to think aloud while filling out the instrument and we asked their opinions on how to administer and distribute the instrument in order to ensure the highest rate and accurateness of response.

The students gave consent to audiotape and transcribe the discussions. The transcripts were analysed by the authors TB, AJ and PB, who individually assigned codes to all issues of interest. They met several times to discuss the coding until consensus was reached on the emerging themes. If necessitated by the outcome of the analysis, supplemental items would be added to the questionnaire.

Internal structure

To explore the third source of validity evidence (construct validity and reliability) we distributed a version of the MCTQ, modified to accommodate the results of the analysis of the response process (Appendix 1) to all fifth- and sixth-year students on clinical rotations during the 2008/2009 academic year. Students who commenced their rotations during the study period were also invited to participate. Approximately 350 students took part in the study.

The MCTQ asks students to rate the performance of one clinical teacher by indicating their agreement, on a 5-point Likert scale (1=fully disagree, 2=disagree, 3=neutral, 4=agree, 5=fully agree), with 24 statements relating to different teaching strategies and learning climate. Additionally, the students are

asked to give an overall judgement of the clinical teacher's performance on a 10-point scale. Two open-ended questions invite suggestions for ways in which teaching might be improved. The students in our study were also asked to answer the supplemental items that could emerge from the pilot study.

The students could evaluate several clinical teachers, using a separate questionnaire for each teacher. They were instructed to complete the questionnaire as soon as possible after a student–teacher encounter. Several reminders were sent in the form of an electronic newsletter.

Participation was voluntary. Students were asked to fill out their name and student number, but they were not obliged to do so because they could also decide to fill out the questionnaire anonymously. Confidentiality was assured and the questionnaire was coded in such a way that data could not be traced back to individual students. This information was provided to students and teachers through a website and email.

Because we were interested in the relationships between teaching strategies at the level of individual teachers, we analysed the aggregate (mean) scores of teachers for whom we had received four or more completed questionnaires.

We used AMOS 18.0 to conduct confirmatory factor analysis to investigate the construct validity of the instrument. We started from the cognitive apprenticeship model which underlies the MCTQ (see Appendix 1 for the domains and reworded items, the 'miscellaneous items' were not included in this analysis). Statistical analysis was performed using robust maximum likelihood estimation, a method less sensitive to violations of the normality assumption than other estimation methods. Inspection of modification indices revealed several items that had a negative impact on the fit of the model. We used several fit indices and criteria as proposed by Byrne (2001)²³ to determine the fit of the model: (1) χ^2 divided by the degrees of freedom (CMIN/df) is <2; (2) the goodness-of-fit index (GFI) is >0.80; (3) the comparative fit index (CFI) is >0.90; (4) and the standardized root mean square residual (SRMR) is <0.10.²³

Cronbach's alphas were computed for each scale (subset of items associated with a factor) to determine internal consistency. A coefficient of 0.70 or higher was considered acceptable.

We wanted to determine the number of student ratings required to provide individual teachers with reliable feedback on the factors resulting from the

VALIDITY AND RELIABILITY

confirmatory factor analysis, as well as on the overall judgement. Therefore, the inter-rater reliability of the teacher evaluation was assessed in a generalizability study for each factor score (mean of subset of items associated with a factor) of the rating form, and for the overall judgement score. Then variance component analysis was applied for the sample of rating forms, thereby decomposing the total variance of each score into the variance component of interest (teacher variance) and error variance components (rater variance, and rater–teacher interaction variance).²⁴ As indices for inter-rater reliability, we estimated the generalizability coefficient (G-coefficient) and the standard error of measurement (SEM) for each factor as well as for overall judgement using the urGENOVA software.²⁵ A G-coefficient of 0.60 was considered acceptable but indicative of a need for improvement and 0.80 was considered very reasonable.²⁶ The SEM was used to estimate confidence intervals for the individual factor scores. A SEM of ≤ 0.26 was considered adequate at the 95% confidence interval, taking into account that we decided to accept a maximum ‘noise level’ of 1.0 on a 5-point scale ($1.96 \times 0.26 \times 2 \approx 1$). An SEM of ≤ 0.52 (0.26×2) was considered satisfactory for the overall judgement because of the use of a 10-point scale.

Relations to other variables

We used SPSS 17.0 to perform bivariate correlation analyses on the factor scores and the scores on the overall judgement score and the supplemental items that could emerge from the pilot study in order to explore the relations of these items with the MCTQ outcomes. Bivariate correlations were obtained for the factor scores and the overall judgement score in order to assess criterion validity.

RESULTS

Response process

All students agreed that the MCTQ is a useful and relevant instrument for providing feedback to teachers. The pilot study led to the rewording of the questionnaire instructions and reformulation of a few items to reduce ambiguity. These changes were mostly induced by the veterinary context. More than 50% of respondents, for example, interpreted a safe learning environment as an environment where there was little chance of being bitten by an animal.

In the pilot study, supervision time was consistently mentioned by the students as a factor that affected their ratings. This comment was evoked by the local context at FVMU where students are assigned to departments and not to individual supervisors, and consequently are supervised by one teacher for only a few hours. While this enables students to observe how different faculty members work within a discipline, the downside is that students do not have the same supervisor for a prolonged period of time. Students said that brief exposure to a supervisor made them hesitant to give an evaluation and prevented judgement of certain aspects of the questionnaire. This result of the pilot study led to the addition of only one supplemental item on supervision time to the questionnaire to enable evaluation of the effect of supervision time on the response process. The item read: how many hours of actual contact time did you have with your supervisor?

According to the majority of the students in the pilot study, filling out paper forms at the clinic – the customary method of collecting student feedback at FVMU – was not the most trusted way regarding anonymity to administer the questionnaire. They suggested that a web-based questionnaire which they could access at home would be the safest and most accurate method to collect evaluation data.

The instrument was therefore turned into a web-based questionnaire and placed on a custom build website with easy access for the participating students. We sent an email to all fifth- and sixth-year students participating in clinical rotations during the 2008/2009 academic year (N=350) inviting them to fill out the questionnaire on the website when we were gathering data for the analysis of internal structure.

Internal structure

Of the 1223 completed questionnaires received by us, 33 were excluded because of incomplete data. The remaining set of 1190 questionnaires evaluated 163 different teachers. Fifty-three teachers were excluded from the analysis because we received fewer than four evaluations for them. This led to inclusion in the study of 1086 questionnaires evaluating 110 teachers. Because not all students filled out their name and student number, we cannot provide an accurate response rate; but at least 208 different students filled out their student number, resulting in a response rate of at least 59%.

All item scores ranged between 1 and 5. 'The clinical teacher allowed me to perform tasks independently' was the highest scoring item ($M=4.11$, $SD=0.57$). The

VALIDITY AND RELIABILITY

lowest scoring item was 'The clinical teacher stimulated me to formulate my own goals' ($M=2.90$, $SD=0.49$).

The confirmatory factor analysis showed that the original model – based on all domains of the cognitive apprenticeship model – (see Appendix 1) did not fit the data. After eliminating and reorganizing the items in accordance with the modification indices, the results of the confirmatory factor analysis revealed a reasonably good fit for a five-factor model comprising 15 items (Table 1).

The results of the five-factor model (Table 1) show the following results: (1) χ^2 divided by degrees of freedom ($CMIN/df$) is 2.594; (2) the GFI is 0.82; (3) the CFI is 0.93; and (4) the SRMR is 0.049. Only the first statistic did not meet the criteria for a reasonable fit. Models with 1, 2, 3 or 4 factors did not show a better fit (Table 2).

	Mean	SD	α
Factor 1: General Learning Climate (GLC)	3.99	0.62	$\alpha = 0.96$
The clinical teacher...			
1. established an environment where I felt free to ask questions or make comments.	4.07	0.71	
2. showed an interest in me as a student.	3.89	0.66	
3. treated me with respect.	4.04	0.57	
Factor 2: Modelling	3.81	0.54	$\alpha = 0.86$
The clinical teacher...			
1. demonstrated how different skills should be performed.	3.79	0.57	
2. explained, while performing a task, which aspects were important and why.	4.04	0.53	
3. was a role model for me.	3.61	0.73	

Table 1 continued			
	Mean	SD	α
Factor 3: Coaching	3.80	0.54	$\alpha = 0.87$
The clinical teacher...			
1. provided me with constructive and concrete feedback during or following direct observation.	3.55	0.62	
2. allowed me to perform tasks independently.	4.11	0.57	
3. was supportive when I experienced difficulties with a task.	3.73	0.61	
Factor 4: Articulation	3.95	0.41	$\alpha = 0.88$
The clinical teacher...			
1. asked me to explain my reasoning and actions.	3.83	0.47	
2. alerted me to gaps in my knowledge and skills.	3.93	0.43	
3. asked questions to increase my knowledge and understanding.	4.09	0.48	
Factor 5: Exploration	3.31	0.45	$\alpha = 0.90$
The clinical teacher...			
1. stimulated me to formulate my own goals.	2.90	0.49	
2. stimulated me to become aware of my own strengths and weaknesses.	3.57	0.49	
3. stimulated me to think about how to improve my own strengths and weaknesses.	3.45	0.50	

VALIDITY AND RELIABILITY

Table 2: Fit indices and criteria for models with 1, 2, 3, 4 and 5 factors

Model	N	CMIN/DF	GFI	CFI	SRMR
Conventional cut-off for adequate fit	-	<2	>.80	>.90	<.10
1 factor	110	6.78	0.560	0.700	0.084
2 factors	110	5.59	0.603	0.764	0.112
3 factors	110	3.94	0.712	0.853	0.074
4 factors	110	2.87	0.790	0.909	0.053
5 factors	110	2.59	0.816	0.926	0.049

Cronbach's alpha reliability coefficients for factors 1–5 were all above 0.70, at 0.96, 0.86, 0.87, 0.88 and 0.90, respectively (Table 1). The mean scores for factor 1 and factor 5 varied between 3.31 for Exploration (F5, scale 1–5) and 3.99 for GLC (F1, scale 1–5), with corresponding standard deviations of 0.45 and 0.62 (N=110).

We examined the inter-rater reliability of the five-factor scores and the overall judgement in a generalizability analysis. Table 3 presents the G-coefficients and SEM per factor as a function of the number of student ratings per clinical teacher. The number of students required to obtain reliable outcomes for the overall judgement of one teacher was six to eight. Between 10 and 12 students were needed for reliable outcomes for each factor. The bilateral correlation coefficients for the five mean factor scores ranged from $r=0.47$ to $r=0.71$ (all $p<0.001$).

RELATIONS TO OTHER VARIABLES

Table 4 shows the means and standard deviations for overall judgement and supervision time and their correlations with the five factors. Mean supervision time was 12.87h (SD=12.81).

There are large and significant correlations between the five factors and overall judgement ($r=0.84, 0.79, 0.70, 0.58$ and 0.62 ($p<0.001$ for all)). Although the correlations of the five-factor scores with supervision time are statistically significant, their values are small ($r=0.10, r=0.15, r=0.20, r=0.12$ and $r=0.13$ (all $p<0.001$)).

Table 3: The generalizability coefficient (G-coefficient) and standard error of measurement (SEM) as a function of the number of student ratings (N) for the five factors (scale 1-5) and the overall judgement (scale 1-10)

Factor	N (ratings)	G-coeff. ¹	SEM ²
F1 General Learning Climate	6	0.70	0.36
	8	0.76	0.32
	10	0.80	0.28
	12	0.82	0.26
F2 Modelling	6	0.66	0.33
	8	0.72	0.29
	10	0.77	0.26
	12	0.80	0.24
F3 Coaching	6	0.61	0.34
	8	0.68	0.29
	10	0.72	0.26
	12	0.76	0.24
F4 Articulation	6	0.53	0.32
	8	0.60	0.27
	10	0.65	0.24
	12	0.69	0.22

VALIDITY AND RELIABILITY

Table 3 continued			
Factor	N (ratings)	G-coeff. ¹	SEM ²
F5 Exploration	6	0.45	0.35
	8	0.52	0.31
	10	0.58	0.27
	12	0.62	0.25
Overall judgement	6	0.71	0.58
	8	0.77	0.50
	10	0.80	0.45
	12	0.83	0.41

1. A G-coefficient of ≥ 0.60 is considered acceptable
 2. F1-F5: a SEM of ≤ 0.26 is considered adequate
 Overall judgement: a SEM of ≤ 0.52 is considered adequate

Table 4: Mean scores and standard deviations (SD) for overall judgement (1-10) and supervision time and the correlations with the five factors							
	Descriptives		Correlations*				
	Mean	SD	F1 GLC	F2 Modelling	F3 Coaching	F4 Articulation	F5 Exploration
Overall judgement of the teacher	7.31	1.69	0.84	0.79	0.70	0.58	0.62
Supervision time	12.87	12.81	0.10	0.15	0.20	0.12	0.13

* All correlations were significant at the 0.01 level (2-tailed)

DISCUSSION

The purpose of this study was to validate the MCTQ as an instrument for measuring clinical teacher effectiveness in a veterinary education context with relatively brief clinical rotations. We sought evidence concerning the first four sources of evidence from the model of standards proposed by Beckman et al. (2005).¹⁹

The results of the pilot study among senior students to test the applicability of the MCTQ in a veterinary context show that students consider all domains and items relevant and useful for providing feedback to clinical teachers. Studies of the validity of the MCTQ conducted by Stalmeijer among educationalists, doctors and students in a medical education setting^{15,27} showed that the cognitive apprenticeship model offers a useful framework for the MCTQ. We therefore conclude that we have met the requirements of the first source of validity evidence: content.

Our analysis of the response process in the pilot study revealed one theme that, according to the students, affected the rating process:¹⁹ supervision time. Students felt that the short periods of supervision by one teacher, which are characteristic of the FVMU programme, prevented them from giving well-considered ratings for some aspects of the MCTQ. This is in line with Stalmeijer's finding that Scaffolding, Reflection and Exploration can be properly evaluated only in longer rotations with one clinical supervisor.²⁷

The validity evidence related to response process in this study led to the addition of one item on supervision time to enable measurement of its effects. We also improved the response process at the suggestion of students by replacing paper forms filled out at the clinic by Internet-based rating forms.

As for the third source of validity evidence, internal structure,¹⁹ the results of the confirmatory factor analysis show that a five-factor model comprising 15 items fits the data reasonably well, with 3 of 4 statistical criteria being met. Only the χ^2 divided by degrees of freedom (CMIN/df=2.594) did not meet the criteria for a reasonable fit (<2). We state that failing to meet this criterion is not problematic. Byrne's (2001)²³ fit indices are quite strict. Other authors propose a CMIN/df<5 as a criterion for an adequate fit.²⁸ The alpha-coefficients demonstrate acceptable levels.

VALIDITY AND RELIABILITY

The five-dimensional scale found by us is in line with findings reported by Stalmeijer et al. (2010) in a medical context.¹⁸ In our study, Coaching and Scaffolding aggregate into one factor as did Exploration and Reflection. This does not surprise us greatly, since Coaching and Scaffolding are both concerned with task performance: giving students the opportunity to perform a task, helping students when they are performing a task and giving students feedback during and following task performance. As for Exploration, this is about setting and pursuing learning goals, and setting (learning) goals is an integral part of reflection.²⁹

The findings of the generalizability study indicate that at least 10–12 student responses are needed per teacher to obtain reliable data at the factor level. Fewer responses (6–8) suffice for reliable overall judgement. In most clinical settings these numbers are quite feasible.

Based on the results, we conclude that the MCTQ has a valid internal structure for evaluating clinical teachers based on student ratings.

As for the relations between the MCTQ scores and other variables with relevance to teacher effectiveness¹⁹ we found large correlations between overall judgement and each of the factor scores. This adds support to the validity of the instrument.

The absence of strong correlations between supervision time and factor scores suggests that, in contrast to the opinions voiced by the students in the pilot study, supervision time does not have an important effect on the outcomes of the MCTQ. We hypothesize that this finding can be contributed to the fact that the specific behaviours reflected by the items in the MCTQ can be displayed by teachers even when their time for supervision is limited.

Based on the above findings, we conclude that this study has met the requirements with regard to four sources of evidence for the validity of an instrument like the MCTQ. We therefore conclude that the 5-domain, 15-item MCTQ, as shown in Table 1, appears to be an instrument with a strong theoretical foundation, which is valid and reliable for the evaluation of clinical teachers in a context with short clinical rotations.

There are some limitations to the study. For one thing, all measurements are based on student perceptions, which is a potential source of resistance from teachers when the MCTQ is used for faculty development purposes.²⁷ We therefore recommend using other evaluations in addition to student perceptions, such as teachers' self-

evaluations and observations. Further research should compare these evaluations with those of students.

Another drawback is that in the analysis we made no distinction between ratings from fifth- and sixth-year students. This is a common limitation in clinical teaching assessments.³⁰ However, the applicability of the cognitive apprenticeship model may vary depending on stages of education. Students who have just started clinical rotations, for example, require more observation and support than more advanced students. This is an issue that should be a topic of future research.

With regard to further research, we would also recommend studies to address the fifth source of validity evidence put forward by the American Psychological and Education Research Associations.²¹ Such studies should investigate which factors influence the outcomes of the MCTQ and the implications for practice. Study questions to be addressed are: Do clinical teachers accept their scores on the MCTQ? How can they be stimulated to reflect on their scores? Does their behaviour really improve as a result of receiving feedback? In conclusion, our study shows that the MCTQ provides teaching staff with a good basis for reflection on their clinical teaching effectiveness. Further studies should focus on the implementation of the MCTQ in practice and its effectiveness in improving teaching.

REFERENCES

1. Smith BP, Walsh DA. Teaching the art of clinical practice: The veterinary medical teaching hospital, private practice, and other externships. *J Vet Med Educ* 2003;30:203-6.
2. Spencer J. ABC of learning and teaching in medicine: Learning and teaching in the clinical environment. *Brit Med J* 2003;326:591-4.
3. Daelmans HEM, Hoogenboom RJI, Donker AJM, Scherpbier AJJA, Stehouwer CDA, van der Vleuten CPM. Effectiveness of clinical rotations as a learning environment for achieving competences. *Med Teach* 2004;26:305-12.
4. Ramani S, Leinster S. AMEE guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30:347-64.
5. Van der Hem-Stokroos HH, van der Vleuten CPM, Daelmans HEM, Haarman HJTM, Scherpbier AJJA. Reliability of the clinical teaching effectiveness instrument. *Med Educ* 2005;39:904-10.
6. Leinster S. Learning in the clinical environment. *Med Teach* 2009;31:79-81.
7. Dolmans DHJM, Wolfhagen HAP, Gerver WJ, de Grave W, Scherpbier AJJA. Providing physicians with feedback on how they supervise students during patient contacts. *Med Teach* 2004;26:409-14.
8. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: Maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000;34:820-6.
9. Lane IF, Strand E. Clinical veterinary education: Insights from faculty and strategies for professional development in clinical teaching. *J Vet Med Educ* 2008;35:397-406.
10. Steinert Y. Staff development for clinical teachers. *The Clinical Teacher* 2005;2:104-10.
11. Steinert Y, Mann K, Centeno A, Dolmans DHJM, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach* 2006 09;28:497-526.
12. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998;73:688-95.
13. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000;75:161-6.
14. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
15. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:272-7.
16. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *American Educator* 1991;15:6-11.
17. Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: A literature review. *Med Educ* 2000;34:827-40.

18. Stalmeijer RE, Dolmans DHJM, Wolfhagen IHAP, Muijtjens AMM, Scherpbier AJJA. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010 ;85:1732-8.
19. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005;20:1159-64.
20. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004;9:971-7.
21. American Education Research Association and American Psychological Association. Standards for educational and psychological testing. Washington DC: American Education Research Association; 1999.
22. Downing SM. Validity: On meaningful interpretation of assessment data. *Med Educ* 2003;37:830-7.
23. Byrne BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum; 2001.
24. Crossley J, Davies H, Humphris G, Jolly B. Generalisability: A key to unlock professional assessment. *Med Educ* 2002;36:972-8.
25. Brennan RL. Generalizability theory. New York: Springer; 2001.
26. Gronlund NE. How to construct achievement tests. Englewood Cliffs, NJ: Prentice-Hall; 1988.
27. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Scherpbier AJJA. Cognitive apprenticeship in clinical practice: Can it stimulate learning in the opinion of students? *Adv Health Sci Educ* 2009;14:535-46.
28. McDonald RP, Ringo Ho M. Principles and practice in reporting structural equation analyses. *Psychol Methods* 2002;7:64-82.
29. Korthagen FAJ, Vasalos A. Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice* 2005;11:47-71.
30. Beckman TJ, Mandrekar JN. The interpersonal, cognitive and efficiency domains of clinical teaching: Construct validity of a multi-dimensional scale. *Med Educ* 2005;39:1221-9.

APPENDICES

1. The modified Maastricht Clinical Teaching Questionnaire

Modelling

The clinical teacher...

- demonstrated how different skills should be performed.
- explained, while performing a task, which aspects were important and why.
- created sufficient opportunities for me to observe him or her.
- was a role model for me.

Coaching

The clinical teacher...

- observed me while I was performing a task.
- provided me with constructive and concrete feedback during or following direct observation.
- gave me a better insight into aspects of my performance that needed improvement.

Scaffolding

The clinical teacher...

- adjusted his/her teaching activities to my level of experience and competence.
- allowed me to perform tasks independently.
- was supportive when I experienced difficulties with a task.
- gradually decreased the amount of guidance in order to bolster my independence.

Articulation

The clinical teacher...

- asked me to explain my reasoning and actions.
- alerted me to gaps in my knowledge and skills.
- asked questions to increase my knowledge and understanding.
- stimulated me to ask questions to increase my knowledge and understanding.

Reflection

The clinical teacher...

- stimulated me to think about my own strengths and weaknesses.
- stimulated me to think about how to improve my own strengths and weaknesses.

Exploration

The clinical teacher...

- stimulated me to formulate my own goals.
- stimulated me to achieve my own goals.
- challenged me to explore new tasks and possibilities.

General Learning Climate

The clinical teacher...

- established an environment where I felt free to ask questions or make comments.
- took enough time to supervise me.
- showed an interest in me as a student.
- treated me with respect.

Miscellaneous

- Give an overall mark (1–10) for this doctor as a clinical teacher.
- What are the strengths of this clinical teacher? (open-ended question)
- Which aspects of the performance of this clinical teacher can be improved? (open-ended question)
- How many hours of actual contact time did you have with your supervisor?

4.

Evaluating clinical teachers with the MCTQ: how much 'teacher' is in student ratings?

T.B.B. Boerboom

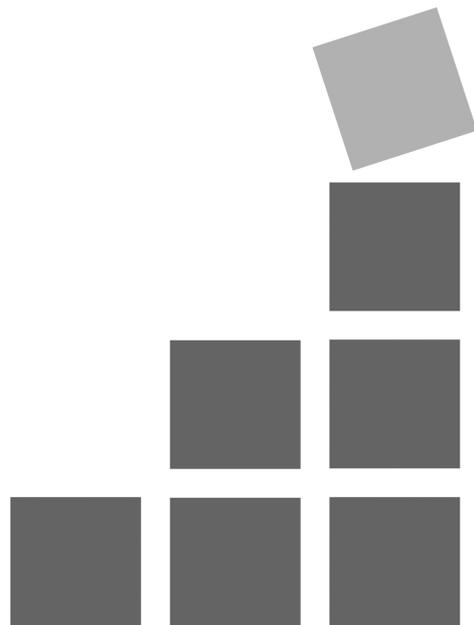
M.T. Mainhard

D.H.J.M. Dolmans

A.J.J.A. Scherpbier

P. van Beukelen

A.D.C. Jaarsma



2011 Under review

ABSTRACT

Context

Students are a popular source of data to evaluate the performance of clinical teachers. Instruments to obtain student evaluations must have proven validity, particularly when evaluations are used to inform high stakes decisions, such as promotion or tenure. One aspect of validity that often remains underexposed is the possibility of effects of between-student differences and teacher and student characteristics not directly related to teaching performance. We examined the occurrence of such effects, using multilevel analysis to analyse data from the Maastricht Clinical Teaching Questionnaire (MCTQ), a validated evaluation instrument, in a veterinary curriculum.

Methods

We analysed a dataset of 1277 MCTQ's completed by 208 students evaluating 168 clinical teachers. The 15-item MCTQ covers five domains: Learning Climate, Modelling, Coaching, Articulation and Exploration. Multilevel analysis was used to divide the variance in the domain scores in components related to respectively teachers and students. Subsequent models were estimated to explore how the MCTQ scores are dependent on teacher and student characteristics, like gender, department, or clerkship year.

Results

Significant amounts of variance in student ratings were due to between-teacher differences, particularly for Learning Climate, Modelling, and Coaching. The effects of teacher and student characteristics were mostly non-significant or small.

Conclusions

Large portions of variance in MCTQ scores were due to differences between teachers, while the contribution of student and teacher characteristics was negligible. The results support the validity of student ratings obtained with the MCTQ for evaluating teacher performance. Future studies should identify student and teacher characteristics that do account for differences in student ratings.

INTRODUCTION

Student ratings are a popular and indispensable source of information about the quality of the teaching of individual clinical teachers.¹⁻³ As the only group with first-hand experience of clinical teaching, students are the obvious source of information about the quality of their teachers.^{4,5} Undoubtedly, the general ease of administration and the affordability of instruments to obtain student ratings also contribute to the popularity of evaluation instruments based on student ratings.³ This has led to widespread interest in these instruments, which is reflected in numerous published studies conducted at universities all over the world.⁶

Evaluations of clinical teaching can serve several purposes. Measurements of teaching effectiveness can be used to provide teachers with feedback to guide, support, and motivate them to improve their teaching.^{4,7,8} Also, they are widely used to underpin decisions about teachers' promotion and tenure.^{3,4} In the latter case in particular, it is of vital importance that these instruments are properly validated.^{2,9}

Many studies have reported on the robustness and validity of instruments to obtain student ratings of clinical teachers.^{2,4,6,7,10-16} The most frequently reported evidence relates to content and construct validity,¹² but some aspects of reliability are regularly overlooked, although they are central to the quality of measurements. One such aspect is the degree to which student evaluations are indeed influenced by the teachers' practices as opposed to other factors.¹⁷ It is important to determine this, because identification of characteristics that influence the outcomes of evaluation instruments have implications for how these outcomes can be used³ and how they should be interpreted to assist clinical teachers in improving their clinical supervising practice. The same information can also be beneficial to coaching and remediation activities as well as selection, promotion and tenure of teaching faculty. Clearly, when an instrument is used to evaluate the quality of clinical teachers, it is vital to know whether significant amounts of variance are due to differences between teachers rather than differences between students. The latter type of differences can be, for example, related to students' individual preferences or learning styles or to student interactions with the scales and items of the instrument at hand.¹⁷

The degree to which variance in evaluations is due to teachers or students has been studied, but mostly in classroom settings in elementary and secondary education and not in health care education.^{3,6,9,18} A subsequent step to determining the

MULTILEVEL ANALYSIS

degree to which variance in evaluations is due to teachers or students, is determining how this variance can be explained. Examples of variables, believed to influence student ratings, are the teacher's experience, certification and speciality, and the interaction between the student's and teacher's gender.^{9,19-22} It is likely that these aspects also have an effect on student ratings in a healthcare setting. The impact of non-teaching related aspects is one of the key issues in the validity of student ratings of teaching, which has remained underexposed in medical education research.

Considering that evidence on the discussed aspects would strengthen the validity of student ratings and elucidate their implications, we designed a study with the following two aims: 1) to determine how much variance in the Maastricht Clinical Teaching Questionnaire (MCTQ) ratings is due to differences between teachers versus differences between students; and 2) to explore the degree to which MCTQ ratings can be explained by specific teacher and student characteristics: the teacher's experience in clinical teaching, the teacher's teaching qualification, teacher and student gender, the teacher's department, hours of clinical teaching and the student's clerkship year.

The MCTQ is a validated 15-item instrument based on the cognitive apprenticeship model, which asks students to rate their clinical teachers' performance in five domains.¹⁴ We chose this particular instrument because the validity and reliability of the MCTQ already has been tested using confirmatory factor analysis and generalizability studies in undergraduate medical and veterinary settings,^{16,23} but it remained unexplored to which degree variance in these scores was due to differences between teachers or students, as well as the influence of teacher and student characteristics, such as gender, experience and year of study.

METHODS

Context

The study was conducted at the Faculty of Veterinary Medicine, Utrecht University (FVMU), the Netherlands. The 6-year undergraduate curriculum offers four years of preclinical training and two years of clinical clerkships. Year 5 consists of thirty weeks of rotations in different clinical departments and in year 6 students undertake rotations in disciplines related to their chosen animal species track. Clinical rotations in veterinary medicine are comparable to rotations in the medical

setting, but they are often shorter, varying between 1 day and 6 weeks, because discipline specific rotations (gynaecology for example) are repeated for different animal species. In order to obtain tenure, teaching faculty can attend a two-year faculty development programme including personal mentoring, which leads to the nationally recognised University Teaching Qualification (UTQ).

Instrument

Data were collected using the Maastricht Clinical Teaching Questionnaire (MCTQ). The MCTQ asks students to rate (on a five-point scale) their clinical teachers' performance in five domains:¹⁴ 1) General Learning Climate: The teacher creates a positive and safe environment where students are treated with respect and feel free to ask questions and make comments. 2) Modelling: The teacher acts as a role model, demonstrating how different skills should be performed and at the same time explaining which aspects are important. 3) Coaching: The teacher allows students to perform tasks independently, is supportive when students experience difficulties with a task and gives constructive and concrete feedback. 4) Articulation: The teacher asks students to articulate their knowledge and reasoning and stimulates and challenges them to deepen and expand their knowledge and understanding. 5) Exploration: The teacher stimulates students to formulate learning goals, based on identified strengths and weaknesses, and to think about how to achieve these goals. The construct validity of the instrument was satisfactory.²³ The fit of the 15-item, 5-domain model for the sample employed here was demonstrated in a confirmatory factor analysis with the following modification indices:²⁴ a) χ^2 divided by degrees of freedom (CMIN/df)=2.594; b) the goodness-of-fit index (GFI)=0.82; c) the comparative fit index (CFI)=0.93 and d) the standardized root mean square residual (SRMR)=0.049. Cronbach's alphas for the five scales were high (0.87-0.96) and the bilateral correlation coefficients ranged from $r=0.47$ to $r=0.71$ (all $ps < 0.001$). A generalizability study on 1190 student ratings pointed out that 10-12 ratings were necessary for reliable outcomes on the domain level. For the present study, we collected MCTQ ratings and additionally asked students to estimate how much time the teacher devoted to supervising them (hours of guidance).

Procedure

All fifth- and sixth-year students undertaking rotations in the Equine Health Department, the Companion Animal Health Department, the Farm Animal Health Department and the Pathology Department between March 2009 and February 2010 were invited by email to evaluate their clinical teachers using an online version of the MCTQ. Students could evaluate several teachers, using a separate questionnaire for each teacher. They were encouraged to complete a questionnaire as soon as possible after a student-teacher encounter.

Sample

A total of 1277 evaluations were handed in (168 different clinical teachers, median=6.0, range=1-40 evaluations per teacher) by a total of 208 students (median=4.0, range=1-48 evaluations per student). We selected the following variables, which have been shown to co-vary with teacher evaluations in primary and secondary education, and added them to the data set: teacher experience,⁶ teacher and student gender,^{3,18} the teacher's department,¹⁵ clerkship year (year 5 or 6),²⁵ and whether a teacher had obtained the UTQ.²⁶

Of the participating students, 82.7% were female, which adequately reflects the gender distribution in the study population. Teachers from all four clinical departments: the Equine Health Department (35 teachers), the Companion Animal Health Department (69 teachers), the Farm Animal Health Department (54 teachers) and the Pathology Department (10 teachers) were evaluated and the numbers of teachers accurately reflect the size of the departments. 66.9% of the evaluations were completed by fifth-year students and 33.1% by sixth-year students. Most clinical teachers supervised students in both years and received evaluations from both groups. Teachers had on average 10.4 years of clinical teaching experience (SD=10 years); 48.8% were female and 45.2% had the UTQ.

Ethics

Participation of the students was voluntary and participants could opt out of the study at any moment. Confidentiality was assured and it was communicated to teachers and students that the questionnaires were coded in such a way that data could not be traced back to individuals and that the data would be used only for research purposes. The researchers took all the precautions to protect both the

students and teachers from harm which could follow from the data collection, interpretation and presentation.

Data analysis

In this study, multilevel analyses were employed to answer the research questions. Multilevel analysis takes interdependence of ratings into account, in the present case for example, students are nested within teachers.²⁷ While, for example, generalizability studies, suitable for determining how many measurements are needed for reliable teacher evaluation outcomes, assume that every single measurement or rating is independent of the other. Multilevel modelling disentangles these dependencies by quantifying the degree to which variance in student ratings is due to differences between individual students or to the higher level construct that is being evaluated (teachers, programmes) and within which ratings are nested.^{28,29} This is also referred to as variance decomposition. The multilevel method also allows the inclusion of explanatory variables (e.g. teacher and student characteristics) that may explain differences between teachers or between students. Another attractive feature of multilevel analysis is that the effects of explanatory variables can be estimated not only as a fixed effect (as in ordinary regression analysis) but also as a random effect. This means that it is possible to estimate not only the average effect of, for example, teacher experience on student ratings, but also the degree to which this effect differs across teachers (random slopes).²⁸ For example, while on average older teachers may become more effective in providing a positive learning climate, some older teachers may however become disengaged and therefore less effective in their teaching. This might lead to significant variation between teachers in the effect of teaching experience on the learning climate.

The dataset in this study was analysed with the multilevel package MLwiN.³⁰ In the current study, the relationships among teacher characteristics, student characteristics and the MCTQ scores were analysed per domain.

Research question 1: teacher versus student variance

We used the MCTQ domain scores as dependent variables to determine how much of the variance in these ratings was due to differences between teachers versus students. For each domain score, a model without explanatory variables (student or teacher characteristics) was estimated (i.e. the empty model). This decomposed the variance in MCTQ domain scores into between-teacher (level 2) and between-

MULTILEVEL ANALYSIS

student (level 1) variance. Subsequently intra-class correlations (ICCs) were calculated, which represent the proportion of the total variance that resided at level 2. ICCs are interpreted as the degree to which individual student ratings are affected by the grouping under the level 2 variable (i.e. the teacher). ICCs of 20% or more are considered to indicate a large grouping effect.²⁸

Research question 2: teacher and student characteristics and MCTQ ratings

In order to explore the degree to which variation in MCTQ scores could be explained by teacher (level 2) and student (level 1) characteristics, these variables were added to the multilevel models: department (dummy coded), teacher experience (in years, grand mean centred), teacher gender (0=female, 1=male), student gender (0=female, 1=male), clerkship year (0=year 5, 1=year 6), and teaching qualification (UTQ; 0=no, 1=yes).

RESULTS

Research question 1: teacher versus student variance

ICCs of 27%, 24% and 24% for the domains Learning Climate, Modelling and Coaching, respectively, indicated that these scores were largely affected by the teachers. For Articulation (16%) and Exploration (12%) the ICCs were somewhat lower (see Table 1 for variance components and ICCs). The remaining variance was due to differences between individual students and random measurement error.

	Climate	Modelling	Coaching	Articulation	Exploration
Level 2 – between teacher variance	0.28***	0.21***	0.21***	0.11***	0.10***
Level 1 – between student variance	0.76***	0.67***	0.68***	0.57***	0.76***
ICC	27%	24%	24%	16%	12%

Note: The ICC represents the proportion of total variance that resides at level 2.
*** p<.001

Table 2: Multilevel models for each of the five factors of the MCTQ including explanatory variables

	Climate		Modelling		Coaching		Articulation		Exploration	
	Fixed effects									
	B	β	B	β	B	β	B	β	B	β
Intercept	3.92***		3.75***		3.68***		3.80***		3.20***	
UTQ	0.13		0.05		0.01		0.14		0.11	
Teacher experience	-0.02**	.19	-0.01*	.10	-0.01*	.10	-0.01		-0.01**	.10
Clerkship year (yr 5-6)	0.10***	.05	0.09		0.20**	.10	0.19***	.11	0.11*	.06
Hours of guidance	0.01		0.01**	.13	0.01*	.14	<0.01		<0.01	
	Random effects									
Level 2										
Between teacher variance	0.24***		0.21***		0.21***		0.11***		0.10**	
Random slope variance Clerkship year	0.23***		n.a.		n.a.		n.a.		n.a.	
% reduction of between teacher variance¶	14.29		<1.00		<1.00		<1.00		<1.00	
Level 1										
Between student variance	0.72**		0.65***		0.66***		0.57***		0.74***	
% reduction of between student variance¶	5.26		2.99		2.94		<1.00		2.63	

Note. N=168 Teachers and 208 students. The dependent variables are the five domains of the MCTQ; their theoretically possible range is 1-5. Teacher experience and Hours of guidance were grand mean centered. Bs represent regression weights and β s the effect sizes (i.e. standardized effects).

¶ reduction of variance as compared to the empty model presented in Table 1

* $p < .05$. ** $p < .01$. *** $p < .001$

Research question 2: teacher and student characteristics and MCTQ scores

The multilevel models for the five domains and the selected explanatory variables are summarised in Table 2. Gender and department variables yielded no significant results for any of the models, and so, for reasons of parsimony, we did not include them in Table 2. There were only relatively small effects of the explanatory variables (UTQ, clerkship year, hours of guidance) on the domain scores. Inclusion of the explanatory variables reduced the between-teacher and between-student variance only marginally, with the exception of the learning climate model, which resulted in a 14% reduction in between-teacher variance (see Table 2). All five models showed a negative relationship between teacher experience and domain ratings. The sixth-year students gave significantly higher ratings compared to the fifth-year students on four domains, and the effects of UTQ and supervision time were either non-significant or significant but very small.

The effect of clerkship year on learning climate ratings differed significantly between teachers (see 'random slope clerkship year' in Table 2). The fixed effect of clerkship year indicated that sixth-year students generally gave higher ratings on learning climate compared to fifth-year students ($B=0.10$), but this effect differed quite substantially across teachers, with the average (fixed) effect of clerkship year ($0.10 \pm$ two SDs; $SD = \sqrt{B_{\text{random}}} = \sqrt{0.23}$) ranging from -0.86 to 1.06 ($=0.10 \pm (2 \times \sqrt{0.23})$); for the computations see Snijders & Bosker, 1999, p. 71).²⁸ This indicates that a significant number of teachers received lower ratings on learning climate from sixth-year students, despite an average positive effect of clerkship year on learning climate.

DISCUSSION

We examined the extent to which student ratings of clinical teachers reflected differences between teachers rather than differences between students and we examined the effects of several teacher and student characteristics on different domains of clinical teaching. The main finding is that student ratings were largely affected by differences between teachers. We discuss the findings in more detail for each of the research questions.

Research question 1: teacher versus student variance

Significant amounts of the variance in the student ratings of all domains were due to differences between teachers. A fairly large amount of the variance ($>20\%$) in

the ratings of General Learning Climate, Modelling and Coaching, in particular, was related to differences between teachers. This means that differences between teachers, with regard to these domains of teaching, had a substantial effect on student ratings. It is therefore reasonable to conclude that student ratings of the MCTQ domains are valid indicators of teaching performance. These results are consistent with findings from other studies exploring which teaching behaviours are important for teaching in the clinical workplace. These studies reported beneficial effects on teaching effectiveness for a positive learning climate,^{31,32} role modelling,³³⁻³⁵ opportunities for students to practice and the provision of rich feedback.^{5,8}

For the domains Exploration and Articulation, the differences between teachers were somewhat less marked, but still substantial. Articulation focuses on the teacher probing the students for knowledge and understanding. Exploration is the domain where students are pushed in a mode of problem solving and reflection.^{14,23} One might speculate that these domains are more susceptible to effects of differences between students in problem-solving skills and reflective attitude, because they contain items relating to active participation by students, such as 'The clinical teacher asked me to explain my reasoning and actions' and 'The clinical teacher stimulated me to formulate my own goals'. Nevertheless, the ICCs of these domain scores indicate that the MCTQ is a useful instrument for evaluating clinical teachers.

The ICCs found in this study do not indicate perfect similarity of scores that any two students would assign to a certain teacher (i.e. an ICC of 1.00), and still significant amounts of variance are due to measurement error and differences between students. This would however be beyond any expectation. In numerous studies in a variety of research area's ICCs are used to reflect the component of variance attributable to a grouping variable in addition to the usual variation attributable to individuals.²⁷ When the results of such studies^{27,36-40} are taken into consideration, the magnitude of the ICCs we found (between 12% and 27%) can be qualified as seemingly substantial.

Research question 2: teacher and student characteristics and MCTQ scores

The teacher's gender and department offered no significant explanation for differences between student ratings, while teaching qualification (UTQ), teacher experience, clerkship year, and hours of guidance were either marginally or non-significantly associated with student ratings. The absence of significant relationships between MCTQ ratings and student- and teacher-characteristics indicates that student ratings reflect teaching practice rather than other teacher characteristics.

MULTILEVEL ANALYSIS

This suggests that our results do not lend support to other studies which discussed the influence of the interaction between student and teacher gender on student ratings of teaching effectiveness.^{9,18,41} However, this result also indicates that there are other possible influencing characteristics which we did not include in this study. Future research should explore these characteristics.

The MCTQ was validated using pooled data from teachers and students in different departments and disciplines.²³ According to Beckman et al., the use of a single instrument to evaluate clinical teaching effectiveness across different departments and disciplines can be a major source of unwanted variance.^{2,42} The absence of a significant department-related effect on MCTQ ratings in our study could be interpreted as an indication that the cognitive apprenticeship model, underlying the MCTQ domains, reflects a uniform principle that fits the different educational settings of the institute.¹⁴ An alternative explanation could be that teaching practices at FVMU are more uniform than at other institutions. The number of hours of teaching was not associated with MCTQ ratings.

There is no clear explanation for the relatively small negative effect of teaching experience on all five domain ratings. Marsh stated that a decline in teaching effectiveness with increasing teaching experience can be partly due to an absence of continuing faculty development activities.⁶ At FVMU, faculty development consists exclusively of a two-year faculty development programme leading to the UTQ,²⁶ which showed no positive relationship with MCTQ ratings. Further research is needed to clarify effects of faculty development on MCTQ ratings.

The results of the analyses indicate that it is worthwhile to examine random slopes. Unlike a random intercept model, in a random slope model each teacher's regression line may have a different slope. This allows explanatory variables (here clerkship year) to have a different effect for each teacher. In other words, if statistically significant random slopes are found, the average effect of a variable is not generalizable across the entire sample. Our results showed that, on average, sixth-year students gave higher ratings than fifth-year students. The higher ratings from sixth-year students could be interpreted as an indication that the required degree of cognitive apprenticeship varies depending on the level of training. In the last phase of their clinical rotations, for example, students may be better able to fully value Articulation and Exploration. However, the random slopes indicated that this trend generalized poorly across teachers. While some teachers received systematically higher ratings for Climate in year 6, others received lower ratings. We will illustrate this by an example of two teachers receiving ratings of 3 (five-

point scale) on General Learning Climate from fifth-year students. Based on the average effect, the obvious conclusion would be that both teachers are likely to get marginally higher ratings for sixth-year rotations, but the significant random slopes indicate that it is also possible that one teacher may be rated almost one scale point higher by sixth-year students, while the other teacher may score a whole scale point lower. An example of how such an effect arises might be that some teachers are exclusively involved in sixth-year's internships that are, for example, shorter and therefore do not allow to create good quality climates. More in-depth research is required to further examine the nature of the random slopes identified here.

Further, employing the cognitive apprenticeship model⁴³ in order to explain variance in MCTQ scores due to students may be a fruitful avenue for further research. This model describes a teaching method where the generally tacit cognitive processes of experts are made explicit in different steps. In sum, students first observe the expert performing a complex task and then try to learn to perform the task themselves under supervision of the expert. As a final step, the students are stimulated to make their own cognitive processes explicit and to master the task through a reflection cycle. This model strongly suggests that the steps are followed in this order. However, it may be possible that this order does not fit every student's learning style.⁴⁴ For example, some students prefer to just try to perform the task prior to receiving an explanation on how to perform it.

A limitation of this study is that only one instrument was used to generate the data, which may have given rise to a halo effect.² It cannot be ruled out that the ratings on Climate and Modelling, which largely reflect interpersonal aspects of teaching, inflated the ratings on the other domains, despite the fact that the MCTQ items were specifically developed to measure concrete teaching behaviours ('The clinical teacher stimulated me to formulate my own goals') rather than student opinions ('The clinical teacher is good in letting us formulate goals'). Future studies should confirm whether and to what extent higher ratings on the different MCTQ domains discriminate between outstanding and weaker clinical teaching performance. Another limitation is the fact that all measurements were conducted at a single institution. This potentially affects the applicability of the findings and necessitates repeating this research in other settings.

A strength of this study is the use of a well validated instrument^{16,23} for the collection of the data. The main conclusions which can be drawn based on this study are twofold. Firstly, teachers have a rather large and significant influence on how

MULTILEVEL ANALYSIS

students value them on an instrument like the MCTQ. Note that this is especially true for General Learning Climate, Modelling and Coaching and to a somewhat lesser degree for Articulation and Exploration. Secondly this study produced a first indication of the fact that gender, department, teaching qualification, and hours of guidance do not influence the student ratings of these domains. Still, large portions of variance remain unexplained. The interpretation of student ratings of their clinical teachers will be further enhanced when future research is able to identify student and teacher characteristics that account for differences in student ratings.

REFERENCES

1. Irby DM, Ramsey PG, Gillmore GM, Schaad D. Characteristics of effective clinical teachers of ambulatory care medicine. *Acad Med* 1991;66:54-5.
2. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004;9:971-7.
3. McOwen KS, Bellini LM, Guerra CE, Shea JA. Evaluation of clinical faculty: Gender and minority implications. *Acad Med* 2007;82:94-6.
4. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
5. Dolmans DHJM, Wolfhagen HAP, Essed GGM, Scherpbier AJJA, van der Vleuten CPM. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. *Acad Med* 2002;77:332-5.
6. Marsh HW. Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *J Educ Psychol* 2007;99:775-90.
7. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000;75:161-6.
8. Ramani S, Leinster S. AMEE guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30:347-64.
9. Centra JA, Gaubatz NB. Is there gender bias in student evaluations of teaching? *J Higher Educ* 2000;71:17-33.
10. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med* 1998;73:688-95.
11. Williams BC, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med* 2002;77:177-80.
12. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005;20:1159-64.
13. Van der Hem-Stokroos HH, van der Vleuten CPM, Daelmans HEM, Haarman HJTM, Scherpbier AJJA. Reliability of the clinical teaching effectiveness instrument. *Med Educ* 2005;39:904-10.
14. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:272-7.
15. Fluit C, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers. *J Gen Intern Med* 2010;25:1337-45.
16. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010 ;85:1732-8.
17. Downing SM. Threats to the validity of clinical teaching assessments: What about rater error? *Med Educ* 2005;39:353-5.

MULTILEVEL ANALYSIS

18. Leone-Perkins M, Schnuth R, Kantner T. Preceptor-student interactions in an ambulatory clerkship: Gender differences in student evaluations of teaching. *Teach Learn Med* 1999;11:164-7.
19. Goldhaber DD, Brewer DJ. Evaluating the effect of teacher degree level on educational performance. *Developments in school finance* 1996;199:197-210.
20. Marsh HW, Roche LA. Making students' evaluations of teaching effectiveness effective. *Am Psychol* 1997;52:1187-97.
21. Aleamoni LM. Student rating myths versus research facts from 1924 to 1998. *J Pers Eval Educ* 1999;13:153-66.
22. Darling-Hammond L, Youngs P. Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher* 2002;31:13-25.
23. Boerboom TBB, Dolmans DHJM, Jaarsma ADC, Muijtjens AMM, van Beukelen P, Scherpbier AJJA. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. *Med Teach* 2011;33:e84-e91.
24. Byrne BM. *Structural equation modeling with AMOS: Basic concepts, Applications, and Programming*. Mahwah, NJ: Lawrence Erlbaum; 2001.
25. Ullian JA, Bland CJ, Simpson DE. An alternative approach to defining the role of the clinical teacher. *Acad Med* 1994;69:832-8.
26. Boerboom TBB, Dolmans DHJM, Muijtjens AMM, Jaarsma ADC, van Beukelen P, Scherpbier AJJA. Does a faculty development programme improve teachers' perceived competence in different teacher roles? *Med Teach* 2009;31:1030-1.
27. Murray DM, Phillips GA, Birnbaum AS, Lytle LA. Intraclass correlation for measures from a middle school nutrition intervention study: Estimates, correlates, and applications. *Health Educ Behav* 2001;28:666-79.
28. Snijders TAB, Bosker RJ. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage; 2004.
29. Dansereau F, Yammarino FJ. *Multi-level issues in strategy and methods*. Bingley, UK: Emerald Group Publishing Limited; 2005.
30. Rasbash J, Browne WJ, Healy M, Cameron B, Charlton C. *MLwiN*. 2005; version 2.02.
31. Boor K, Scheele F, van der Vleuten CPM, Teunissen PW, den Breejen EME, Scherpbier AJJA. How undergraduate clinical learning climates differ: A multi-method case study. *Med Educ* 2008;42:1029-36.
32. Dolmans DHJM, Wolfhagen HAP, Heineman E, Scherpbier AJJA. Factors adversely affecting student learning in the clinical learning environment: A student perspective. *Educ Health* 2008;20:1-10.
33. Mullan P, Sullivan D, Dielman T. What are raters rating? Predicting medical student, pediatric resident, and faculty ratings of clinical teachers. *Teach Learn Med* 1993;5:221-6.
34. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: Maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000;34:820-6.
35. Elnicki DM, Cooper A. Medical students' perceptions of the elements of effective inpatient teaching by attending physicians and housestaff. *J Gen Intern Med* 2005;20:635-9.

36. Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: Data from the health survey for England 1994. *Am J Epidemiol* 1999;149:876-83.
37. Scheier LM, Griffin KW, Doyle MM, Botvin GJ. Estimates of intragroup dependence for drug use and skill measures in school-based drug abuse prevention trials: An empirical study of three independent samples. *Health Educ Behav* 2002;29:85-103.
38. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005;58:246-51.
39. Littenberg B, MacLean CD. Intra-cluster correlation coefficients in adults with diabetes in primary care practices: The Vermont diabetes information system field survey. *BMC Med Res Methodol* 2006;6:20.
40. Pals SL, Beaty BL, Posner SF, Bull SS. Estimates of intraclass correlation for variables related to behavioral HIV/STD prevention in a predominantly African American and Hispanic sample of young women. *Health Educ Behav* 2009;36:182-94.
41. Beckman TJ, Mandrekar JN. The interpersonal, cognitive and efficiency domains of clinical teaching: Construct validity of a multi-dimensional scale. *Med Educ* 2005;39:1221-9.
42. Beckman TJ, Cook DA, Mandrekar JN. Factor instability of clinical teaching assessment scores among general internists and cardiologists. *Med Educ* 2006;40:1209-16.
43. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *American Educator* 1991;15:6-11.
44. Hawk TF, Shah AJ. Using learning style instruments to enhance student learning. *Decision Sciences Journal of Innovative Education* 2007;5:1-19.

5.

Peer group reflection helps clinical teachers to critically reflect on their teaching

T.B.B. Boerboom

A.D.C. Jaarsma

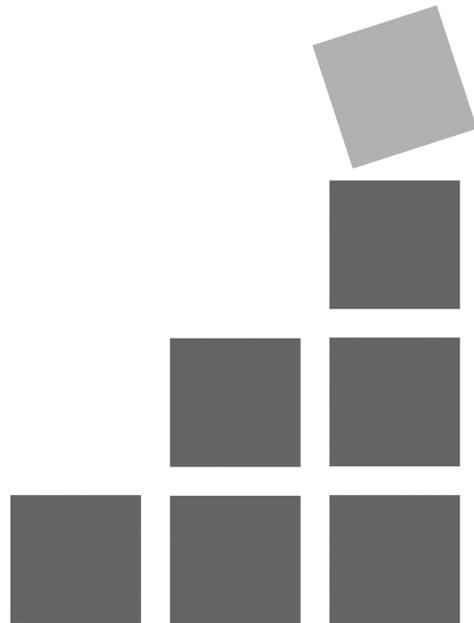
D.H.J.M. Dolmans

A.J.J.A. Scherpbier

N.J.J.M. Mastenbroek

P. van Beukelen

2011 Under review



ABSTRACT

Context

Student evaluations can help clinical teachers to reflect on their teaching skills and find ways to improve their teaching. Studies have shown that the mere presentation of student evaluations is not a sufficient incentive for teachers to critically reflect on their teaching. We evaluated and compared the effectiveness of two feedback facilitation strategies that were identical except for a peer reflection meeting.

Methods

Fifty-four clinical teachers were randomly assigned to two feedback strategies. In one strategy, a peer reflection was added as an additional step. All teachers completed a questionnaire evaluating the strategy they had experienced. We analysed the reflection reports and the evaluation questionnaire.

Results

Both strategies stimulated teachers to reflect on feedback and formulate alternative actions for their teaching practice. The teachers who had participated in the peer reflection meeting showed deeper critical reflection and more concrete plans to change their teaching. All feedback strategies were considered effective by the majority of the teachers.

Conclusions

Strategies with student feedback and self-assessment stimulated reflection on teaching and helped clinical teachers to formulate plans for improvement. A peer reflection meeting seemed to enhance reflection quality. Further research should establish whether it can have lasting effects on teaching quality.

INTRODUCTION

Student evaluations are one of the methods commonly used to assess clinical teaching skills.¹⁻⁷ They can be used for performance appraisal and promotion up the academic ladder,⁸⁻¹⁰ but also to improve teaching practice.¹¹ In a review study, Marsh and Roche found evidence of the usefulness of student evaluations in improving teaching effectiveness, provided the utilised instruments were thoroughly validated.¹² Student evaluations can not only help teachers to identify areas for improvement,^{11,13} but also stimulate teachers to reflect on their teaching practice and plan concrete actions for improvement.

Many authors have endorsed the importance of reflection for enhancing professional growth.¹⁴⁻¹⁸ Especially Schön's concept that professionals who reflect on their actions learn in a more profound way has been influential in the literature.¹⁹ There is still some debate over the definition of reflection.¹⁵ In this paper we use a straightforward and useful one given by Hatton and Smith: "Reflection is deliberate thinking about action with a view to its improvement".¹⁴

A widely used model for reflection is Korthagen's ALACT model,¹⁶ which proposes a three-step process to promote reflection after experiencing a certain event: 1) Looking back on the action; 2) Awareness of essential aspects; 3) Creating alternative methods of action. This last step is an important aspect of reflection.¹⁶ When teachers are able to critically reflect upon their practice, they are more likely to take action to improve it.¹⁸ In studies on reflection, however, the link between reflection and practice improvement often remains underexposed.¹⁵

Various authors have pointed out that merely reading an analysis of ratings by students or being aware of weaknesses does not automatically incite teachers to reflect on and look for ways to improve their teaching practice.²⁰⁻²² Moreover, there is evidence that unfacilitated feedback can cause negative emotional reactions, like denial and defensiveness, which can even raise barriers to reflection²³⁻²⁵ and behavioural change.²⁴ This problem can be overcome by facilitating feedback²⁴ so as to stimulate feedback recipients to engage in constructive reflection on their practice.^{22,23} Self-assessment and peer meetings are ways to facilitate feedback^{26,27} and several authors have highlighted their importance in facilitating teachers' professional growth.²⁶⁻³⁰ There are, however, no published studies comparing the impact of different feedback strategies on the quality of self-reflection. Since peer reflection meetings are generally more time consuming and costly than self-assessment, it can be a challenge to gain acceptance

TEACHER REFLECTION

and funding for this type of faculty development.³⁰ It is therefore important to provide evidence of its impact.

The present study

Hatton & Smith defined four levels of reflective writing, ranging from descriptive writing to critical reflection.¹⁴ Critical reflection is considered especially beneficial, because it takes account of aspects of the working environment, such as curriculum design, student characteristics and institutional culture.¹⁴ These aspects are considered highly relevant for teachers' functioning.¹⁷

We conducted a mixed method study to compare the effects of two feedback facilitation strategies on the quality of reflection by teachers. Both strategies consisted of a self-evaluation questionnaire, combined feedback from student evaluations and the self-assessment, and writing a self-reflection report on teaching practice. The strategies differed in the addition of a peer reflection meeting to one of the strategies. We used the modified Maastricht Clinical Teaching Questionnaire (MCTQ) to elicit student feedback and the self-evaluation. The MCTQ is a validated instrument, based on the Cognitive Apprenticeship model and consisting of fifteen items in five domains: 1) General Learning Climate, 2) Modelling, 3) Coaching, 4) Articulation and 5) Exploration.^{31,32}

We sought answers to the following research questions:

1. What types of reflection do clinical teachers demonstrate in a pre-structured, written reflection report after a feedback facilitating strategy comprising self-assessment and student feedback and does reflection change when a peer reflection meeting is added to the feedback facilitation strategy?
2. How do clinical teachers evaluate the two different feedback facilitation strategies?

METHODS

Context

The study was conducted at the Faculty of Veterinary Medicine, Utrecht University, the Netherlands (FVMU) between November 2009 and February 2010. FVMU offers a six-year undergraduate curriculum with four years of preclinical training followed by two years of clinical clerkships.

Participant selection

Previous research has shown that 6-8 MCTQ questionnaires suffice for a reliable overall judgement of a teacher.³² We therefore emailed an invitation to participate in the study to all the clinical teachers for whom we had received six or more MCTQ student ratings between March and November 2009 (76 teachers). In order to investigate the added value of a peer reflection meeting, we allocated the teachers to two feedback facilitation strategies (FFS1 and FFS2). The strategies were identical except for addition of a peer reflection meeting to FFS2. In order to ensure optimal comparability of the groups, we used matched random sampling, in which participants are first purposefully matched on specific characteristics, before being individually and randomly assigned to one of two groups.³³ The participants were matched on age, department, gender, participation in a faculty development programme and mean score on the MCTQ. Since this might result in an uneven number of teachers in a sub-group, it was possible that the final groups differed in size.

Feedback facilitation strategies

The steps of the feedback facilitation strategies are presented in Figure 1. The participants in both groups completed an online self-assessment questionnaire based on the MCTQ,^{31,32} after which they received an email with an individual feedback report, containing the student ratings (mean and standard deviation) collected between March and November 2009 as well as their own self-assessment ratings. Next, the FFS2 group, but not the FFS1 group, participated in a peer group reflection meeting, conducted in accordance with a modified version of the Critical Incident Method (Appendix 1).³⁴ After receiving instructions for this meeting by email, all FFS2 group members were assigned to one of five sessions (five teachers per meeting) held in January 2010. In composing these groups we took care to ensure diversity of departments. The peer reflection meetings lasted between 1.5 and 2 hours and were moderated by a psychologist who was an experienced facilitator.

TEACHER REFLECTION

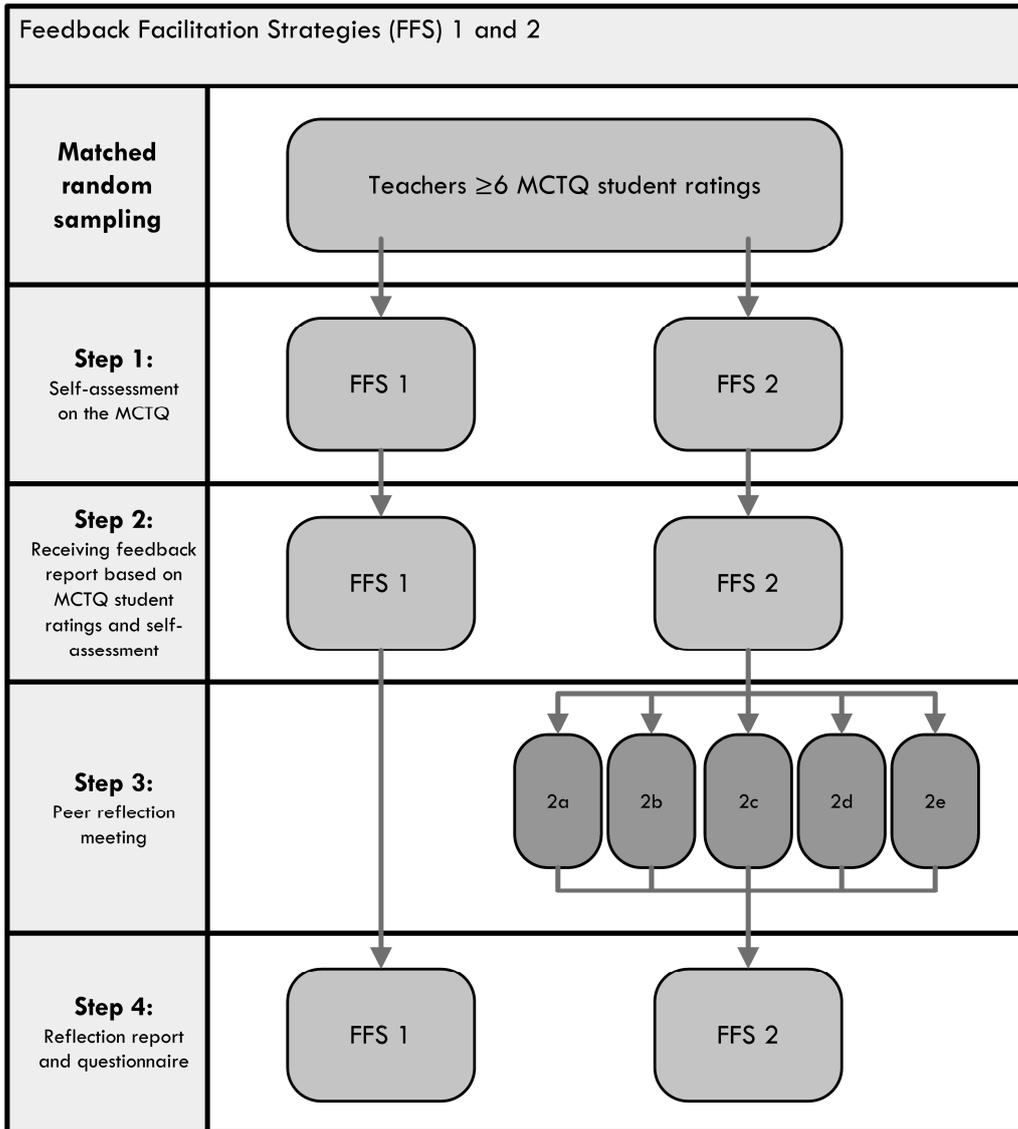


Figure 1: Workflow of the two Feedback Facilitation Strategies

All participants (FFS1 and FFS2) completed an online editable version of a reflection report with pre-structured questions based on Korthagen’s ALACT model.¹⁶ The questions are presented in Appendix 2. After submitting their report, the participants were automatically directed to an online questionnaire about the

feedback facilitation strategy in which they had participated. An open question asked them to describe their experiences. Additionally, they were asked to rate, on a five-point Likert scale (1=fully disagree; 5=fully agree), for each step of the strategy, their agreement with four statements: 1. This step stimulated me to reflect on my teaching practice; 2. This step stimulated me to formulate alternative methods of action for my teaching practice; 3. This step stimulated me to actually alter my methods of action concerning my teaching practice; 4. This is an effective step of the feedback facilitation strategy. The first author analysed the descriptions for recurring themes and coded them as positive or negative. He discussed the emerging themes with the second author. Using SPSS 16.0, we calculated mean scores and standard deviations for each step and performed an independent samples t-test to determine any significant differences between the two groups.

Analysis of the reflection reports

The first author analysed the reflection reports. The level of reflection was determined using a framework developed by Hatton & Smith.¹⁴ Based on a summary of this framework,³⁵ we formulated criteria to distinguish four levels of reflection:

1. 'Descriptive writing'

The clinical teacher is not reflective, he or she is merely reporting events with no attempt to provide reasons.

2. 'Descriptive reflection'

The clinical teacher provides reasons (often based on personal judgement) for events or actions, but only in a reportive way.

3. 'Dialogic reflection'

The clinical teacher demonstrates a form of discourse with him/herself, mulling over reasons for events or actions. The clinical teacher also explores alternative methods of action.

4. 'Critical reflection'

The clinical teacher takes account of the (socio-political) context in which the events took place and decisions were made.

The alternative methods of action (step 3 in Korthagen's ALACT model)¹⁶ were qualified as follows: 1) No alternative methods: The clinical teacher does not

TEACHER REFLECTION

provide alternatives for his/her methods of action; 2) Aspects in need of improvement: The clinical teacher does not provide alternatives, but indicates for which aspects improvement is needed; 3) Alternative methods: The clinical teacher provides alternatives, but no concrete plan of action; 4) Alternative methods and plan: The clinical teacher provides achievable alternatives and a plan of action for achieving them.

After the analysis by the first author, who selected quotes to illustrate the results, five reports were randomly selected from the reports of each group (23%) and independently analysed by the second author. The two authors met once to discuss their interpretations and judgements. The reflection reports were anonymised during this part of the study. Inter-judge agreement (IJA) was calculated as follows:³⁶ $IJA = (\text{number of agreements} / (\text{number of agreements} + \text{number of disagreements})) * 100$. With respect to reflection the researchers agreed on nine reports and disagreed on one (inter-judge agreement 90%). The analysis of the alternatives showed agreement on eight reports and disagreement on two (inter-judge agreement 80%). Using a completed reflection report as the unit of analysis, the highest levels of reflection and proposals for alternatives were recorded and used to make a comparison between the groups.

Ethical considerations

A digitally signed informed consent form was obtained from all participants. Participation was voluntary and participants could opt out of the study at any moment. Any information that could be traced to individual students was removed from the feedback reports. The names and scores of the participants were only known to the first author, but the moderator also knew the names of the participants in the peer reflection meetings. The participants had no information about the feedback reports of the other participants. Confidentiality was assured and the reflection reports were coded in such a way that it was impossible for the second researcher to link them to an individual teacher. This information was provided to all participants by email.

RESULTS

Of the 76 clinical teachers we invited, 54 agreed to participate. Of the participants, 29 and 25 were randomly assigned to FFS1 and FFS2, respectively. All participants completed steps 1 and 2 (Figure 1). Four participants of the FFS2 group did not attend the peer reflection meeting (step 3) due to time constraints, which left four sessions with four participants and one session with five participants. 23 completed reflection reports (79.3%) (step 4) were submitted by the FFS1 group and 19 by the FFS2 group (90.5%).

Qualitative analyses of the reflection reports: Levels of reflection

Not all teachers showed evidence of reflection in their reflection reports. The following excerpt from the reflection report of Teacher 12 of the FFS1 group is illustrative of level 1, 'descriptive writing':

...MY CLINICAL COMPETENCE IS JUDGED AS GOOD BY THE STUDENTS, ALTHOUGH THEY LACK THE EXPERIENCE TO RATE THIS THOROUGHLY [DESCRIPTIVE WRITING]. MOST STUDENTS ALSO THINK THAT MY COMMUNICATION SKILLS ARE FINE, JUST LIKE MY EXPLANATION SKILLS [DESCRIPTIVE WRITING]. ON THE OTHER HAND, THERE ARE ALSO STUDENTS WHO THINK THAT I PERFORM POORLY IN THESE DOMAINS [DESCRIPTIVE WRITING]. SOME STUDENTS SAY THAT I SHOULD GIVE THEM MORE RESPONSIBILITIES AND THAT I SHOULD PAY MORE ATTENTION TO TEACHING [DESCRIPTIVE WRITING]...

This reflection report was qualified as 'descriptive writing' because the teacher merely reported the key points from the feedback report in his/her own words without attempting to provide reasons for his/her teaching practice.

Most participants exceeded the level of mere description. A teacher from the FFS1 group, for example, described experiencing difficulties in giving concrete feedback to individual students after a training session in the clinic.

...I WANT TO IMPROVE THE QUALITY OF MY FEEDBACK [DESCRIPTIVE WRITING]. [...] PROVIDING STUDENTS WITH CONCRETE FEEDBACK AFTER ONE DAY'S WORK IS HARD [DESCRIPTIVE WRITING]. [...] MOST OF THE TIME I AM SO BUSY EXPLAINING HOW THINGS WORK THAT I DO NOT HAVE TIME TO OBSERVE THE OTHER STUDENTS. THEREFORE I DON'T REMEMBER WHAT INDIVIDUAL STUDENTS DID RIGHT OR WRONG [DESCRIPTIVE REFLECTION]. [...] I AM JUST TOO FOCUSED ON TEACHING AND EXPLAINING INSTEAD OF OBSERVING

TEACHER REFLECTION

[DESCRIPTIVE REFLECTION]. MOREOVER, I SEE THE STUDENTS FOR SUCH A SHORT PERIOD THAT I CAN HARDLY REMEMBER THEIR NAMES [DESCRIPTIVE REFLECTION]...

This report was interpreted as 'descriptive reflection' (level 2). The teacher reports the key point from the feedback report, but also tries to explain his/her teaching practice.

The following excerpt from a reflection report in which a teacher from the FFS2 group responds to a low feedback rating in relation to asking students about their learning goals is an example of level 4 (critical reflection). Not only is this teacher mulling over reasons for not asking about learning goals, he or she also takes account of the teaching context.

...I THINK THE PROBLEM IS THAT I DO NOT ASK INDIVIDUAL STUDENTS ABOUT THEIR SPECIFIC LEARNING GOALS [DESCRIPTIVE REFLECTION]. I DO TRY TO MAKE CLEAR WHICH ASPECTS NEED IMPROVEMENT BY ASKING THEM QUESTIONS. SUBSEQUENTLY I CONCENTRATE ON THESE ASPECTS DURING CLINICAL WORK. PRESUMABLY I WAS UNDER THE IMPRESSION THAT ALL STUDENTS HAVE THE SAME LEARNING GOALS DURING A SPECIFIC PART OF THEIR ROTATIONS [DIALOGIC REFLECTION]. HOWEVER, THE ESSENTIAL FACTOR IS THAT STUDENTS HAVE DIFFERENT NEEDS WITH RESPECT TO WHAT THEY WANT TO LEARN OR PRACTISE DURING MY ROTATION [CRITICAL REFLECTION]. THE FACT THAT I DO NOT ASK FOR INDIVIDUAL LEARNING GOALS IS PROBABLY FIRST OF ALL MY FAULT [DIALOGIC REFLECTION]. ON THE OTHER HAND, ASKING FOR LEARNING GOALS IS NOT CUSTOMARY IN MY DEPARTMENT, SO THAT IS PROBABLY THE REASON WHY I AM NOT USED TO DOING THAT [CRITICAL REFLECTION]...

Qualitative analyses of the reflection reports: Alternative methods of action

Most teachers from both groups formulated alternatives for their teaching practice, and only a few did not go beyond the level of merely mentioning aspects where improvement was needed. Two participants indicated being unable to give alternatives for their teaching practice. The following excerpt from a reflection report in which a participant of the FFS1 group responds to low ratings on communication skills illustrates this level of 'No alternative methods of action'.

...I RECEIVED LOW RATINGS ON MY COMMUNICATION SKILLS IN MY FEEDBACK REPORT. I REALLY DO NOT KNOW HOW TO IMPROVE THIS ASPECT. I JUST HAVE TOO LITTLE EXPERIENCE WITH TEACHING IN THE CLINICAL SETTING...

Level 3: 'The clinical teacher provides alternatives, but no concrete plan of action' was mainly demonstrated by participants who thought that negative feedback was due to clinic organisation. It is noticeable that these teachers did not propose concrete changes concerning their own behaviour.

...I WANT TO SUPERVISE A WHOLE WEEK INSTEAD OF ONE MORNING [ALTERNATIVE]. THE PROBLEM IS THAT THIS IS NOT FOR ME TO DECIDE...

...WE HAVE TO DETACH EDUCATION FROM CLINIC WORK [ALTERNATIVE]. THE ORGANISATION SHOULD THINK ABOUT THIS IDEA...

Level 4 'alternative methods and plan' is exemplified by a teacher from the FFS2 group, who wishes to act less like a schoolteacher and describes a concrete alternative method of action.

...I WANT TO IMPROVE THIS ASPECT BY MAKING STUDENTS TAKE THE CONSEQUENCES OF DOING NO PREPARATION (FOR CLINICAL WORK) [ALTERNATIVE]. IF AS A CONSEQUENCE A TOPIC CANNOT BE DISCUSSED I SHOULD NOT BLAME MYSELF [ALTERNATIVE]. BECAUSE I TEND TO REACT LIKE A SCHOOLTEACHER I AM GOING TO ASK BOTH STUDENTS AND COLLEAGUES TO GIVE ME FEEDBACK ON THIS. I WILL EXPLAIN THIS TO THE STUDENTS I SUPERVISE. THIS WAY I WILL BECOME MORE AWARE OF MY TEACHING HABITS [PLAN]...

Quantitative analyses of the reflection reports: Differences between the two feedback facilitation groups

Table 1 shows the quantitative results for the levels of reflection and alternative action plans of both groups. With respect to reflection, most participants scored above the lowest level of mere description and the highest level was attained by eight participants in the FFS1 group and 15 participants in the FFS2 group. As for alternative methods of action, 13 and 17 participants in the FFS1 and FFS2 group respectively formulated alternative methods of action to improve their teaching practice, while a concrete plan of action was described by eight participants in FFS1 and 13 participants in group FFS2.

TEACHER REFLECTION

Table 1: Numbers and percentages of participants in both groups according to the highest level of reflection and alternative methods of action they attained				
Hatton & Smith's levels of reflection	FFS*1 (N=23)		FFS2 (N=19)	
	N	%	N	%
1. Descriptive writing	2	8.7%	0	0%
2. Descriptive reflection	7	30.4%	3	15.8%
3. Dialogic reflection	6	26.1%	1	5.2%
4. Critical reflection	8	34.8%	15	79%
Alternative methods of action	FFS1 (N=23)		FFS2 (N=19)	
	N	%	N	%
1. No alternative methods of action	1	4.3%	1	5.3%
2. Aspects in need of improvement	9	39.1%	1	5.3%
3. Alternative methods	5	21.7%	4	21.0%
4. Alternative methods and plan	8	34.8%	13	68.4%
*FFS = Feedback Facilitation Strategy				

Perceived effectiveness of the two feedback facilitation strategies

All the participants who submitted a reflection report also submitted the completed effectiveness questionnaire. The means and standard deviations are shown in Table 2. Because an independent samples t-test did not indicate any significant difference between the two groups, we present aggregated scores, except for the question about the peer reflection meeting, which was put to group FFS2 only.

Table 2: Perceived effectiveness of the steps of the feedback facilitation strategies (FFS) (mean ratings and SDs on a Likert scale from 1 = fully disagree to 5 = fully agree)

Statements	Self-assessment (N=42)		Feedback report (N=42)		Peer meeting (N=19)		Reflection report (N=42)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1. Stimulated to reflect on practice	3.9	0.7	4.5	0.6	4.3	0.6	3.8	0.9
2. Stimulated the formulation of alternative methods	3.5	0.8	4.2	0.7	4.3	0.7	3.6	0.9
3. Stimulated alterations of teaching practice	3.3	0.9	4.0	1.0	4.1	1.0	3.7	1.0
4. This was an effective step of the FFS	4.1	0.7	4.5	0.6	4.2	0.9	3.6	0.9

Experiences with the two feedback facilitation strategies

The clinical teachers reported 119 positive experiences and 43 negative experiences. A summary is shown in Table 3. Most teachers indicated that the steps helped them to gain insight into the strengths and weaknesses of their teaching. The majority of the teachers from the FFS2 group stated that the peer reflection meeting helped them to translate the student feedback into concrete alternatives for their teaching practice.

TEACHER REFLECTION

Table 3: Positive and negative experiences concerning the steps of the feedback facilitation strategies reported by participants in both feedback facilitation strategies (FFS)

		FFS1 (N=23)	FFS2 (N=19)
Step 1: Self-assessment	Pos.	<ul style="list-style-type: none"> • Helped me to understand what is important for good workplace learning (4 comments) • Stimulated me to think about my teaching practice (2 comments) • Helped me to define the problems I encounter while teaching (2 comments) 	<ul style="list-style-type: none"> • Stimulated me to think about my teaching practice (6 comments) • Helped me to understand what is important for good workplace learning (4 comments)
	Neg.	<ul style="list-style-type: none"> • Did not stimulate reflection. For that you need feedback from students. (2 comments) 	<ul style="list-style-type: none"> • Did not stimulate reflection. For that you need feedback from students (3 comments) • This was the least effective step (1 comment)
Step 2: Feedback report	Pos.	<ul style="list-style-type: none"> • Offered a great opportunity to learn about positive and negative aspects of my clinical teaching (15 comments), especially because the student feedback was combined with self-assessment (6 comments) • Students usually do not give concrete verbal feedback. The information in the feedback report is far more valuable. (2 comments) 	<ul style="list-style-type: none"> • Offered a great opportunity to learn about positive and negative aspects of my clinical teaching (12 comments), especially because the student feedback was combined with self-assessment (3 comments) • Helped me to think about ways to alter my teaching practice (8 comments) • Students usually do not give concrete verbal feedback. The information in the feedback report is more informative. (3 comments) • It stimulated me to take time to think about my teaching task (2 comments)
	Neg.	<ul style="list-style-type: none"> • It would be better if feedback reports were provided more often and from more students (3 comments), • I question the reliability of this report (4 comments) • Just a report does not stimulate me to reflect. For that you need coaching. (3 comments) • It would be nice to see the (overall) results of the other teachers from one's department (1 comment) 	<ul style="list-style-type: none"> • Just reading the feedback report does not suffice. It was the peer reflection meeting that stimulated me to reflect. (2 comments) • I question the reliability of this report (1 comment) • Interpreting the feedback report was sometimes difficult. I would like more oral feedback. (1 comment)

Table 3 continued			
Step 3: Peer reflection meeting	Pos.		<ul style="list-style-type: none"> • It really helped to translate the student feedback into concrete alternatives for teaching practice (14 comments) • This meeting helped me to look at my teaching practice from a different perspective (10 comments) • Listening to and helping other participants really helped me to reflect (5 comments) • It was the step with the biggest impact (2 comments)
	Neg.		<ul style="list-style-type: none"> • Peer meetings are quite time consuming (2 comments) • My peers were facing the same problems, this did not help me to reflect (2 comments) • Only time can tell if the peer reflection meeting really helped me to improve my teaching practice (1 comment)
Step 4: Reflection report	Pos.	<ul style="list-style-type: none"> • Helped me to make the feedback more concrete and stimulated me to think about alternatives for my teaching practice (6 comments) • It was useful that the reflection report was pre-structured (2 comments) 	<ul style="list-style-type: none"> • Writing a reflection report helped me to make my reflection more concrete (8 comments) • It really helped me to alter my teaching practice (3 comments)
	Neg.	<ul style="list-style-type: none"> • The reflection report in itself did not stimulate me to reflect, that happened earlier (6 comments) • Writing the reflection report takes a lot of time (3 comments) • Coaching would be useful when writing a reflection report (2 comments) • Some of the pre-structured questions were difficult to understand (2 comments) 	<ul style="list-style-type: none"> • This step is more about recording your reflection. The actual reflection takes place during the peer reflection meeting. (2 comments) • Coaching would be useful when writing a reflection report (1 comment) • Some of the pre-structured questions were difficult to understand (1 comment)

*FFS = Feedback Facilitation Strategy

DISCUSSION

We explored the levels of reflection demonstrated by clinical teachers after receiving feedback in a facilitated manner and examined if participation in peer group reflection made a difference. We also examined the participants' evaluation of the two feedback facilitation strategies.

Most of the clinical teachers reflected on the dialogic or even the critical level. The reflection reports showed that teachers analysed the student feedback from different perspectives and took account of the context of the teaching. According to Hatton & Smith, this is beneficial for professional growth.¹⁴ The findings from the reflection reports are in line with earlier studies by Pee et al.³⁵ and Tigelaar et al.,¹⁷ who also reported on activities to stimulate reflection.

An innovative aspect of our study is that we also examined whether teachers formulated concrete alternatives and plans for teaching practice. This was done by the majority of the teachers but more markedly by the teachers who participated in the peer reflection meeting. Formulating alternatives for practice is an essential aspect of reflection, because concrete plans of action have been shown to be related to actual improvement of teaching, which should be the ultimate goal of providing teachers with student feedback.¹⁵

The evaluation of the feedback facilitation strategies by the teachers underlines the results from the reflection reports. According to the teachers, all steps of the feedback facilitation strategies were effective. The teachers who participated in the peer reflection meeting indicated that this step was very useful for translating the student feedback into concrete alternatives and plans for their teaching practice and helped them to take different perspectives. Comparison between the two feedback facilitation groups demonstrated that the peer reflection meeting had a positive impact on engaging in dialogic and critical reflection and also led to more alternatives and concrete action plans. This suggests that peer discussion can further enhance the reflection process. We hypothesise that learning about their peers' alternative teaching methods can help teachers to think of practical applications of this knowledge in their own teaching practice. This hypothesis is supported by the findings of Meirink, Meijer and Verloop (2007)³⁷ in a study of teachers' individual learning in a collaborative setting.

In conclusion, a feedback facilitation strategy comprising self-assessment, a feedback report and a pre-structured reflection report promoted dialogic and

critical reflection. Adding peer group reflection resulted in deeper critical reflection and more concrete plans for change. These findings are in line with other studies^{23,26,28,30,38} which showed that a limited effect of individual self-reflection can be enhanced by peer meetings. However, no previous studies have compared strategies that differed only in the inclusion of a peer meeting. What our study adds to the literature is some evidence that peer group reflection can offer added value.

This study has some limitations. Any comparison between groups carries the potential risk of sampling bias, and we cannot rule out the possibility that the teachers attending the peer reflection meeting were more competent with respect to reflection than the comparison group. This seems rather unlikely, however, since we used matched random sampling,³³ which ensured that the teachers assigned to each strategy were comparable with respect to age, department, gender, participation in a faculty development programme and mean score on the MCTQ. Another limitation is the absence of a control group which wrote a reflection report after receiving student feedback without feedback facilitation. A future study should also investigate the results of such a group with regard to critical reflection and formulation of plans for improvement.

Although we know from the literature that reflection is crucial for enhancing teaching practice,¹⁴⁻¹⁸ the results of our study provide no evidence to ascertain whether the teachers did actually change their teaching practices. This remains to be examined in further studies. Are proposed changes actually implemented? On what level do these changes occur? What is the durability of these changes in teaching behaviour and do student ratings/evaluations improve as a result? Nevertheless, the results provide evidence that a feedback facilitation strategy can be effective, especially if it includes a peer reflection meeting.

REFERENCES

1. Ramsey PG, Gillmore GM, Irby DM. Evaluating clinical teaching in the medicine clerkship: Relationship of instructor experience and training setting to ratings of teaching effectiveness. *J Gen Intern Med* 1988;3:351-5.
2. Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. *Acad Med* 1998;73:324-32.
3. Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 2000;75:161-6.
4. Hamdy H, Williams R, Tekian A. Application of "VITALS": Visual indicators of teaching and learning success in reporting student evaluations of clinical teachers. *Educ Health* 2001;14:267-76.
5. Dolmans DHJM, Wolfhagen HAP, Gerver WJ, de Grave W, Scherpbier AJJA. Providing physicians with feedback on how they supervise students during patient contacts. *Med Teach* 2004;26:409-14.
6. McGrath C, Wai Kit Yeung R, Comfort MB. Development and evaluation of a questionnaire to evaluate clinical dental teachers (ECDT). *Br Dent J* 2005;198:45-8.
7. Moore S, Kuol N. Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education* 2005;10:57-73.
8. Cashin WE. Student ratings of professors: Uses and misuses. In: Seldin P, Hutchings P. *Changing practices in evaluating teaching: a practical guide to improved faculty performance and promotion/tenure decisions*. 1st ed. San Francisco, CA: Jossey-Bass; 1999. 25-44.
9. Berk RA. Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education* 2005;17:48-62.
10. Ramani S. Twelve tips to promote excellence in medical teaching. *Med Teach* 2006;28:19-23.
11. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
12. Marsh HW, Roche LA. Making students' evaluations of teaching effectiveness effective. *Am Psychol* 1997;52:1187-97.
13. Morrison J. ABC of learning and teaching in medicine: Evaluation. *Brit Med J* 2003;326:385-7.
14. Hatton N, Smith D. Reflection in teacher education: Towards definition and implementation. *Teach Teach Educ* 1995;11:33-49.
15. Ward JR, McCotter SS. Reflection as a visible outcome for preservice teachers. *Teach Teach Educ* 2004;20:243-57.
16. Korthagen FAJ, Vasalos A. Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice* 2005;11:47-71.
17. Tigelaar DEH, Dolmans DHJM, de Grave WS, Wolfhagen IHAP, van der Vleuten CPM. Portfolio as a tool to stimulate teachers' reflections. *Med Teach* 2006;28:277-82.
18. Watts M, Lawson M. Using a meta-analysis activity to make critical reflection explicit in teacher education. *Teach Teach Educ* 2009;25:609-16.

19. Schön DA. *Educating the reflective practitioner: Towards a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass; 1987.
20. Atwater LE, Waldman DA, Brett JF. Understanding and optimizing multisource feedback. *Hum Resour Manage* 2002;41:193-208.
21. Seifert CF, Yukl G, McDonald RA. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *J Appl Psychol* 2003;88:561-9.
22. Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Person Psychol* 2005;58:33-66.
23. DeNisi AS, Kluger AN. Feedback effectiveness: Can 360-degree appraisals be improved? *The Acad Manage Exec* 2000;14:129-39.
24. Sargeant JM, Mann KV, Sinclair D, Van der Vleuten CPM, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv in Health Sci Educ* 2008;13:275-88.
25. Overeem K, Wollersheim H, Driessen E, Lombarts K, van de Ven G, Grol R, et al. Doctors' perceptions of why 360-degree feedback does (not) work: A qualitative study. *Med Educ* 2009;43:874-82.
26. Tigelaar DEH, Dolmans DHJM, Meijer PC, de Grave WS, van der Vleuten CPM. Teachers' interactions and their collaborative reflection processes during peer meetings. *Adv in Health Sci Educ* 2008;13:289-308.
27. Ross JA, Bruce CD. Teacher self-assessment: A mechanism for facilitating professional growth. *Teach Teach Educ* 2007;23:146-59.
28. Schaub-de Jong MA, Cohen-Schotanus J, Dekker H, Verkerk M. The role of peer meetings for professional development in health science education: A qualitative analysis of reflective essays. *Adv in Health Sci Educ* 2009;14:503-13.
29. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Peters WG, van Coppenolle L, Scherpbier AJJA. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv in Health Sci Educ* 2009;15:315-28.
30. Mcleod PJ, Steinert Y. Peer coaching as an approach to faculty development. *Med Teach* 2009;31:1043-4.
31. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:e270-7.
32. Boerboom TBB, Dolmans DHJM, Jaarsma ADC, Muijtjens AMM, van Beukelen P, Scherpbier AJJA. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. *Med Teach* 2011;33:e84-91.
33. Brown KW, Cozby PC, Kee DW, Worden PE. *Research methods in human development*. Mayfield: Mountain View; 1999.
34. Hendriksen J. *Begeleid intervisiemodel, collegiale advisering en probleemoplossing*. [guided peer model, advising colleagues and problem solving]. Baarn, NL: Nelissen; 1997.
35. Pee B, Woodman T, Fry H, Davenport ES. Appraising and assessing reflection in students' writing on a structured worksheet. *Med Educ* 2002;36:575-85.

TEACHER REFLECTION

36. Miles MB, Huberman AM. Qualitative data analysis: A sourcebook of new methods. Beverly Hills, Ca: Sage; 1984.
37. Meirink JA, Meijer PC, Verloop N. A closer look at teachers' individual learning in collaborative settings. *Teachers and Teaching: Theory and Practice* 2007;13:145-64.
38. Steinert Y, Mann KV. Faculty development: Principles and practices. *J Vet Med Educ* 2006;33:317-24.

APPENDICES

1. The modified Critical Incident Method used in the peer reflection meetings

1. The moderator interviews one participant on his personal learning question while the other participants listen. The goal of this phase is to explore all aspects of the learning question. Example questions are: Why did the interviewee select this question? Why is it a problem for the interviewee? What is the context of this problem? What has the interviewee already done to solve this problem?
2. The other participants then reflect on the interviewee's personal learning question. Emphasis in this phase is on participants engaging in brainstorming to generate solutions for the interviewee's problem, using their own experiences. The interviewee is encouraged to listen carefully and not to intervene in the discussion.
3. The moderator then interviews the interviewee to elucidate which reflection/solution, mentioned during phase 2, appeals the most. The interviewee is also asked to reflect on how he or she is going to implement this solution in his/her teaching practice.

2. The questions, based on Korthagen's ALACT-model, used in the pre-structured reflection reports:**Looking back on action:**

1. Which aspects of your clinical teaching practice are evaluated positively in your feedback report?
2. Which aspects of your clinical teaching practice are evaluated negatively in your feedback report?
3. On which aspects of your clinical teaching practice do you want to reflect in this report? Why did you choose these aspects?

Awareness of essential aspects:

4. What do these aspects (see question 3) mean for you as a clinical teacher?
5. Which essential factors have an influence on these aspects?
6. How do these factors influence these aspects?

Creating alternative methods of action:

7. What are alternative methods of action regarding these aspects?
8. How are you going to implement these alternative methods of action?

TEACHER REFLECTION

6.

The effect of facilitated student feedback on ratings of clinical teaching effectiveness

T.B.B. Boerboom

D.H.J.M. Dolmans

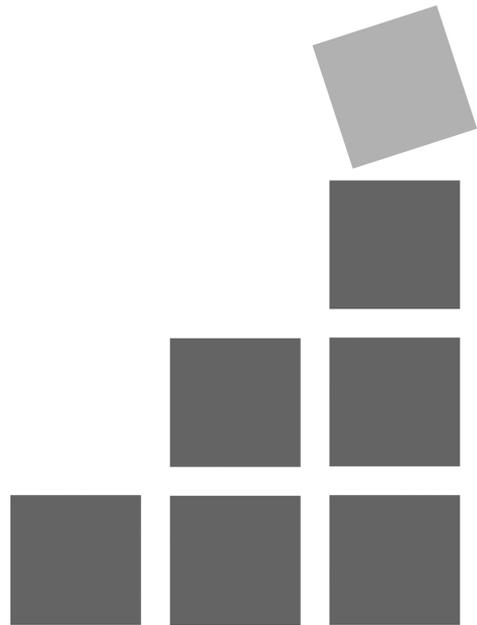
A.M.M. Muijtjens

P. van Beukelen

A.D.C. Jaarsma

A.J.J.A. Scherpbier

2011 Under review



ABSTRACT

Context

Student feedback plays an important role in faculty development. Many publications have reported on best practices for validation of feedback instruments and facilitation of feedback. There is little evidence, however, that such initiatives actually lead to more effective clinical teaching or higher student ratings. We investigated the effects of facilitated feedback and the additional effect of peer reflection on the teaching performance of clinical teachers.

Methods

In a pre- post study design student evaluations of veterinary clinical teachers were obtained with a validated instrument (the Maastricht Clinical Teaching Questionnaire (MCTQ)) before and after clinical teachers participated in one of two interventions. Both interventions comprised self-assessment, student ratings and a pre-structured reflection report while one also included peer reflection. A control group received no intervention. Regression analysis was performed to examine the differences between the three groups.

Results

The two intervention groups and the control group consisted of 29, 25 and 18 teachers, respectively. Following feedback facilitation there were significant positive effects on overall teaching effectiveness and learning climate, while after peer reflection there was an additional effect for items relating to students' active engagement in learning.

Conclusions

Facilitated feedback has a positive effect on clinical teaching, while peer reflection showed a small additional effect on MCTQ scores concerning students' active engagement in learning. Further research should investigate how feedback facilitating strategies can improve clinical teaching in all the MCTQ domains and whether positive effects can be sustained for a prolonged period of time.

INTRODUCTION

There is widespread agreement that clinical rotations are a crucial component of the undergraduate medical¹⁻⁴ and veterinary curriculum.⁵ Clinical clerkships in both curricula show strong similarities.⁶ Students participate in clinical practice where they are supervised by expert physicians⁷ and learn to integrate knowledge with complex professional skills, including communication and collaboration skills.^{3,8}

Although the goal of clinical training is clear, the road towards the goal remains something of a mystery. Clerkships are like a black box^{5,7} with students entering on one side and junior (veterinary) doctors emerging on the other side. Many authors have tried to pinch holes in the box to take a closer look at the factors and processes that enable this transformation. They have focused on factors like assessment, motivation and learning climate but above all on the role of the clinical teacher.^{1,2,5,7,9-13}

Several authors have stated that clinical teachers play a key role in the transformation from student to doctor.^{2,8,14,15} Clinical teachers assess students' performance, act as role models, supervise and coach students and help them to reflect on their actions.^{2,14,16} This central role has prompted a great deal of interest in instruments to evaluate clinical teachers' supervisory skills.¹⁷

Of the numerous instruments developed in recent years to examine clinical teaching effectiveness,^{18,19} quite a few were constructed and validated to collect feedback for clinical teachers based on student ratings and intended for use in faculty development.¹⁷ Beckman and Fluit examined the quality of a range of these instruments.^{18,19} While Beckman focused mainly on psychometric qualities,^{18,20} Fluit concentrated on the content validity of the instruments and their usability.¹⁹ A study by Stalmeijer et al. focused on theoretical dimensions to underpin the construction of a student evaluation instrument.²¹ Based on the theory of cognitive apprenticeship,²² these authors developed and validated the Maastricht Clinical Teaching Questionnaire,¹⁶ an instrument to evaluate the performances of individual clinical teachers.

If feedback instruments are to supply data that can be used for faculty development purposes, they should be valid and reliable.²⁰ It should be noted, however, that using a validated instrument does not automatically guarantee that feedback is accepted by teachers let alone that it promotes learning and leads to measurable improvement of teaching.²³ For teachers to be able to effectively use

EFFECT OF FACILITATED FEEDBACK

feedback from student evaluations to adapt and improve their teaching strategies, reflection on the feedback is indispensable.^{24,25} In an earlier study we demonstrated that teachers receiving feedback from student MCTQ ratings reflected more deeply when the feedback was discussed in a peer reflection meeting.²⁶ Other authors have also reported that peer reflection can enhance teachers' reflective skills.²⁷⁻²⁹ Although feedback facilitation can foster deeper reflection and initiatives to improve teaching behaviour,²⁶ there is scant evidence that such interventions do indeed result in more effective clinical teaching from the students' perspective.

In the current paper we report a study designed to answer the following research questions:

1. Do student ratings of clinical teachers' performance improve after a feedback strategy consisting of self-assessment, a feedback report containing self-assessment and student ratings, and a written self-reflection assignment?
2. Is further improvement of student ratings achieved when teachers also discuss their ratings in a peer reflection group?

METHODS

Study design

We used a pre-post study design with a control group and two intervention groups. One intervention group received student ratings, self-assessment ratings and a written self-reflection assignment and the other group received the same intervention with an additional peer reflection session after receiving the feedback report and before the self-reflection assignment.

Context

The study was conducted at the Faculty of Veterinary Medicine, Utrecht University (FVMU), the Netherlands. The curriculum offers four years of preclinical training followed by two years of clinical clerkships. Clinical rotations are relatively short, varying from one to six weeks, because rotations in the same discipline are repeated for different animal species. Clinical teaching is provided by four clinical departments: the Equine Health Department, the Companion Animal Health Department, the Farm Animal Health Department and the Pathology Department.

Instrument

Both the pre- and post-measurements were performed using the MCTQ,²¹ which has been validated for short veterinary rotations.⁶ The MCTQ contains fifteen items (statements, five-point Likert scale) inviting ratings of individual teachers' performance in the five domains of the cognitive apprenticeship model³⁰ (Modelling, Coaching, Articulation and Exploration) and General Learning Climate (Climate) (Table 1) and one item inviting an overall rating (Overall effectiveness) on a ten-point scale (1=very poor, 10=excellent) of the teacher's performance. An earlier reliability study showed that six to eight student ratings are required for a reliable evaluation of one individual teacher.⁶ The development and validation of the MCTQ are reported elsewhere.^{6,21}

Pre-measurement

For the pre-measurement, all students undertaking veterinary clinical rotations between March and November 2009 were asked to complete the MCTQ online as soon as possible after each student-teacher encounter.

Intervention

The intervention consisted of a feedback strategy (FBK) and a peer reflection meeting (RFL). Intervention group 1 participated only in the FBK strategy, consisting of a self-assessment, a feedback report containing the teacher's self-assessment ratings and student ratings on the MCTQ, followed by a pre-structured written reflection assignment. For this assignment the teachers were asked to write a reflection report on their teaching effectiveness based on the feedback report using a pre-structured form designed to facilitate reflection.

Intervention group 2 received the same intervention but additionally participated in a peer reflection meeting in which the feedback report was discussed (RFL strategy). The meeting took place before the teachers wrote the reflection report. Access to the online pre-structured reflection form was only granted after the teachers had attended the peer reflection meeting. The meeting was structured according to the Critical Incident Method³¹ and moderated by an experienced facilitator. An extensive description of the intervention can be found elsewhere.²⁶

EFFECT OF FACILITATED FEEDBACK

The third group was a control group, which received no feedback. The pre- and post-measurements were completed for all the participating teachers. In sum, three groups of teachers participated in the study. Intervention groups 1 and 2 were exposed to FBK only and FBK plus RFL, respectively.

We sent invitations to participate in the study to teachers who met the following inclusion criteria: (1) completed MCTQs from at least six students to ensure reliable ratings and (2) the teachers were expected to hold their current post at FVMU for more than one year after the pre-measurement to ensure participation in the post-measurement. Of the invited teachers (N=76), 54 consented to participate. In order to optimize comparability of the groups we purposely composed subgroups of participants matched on specific characteristics (age, department, gender, participation in a faculty development programme and mean score on the MCTQ pre-measurement) before randomly assigning the teachers to one of the intervention groups. Twenty-nine teachers were assigned to group 1 and 25 to group 2. Teachers who were willing to participate but unable to do so because of time constraints were included in the control group (N=18). Four of the invited teachers refused to participate. Comparability of the control group with the two treatment groups based on the matching criteria mentioned above was found to be satisfactory. The interventions were carried out between November 2009 and February 2010.

Post-measurement

After the intervention we again distributed an online version of the MCTQ to all clinical clerks in years 5 and 6. The procedure was identical to that of the pre-measurement. The students were not informed about the different study groups. They could evaluate every clinical teacher of the participating departments. We sent several reminders to stimulate response. Data collection continued until February 2011. Post-measurement data were analyzed for the teachers for whom six or more MCTQ ratings were received after the feedback intervention.

Analysis

For each participating teacher the pre- and post-measurement values for Overall effectiveness and the domains Climate, Modelling, Coaching, Articulation and Exploration were obtained by calculating the mean values for the fifteen items and

the five scales of the MCTQ. For these variables we estimated the effects of the FBK and RFL components of the intervention using linear regression according to

$$postmeas = b_0 + b_1 \cdot premeas + b_2 \cdot FBK + b_3 \cdot RFL + error$$

where *postmeas* and *premeas* are the post- and pre-measurements of a variable, *FBK* and *RFL* are dummy variables indicating the intervention to which the teacher had been exposed (intervention 1: *FBK*=1, *RFL*=0; intervention 2: *FBK*=1, *RFL*=1; control: *FBK*=0, *RFL*=0), b_0 is the intercept, b_1 , b_2 , and b_3 represent the effects of the pre-measurement, FBK and RFL, respectively, and error is the residual. The expression represents an analysis of covariance model with post-measurement as the dependent variable, pre-measurement as the covariate and the indicator variables FBK and RFL as the effects of interest. The statistical package SPSS 17.0 was used to perform the analyses.

Ethics

Participation was voluntary for both students and teachers. Confidentiality was ensured and a digitally signed informed consent form was obtained from all teachers participating in the intervention groups. All clinical teachers of the participating departments were informed about the MCTQ questionnaire by the head of their department. The MCTQ questionnaires were filled out anonymously by students and the ratings were only known to the first author. The teachers and the students were kept informed about the research by means of a website and newsletters.

RESULTS

A total of 42 clinical teachers in intervention group 1 (23, 80%) and 2 (19, 76%) completed the interventions. Data for those who did not complete the intervention were not included in the analysis. The post-measurement yielded data for fourteen teachers in intervention group 1 (61%), twelve teachers in intervention group 2 (63%) and fifteen teachers in the control group.

Table 1 shows descriptive data of the pre-measurement and post-measurement for Overall effectiveness and the item and scale scores of the five MCTQ domains. The results of the regression analyses are shown in the last two columns of Table 1 (effects of FBK and RFL) and in Figure 1. For Overall effectiveness and the MCTQ scales Climate, Modelling, Coaching, Articulation and Exploration the figure shows scatter plots and regression lines for post- versus pre-measurement.

The descriptive data in Table 1 show that the Overall effectiveness scores of intervention groups 1 and 2 are 7.26 (SD=1.08) and 7.56 (SD=0.90) respectively at pre-measurement and 7.43 (SD=1.03) and 7.65 (SD=0.76), respectively at post-measurement. Overall effectiveness declined in the control group from 7.23 (SD=1.05) to 6.93 (SD=0.86). As can be seen in Table 1, the majority of the differences between post- and pre-measurement were small but positive for the intervention groups and generally negative for the control group.

The regression results show that the net effects of the intervention components FBK and RFL were generally positive. Compared to the control group, there were significant positive effects of the standard feedback strategy (FBK) for Overall effectiveness (standardized regression coefficient, $\beta=0.25$) and Climate ($\beta=0.17$). Almost significant effects ($p=0.07, 0.06$; $\beta=0.19, 0.20$) were found for the domain scores on Modelling and Coaching. On the item level, significant positive effects of FBK were found for item 6 'The clinical teacher was a role model for me' ($b=0.30, p<0.05, \beta=0.23$) and item 7 'The clinical teacher provided me with constructive and concrete feedback during or following direct observation' ($b=0.38, p<0.05, \beta=0.34$). Using the standard regression coefficient as an indicator of effect size, according to Cohen's classification, the coefficients represent small to medium, and medium effect sizes (0.1, 0.3, 0.5 corresponding to small, medium and large effect).³²

For RFL, no significant positive effects were found on the scale level, except for an almost significant effect ($b=0.20, p=0.07, \beta=0.24$) for Exploration, which was mainly due to significant positive effects for items 14 'The clinical teacher stimulated me to become aware of my own strengths and weaknesses' and 15 'The clinical teacher stimulated me to think about how to improve my own strengths and weaknesses'.

Table 1: Descriptives (mean and SD) of the pre and post-test and the results of the regression analysis for the 5 domains (scales) and the overall effectiveness of the MCTQ

Descriptives							Regression coefficients	
Domain and Items	Intervention 1 (FBK) N=14		Intervention 2 (FBK+RFL) N=12		Control N=15		Effect of standard strategy (FBK) b_2	Additional effect refl. meeting (RFL) b_3
	Pre	Post	Pre	Post	Pre	Post		
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)		
Climate: The clinical teacher...								
1. established an environment where I felt free to ask questions or make comments.	4.15 (.74)	4.32 (.58)	4.19 (.76)	4.27 (.62)	3.88 (.80)	3.92 (.72)	0.21	-0.06
2. showed an interest in me as a student.	3.82 (.78)	4.00 (.67)	4.08 (.57)	3.97 (.60)	3.77 (.64)	3.70 (.63)	0.26*	-0.24
3. treated me with respect.	4.07 (.60)	4.16 (.48)	4.17 (.49)	4.18 (.63)	3.92 (.61)	3.88 (.58)	0.17*	-0.06
Scale score	4.01 (.69)	4.16 (.57)	4.14 (.59)	4.14 (.54)	3.86 (.67)	3.84 (.63)	0.20*	-0.12
Modelling: The clinical teacher...								
4. demonstrated how different skills should be performed.	3.71 (.41)	3.78 (.57)	3.62 (.62)	3.93 (.58)	4.02 (.35)	3.88 (.69)	0.10	0.21
5. explained, while performing a task, which aspects were important and why.	3.99 (.60)	4.06 (.67)	4.13 (.54)	4.26 (.28)	4.15 (.42)	3.99 (.48)	0.18	0.10
6. was a role model for me.	3.54 (.78)	3.66 (.71)	3.68 (.80)	3.82 (.49)	3.57 (.72)	3.38 (.65)	0.30*	0.08
Scale score	3.75 (.53)	3.83 (.60)	3.81 (.61)	4.00 (.39)	3.91 (.45)	3.75 (.54)	0.20 ^b	0.12

EFFECT OF FACILITATED FEEDBACK

Table 1 continued								
Coaching: The clinical teacher...								
7. provided me with constructive and concrete feedback during or following direct observation.	3.49 (.66)	3.78 (.61)	3.47 (.61)	3.83 (.53)	3.74 (.46)	3.57 (.49)	0.38*	0.07
8. allowed me to perform tasks independently.	4.11 (.53)	4.04 (.59)	4.13 (.55)	4.26 (.32)	4.13 (.53)	3.98 (.50)	0.07	0.21
9. was supportive when I experienced difficulties with a task.	3.70 (.59)	3.63 (.59)	3.76 (.44)	3.89 (.37)	3.86 (.43)	3.63 (.60)	0.11	0.22
Scale score	3.77 (.50)	3.82 (.56)	3.78 (.47)	4.00 (.33)	3.91 (.42)	3.72 (.49)	0.20 ^a	0.17
Articulation: The clinical teacher...								
10. asked me to explain my reasoning and actions.	3.82 (.51)	3.58 (.52)	3.88 (.49)	4.05 (.47)	3.88 (.37)	3.76 (.39)	-0.16	0.44*
11. alerted me to gaps in my knowledge and skills.	3.74 (.55)	3.78 (.59)	4.02 (.39)	4.05 (.37)	4.02 (.29)	3.79 (.39)	0.16	0.10
12. asked questions to increase my knowledge and understanding.	3.99 (.57)	4.06 (.57)	4.26 (.33)	4.23 (.31)	4.23 (.34)	3.89 (.50)	0.35	-0.03
Scale score	3.85 (.53)	3.80 (.52)	4.05 (.35)	4.11 (.37)	4.04 (.28)	3.81 (.39)	0.13	0.16
Exploration: The clinical teacher...								
13. stimulated me to formulate my own goals.	2.88 (.34)	3.12 (.39)	2.86 (.38)	3.17 (.41)	2.84 (.38)	2.91 (.46)	0.19	0.06
14. stimulated me to become aware of my own strengths and weaknesses.	3.49 (.50)	3.27 (.42)	3.59 (.44)	3.58 (.39)	3.57 (.41)	3.30 (.42)	0.02	0.25*
15. stimulated me to think about how to improve my own strengths and weaknesses.	3.41 (.42)	3.16 (.43)	3.42 (.46)	3.44 (.33)	3.46 (.47)	3.16 (.41)	0.02	0.28*
Scale score	3.26 (.39)	3.18 (.38)	3.29 (.37)	3.40 (.34)	3.29 (.38)	3.13 (.40)	0.07	0.20 ^b
Overall effectiveness (1-10)	7.26 (1.08)	7.43 (1.03)	7.56 (.90)	7.65 (.76)	7.23 (1.05)	6.93 (.86)	0.48*	-0.00
* p<0.05, a p=0.06, b p=0.07								

A significant positive effect was also found for item 10 ‘The clinical teacher asked me to explain my reasoning and actions’ in the Articulation domain. The results of the regression analyses on scale level are presented in Figure 1, where the differences in height between the regression lines indicate the effects of FBK (intervention group 1 versus control: broken line versus solid line) and the additional effect of RFL (intervention group 2 versus intervention group 1: dotted line versus broken line).

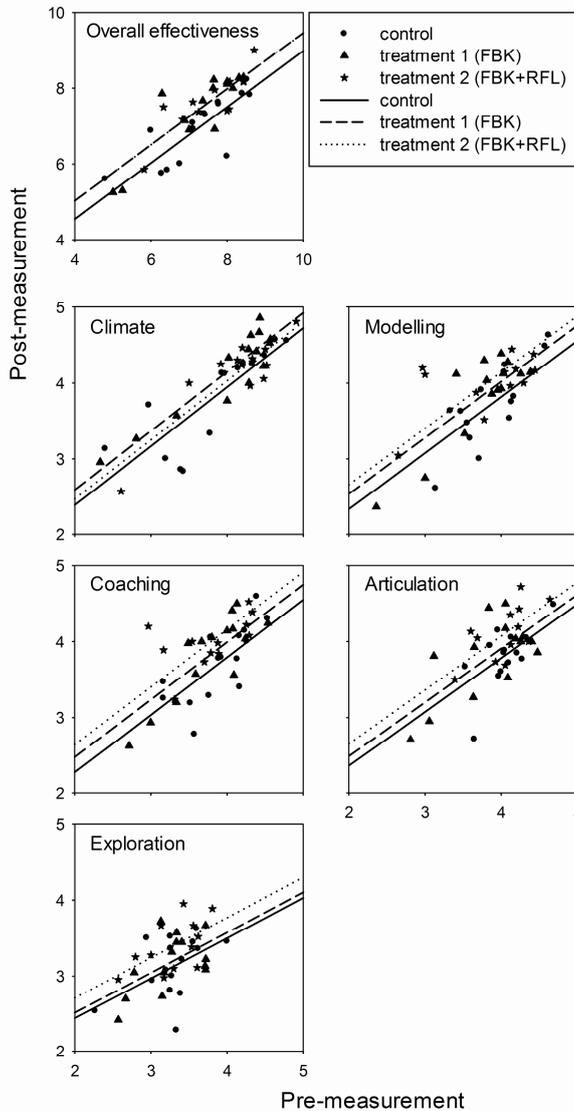


Figure 1: Scatter plots of post- versus pre-measurement of student ratings consisting of an Overall effectiveness score (10 point scale) and scores on the items of the MCTQ (five-point scales). The latter were summarized into five scale scores (Climate, Modelling, Coaching, Articulation, and Exploration). Regression lines are indicated for the control group, and for the two treatment groups: feedback (FBK) and feedback plus reflection (FBK+RFL). The differences in height indicate the effect of feedback (broken versus solid) and the additional effect of reflection (dotted versus broken).

DISCUSSION

This study explored whether student ratings of clinical teachers' performances improved after facilitated feedback (FBK) and whether there was a further improvement after a peer reflection meeting (RFL). The results show that the feedback strategy comprising self-assessment, feedback report with self-assessment ratings and student ratings and a pre-structured reflection assignment (FBK) had a positive significant effect on Overall effectiveness and Climate. There were also positive significant effects for items of the Modelling and Coaching scales. These findings are particularly of importance, because the behaviours encompassed by these domains: providing a positive learning climate,^{12,33} acting as a role model^{14,34,35} and providing students with opportunities to practice and the provision of rich feedback,^{8,36} are all reported to be highly beneficial for teaching effectiveness in the clinical context.

The peer reflection meeting (RFL) showed no significant effects in addition to those of the standard feedback strategy (FBK) on the domain level or for Overall effectiveness. There were, however, significant positive effects for three items of the Articulation and Exploration domains: asking students to explain their reasoning and skills and stimulating students to become aware of and think about improvement of their strengths and weaknesses. These items relate to teaching strategies that require active student participation. The domain of Articulation focuses on probing students for knowledge and understanding, while Exploration is the domain where students are encouraged to reflect on their performance in the clinical context.^{6,21} Success of these teaching strategies thus depends not exclusively on what the teacher does, but also on students' active participation, learning styles and reflective skills. These characteristics are known to vary between students.³⁷⁻³⁹ The results of the current study indicate that participating in a peer reflection meeting helps teachers to stimulate students to become actively involved in their learning during clinical rotations. This beneficial effect of a peer reflection meeting is in line with other studies exploring the advantages of this type of interventions.^{27,40} It may be explained by the fact that teachers who participated in the peer reflection meeting discussed and shared tips and tricks on how to stimulate active student participation as part of the teaching strategies for Articulation and Exploration.

The findings of this study suggest that providing veterinary clinical teachers with MCTQ feedback in a facilitated manner can lead to improved teaching effectiveness. Other authors have also stated that facilitated feedback is essential

for the learning of professionals.^{23,41,42} When teachers learn from feedback and critically reflect upon their practices, they are more likely to be able to achieve real improvement in their teaching.⁴³ This is the very result that feedback facilitation is expected to achieve.

This study has some strengths and limitations. A strong asset of the study is that effects were measured by student evaluations obtained with a validated instrument⁶ and did not rely exclusively on teachers' self-perceptions. It is known that professionals tend to rate their own performances too low or too high.⁴⁴ In this study, students observed teachers in practice and rated their performance on the MCTQ both before and after the intervention. At first sight, the differences between the pre- and post-measurements appear to be relatively small, albeit significant, but in the regression analysis where the control group score was used as reference the effects of the interventions were of small to medium and medium size. These results can be considered favourable, particularly when we consider that pre-measurement scores were already high and that only one intervention was performed in a one-year period. Longitudinal studies with larger samples are recommended to replicate the findings of this study.

The control group was composed of teachers who reacted positively to the invitation to participate but were not able to do so due to time constraints. Although the control group's pre-test scores did not differ dramatically from those of the treatment group, the possibility cannot be ruled out that the composition of this group affected the results due to sampling bias. Maybe the teachers in the control group were less interested in clinical teaching. The positive results of the intervention groups can be partly explained by the decline in teaching performance in the control group. These results suggest that performance in the control group decreased over time, while the interventions prevented such a decrease and turned it into a small, but significant, improvement. Marsh stated that a decline in teaching effectiveness can be partly due to an absence of faculty development activities.⁴¹

This study provides substantial indications for a beneficial effect of facilitated feedback on teaching performance even with a relatively small sample of 29, 25 and 18 teachers in the two experimental groups and the control group, which sounds promising. Nevertheless, further research is needed with larger samples of teachers. Furthermore, it still remains unclear how feedback facilitation aids teachers in improving their teaching in the different domains of the MCTQ. More qualitative research should explore this question. Another aspect that requires

EFFECT OF FACILITATED FEEDBACK

further study is maintenance of performance improvement over time. Longitudinal studies with more participating clinical teachers are recommended to shed light on this issue.

REFERENCES

1. Spencer J. ABC of learning and teaching in medicine: Learning and teaching in the clinical environment. *Brit Med J* 2003;326:591-4.
2. Deketelaere A, Kelchtermans G, Struyf E. Disentangling clinical learning experiences: An exploratory study on the dynamic tensions in internship. *Med Educ* 2006;40:908-15.
3. Leinster S. Learning in the clinical environment. *Med Teach* 2009;31:79-81.
4. Pratt DD, Harris P, Collins JB. The power of one: Looking beyond the teacher in clinical instruction. *Med Teach* 2009;31:133-7.
5. Lane IF, Strand E. Clinical veterinary education: Insights from faculty and strategies for professional development in clinical teaching. *J Vet Med Educ* 2008;35:397-406.
6. Boerboom TBB, Dolmans DHJM, Jaarsma ADC, Muijtjens AMM, van Beukelen P, Scherpbier AJJA. Exploring the validity and reliability of a questionnaire for evaluating veterinary clinical teachers' supervisory skills during clinical rotations. *Med Teach* 2011;33:e84-91.
7. Van der Hem-Stokroos HH, Scherpbier AJJA, Van der Vleuten CPM, De Vries H, Haarman HJTM. How effective is a clerkship as a learning environment? *Med Teach* 2001;23:599-604.
8. Ramani S, Leinster S. AMEE guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30:347-64.
9. Remmen R, Denekens J, Scherpbier AJJA, Hermann I, Van der Vleuten CPM, Royen PV, et al. An evaluation study of the didactic quality of clerkships. *Med Educ* 2000;34:460-4.
10. Van der Vleuten CPM, Scherpbier AJJA, Dolmans DHJM, Schuwirth LWT, Verwijnen GM, Wolfhagen HAP. Clerkship assessment assessed. *Med Teach* 2000;22:592-600.
11. Smith CA, Varkey AB, Evans AT, Reilly BM. Evaluating the performance of inpatient attending physicians: A new instrument for today's teaching hospitals. *J Gen Intern Med* 2004;19:766-71.
12. Boor K, Scheele F, Van der Vleuten CPM, Teunissen PW, Den Breejen EME, Scherpbier AJJA. How undergraduate clinical learning climates differ: A multi-method case study. *Med Educ* 2008;42:1029-36.
13. Yeates PJA, Stewart J, Barton JR. What can we expect of clinical teachers? Establishing consensus on applicable skills, attitudes and practices. *Med Educ* 2008;42:134-42.
14. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: Maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000;34:820-6.
15. Prideaux D, Spencer J. On theory in medical education. *Med Educ* 2000;34:888-9.
16. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010;85:1732-8.
17. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
18. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004;9:971-7.

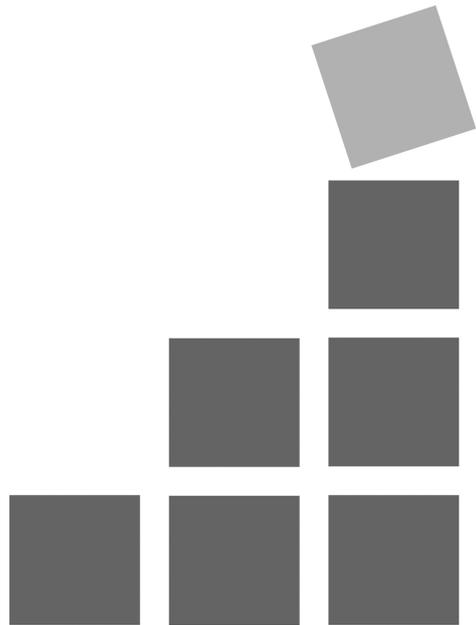
EFFECT OF FACILITATED FEEDBACK

19. Fluit C, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers. *J Gen Intern Med* 2010;25:1337-45.
20. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005;20:1159-64.
21. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:272-7.
22. Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. *Knowing, learning and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum; 1989. p. 453-94.
23. Sargeant JM, Mann KV, Sinclair D, Van der Vleuten CPM, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv in Health Sci Educ* 2008;13:275-88.
24. Hatton N, Smith D. Reflection in teacher education: Towards definition and implementation. *Teach Teach Educ* 1995;11:33-49.
25. Korthagen FAJ, Vasalos A. Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice* 2005;11:47-71.
26. Boerboom TBB, Jaarsma ADC, Dolmans DHJM, Scherpbier AJJA, Mastenbroek NJJM, van Beukelen P. Peer group reflection helps clinical teachers to critically reflect on their teaching. *Med Teach* 2011;Submitted.
27. Tigelaar DEH, Dolmans DHJM, Meijer PC, de Grave WS, van der Vleuten CPM. Teachers' interactions and their collaborative reflection processes during peer meetings. *Adv in Health Sci Educ* 2008;13:289-308.
28. Mcleod PJ, Steinert Y. Peer coaching as an approach to faculty development. *Med Teach* 2009;31:1043-4.
29. Schaub-de Jong MA, Cohen-Schotanus J, Dekker H, Verkerk M. The role of peer meetings for professional development in health science education: A qualitative analysis of reflective essays. *Adv in Health Sci Educ* 2009;14:503-13.
30. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *American Educator* 1991;15:6-11.
31. Hendriksen J. *Begeleid intervisiemodel, collegiale advisering en probleemoplossing*. [guided peer model, advising colleagues and problem solving]. Baarn, NL: Nelissen; 1997.
32. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum; 1988.
33. Dolmans DHJM, Wolfhagen HAP, Heineman E, Scherpbier AJJA. Factors adversely affecting student learning in the clinical learning environment: A student perspective. *Educ Health* 2008;20:1-10.
34. Mullan P, Sullivan D, Dielman T. What are raters rating? Predicting medical student, pediatric resident, and faculty ratings of clinical teachers. *Teach Learn Med* 1993;5:221-6.
35. Elnicki DM, Cooper A. Medical students' perceptions of the elements of effective inpatient teaching by attending physicians and housestaff. *J Gen Intern Med* 2005;20:635-9.

36. Dolmans DHJM, Wolfhagen HAP, Essed GGM, Scherpbier AJJA, van der Vleuten CPM. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. *Acad Med* 2002;77:332-5.
37. Curry L. Cognitive and learning styles in medical education. *Acad Med* 1999;74:409-13.
38. Kember D, Leung DYP, Jones A, Loke AY, McKay J, Sinclair K, et al. Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education* 2000;25:381-95.
39. Aukes LC, Geertsma J, Cohen-Schotanus J, Zwierstra RP, Slaets JP. The development of a scale to measure personal reflection in medical practice and education. *Med Teach* 2007;29:177-82.
40. Meirink JA, Meijer PC, Verloop N. A closer look at teachers' individual learning in collaborative settings. *Teachers & teaching* 2007;13:145-64.
41. Marsh HW. Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *J Educ Psychol* 2007;99:775-90.
42. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Peters WG, van Coppenolle L, Scherpbier AJJA. Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Adv in Health Sci Educ* 2009;15:315-28.
43. Watts M, Lawson M. Using a meta-analysis activity to make critical reflection explicit in teacher education. *Teach Teach Educ* 2009;25:609-16.
44. Ross JA, Bruce CD. Teacher self-assessment: A mechanism for facilitating professional growth. *Teach Teach Educ* 2007;23:146-59.

7.

General Discussion



INTRODUCTION

Clinical workplace learning is an essential component of students' preparation for entry into the veterinary or medical profession. For good clinical teaching, the fact that learning takes place in an authentic context is both essential and limiting.¹⁻³ Clinical teachers are not only responsible for the education and training of students, they also have roles to fulfil in patient care and safety, research and administration.^{4,5} The supervisory skills and teaching effectiveness of clinical teachers have a decisive impact on the learning of students in this challenging working environment.^{4,6-9} To enhance clinical teaching it is important that teachers are supported in training their future colleagues through feedback and opportunities to learn from feedback through reflection.⁴ That is why feedback and reflection cycles are crucial aspects of faculty development programmes.

Evaluation instruments to provide teachers with feedback on their teaching effectiveness are indispensable to improve clinical teaching¹⁰ and many feedback instruments have been developed for the clinical context.^{11,12} However, the theoretical dimensions underlying their composition are often not described.^{13,14} The central instrument used in this thesis to generate feedback on teaching effectiveness in the clinical environment is the Maastricht Clinical Teaching Questionnaire (MCTQ), which is based on the theory of cognitive apprenticeship.¹⁵ Although it has been thoroughly validated for the medical context,^{14,16,17} validity evidence for the veterinary context is lacking. To explore the implementation and role of a feedback and reflection cycle in a veterinary context we formulated the following central research question.

What is the validity evidence for the MCTQ in a veterinary context?

Because of the breadth of the concept of validity evidence we formulated four specific research questions to address our general research question. These research questions represent the conditions for the introduction of an instrument like the MCTQ in a new context:

- **Does a faculty development programme improve teachers' perceived competence in different teaching roles? (Chapter 2)**
- **What is the validity evidence for the MCTQ in a veterinary curriculum? (Chapter 3 and 4)**

GENERAL DISCUSSION

- Does facilitated feedback from the MCTQ stimulate clinical teachers to critically reflect on their teaching behaviour and to formulate plans for improvement of their behaviour? (Chapter 5)
- What is the effect of facilitated feedback on clinical teaching performance? (Chapter 6)

In this chapter the conclusions of the studies are summarized and discussed. We discuss the limitations of the studies and propose recommendations for future research. Finally, practical implications with respect to faculty development for veterinary clinical teachers are considered.

MAIN FINDINGS AND DISCUSSION

Following the General Introduction, **Chapter 2** presented a study exploring teachers' perceptions concerning the usefulness of a faculty development programme. There are quite a few studies published on this topic,^{18,19} but the main focus is generally on gains in classroom didactic skills. Using a questionnaire, we investigated if the nationally recognized University Teaching Qualification (UTQ) improved teachers' perceived competence in a wide range of teacher roles described in the teaching framework developed by Tigelaar et al.²⁰ The UTQ requires attendance of a two-year faculty development programme comprising workshops/seminars and personal mentoring. Participants keep a portfolio to demonstrate their gains in teaching effectiveness. We found a significant positive effect of the UTQ on teachers' perceived competence in the majority of the teaching roles for teachers who had obtained the UTQ compared to teachers who had not obtained the UTQ. This effect was particularly evident for the roles of the Person as Teacher, Content Expert, Designer, Counsellor, Organizer and Scholar. No effect was found for the Evaluator role, which may be explained by the absence of this role in the UTQ programme at the time of the study.

The main conclusion of the study in Chapter 2 is that training, feedback and opportunities for reflection have a positive effect on teachers' perceived teaching competence in a wide range of roles. This is in line with a study performed by Elliot et al., which described gains in instructional skills and greater involvement in the institution's teaching activities after an eighteen-month faculty development programme²¹ that was similar to the UTQ programme. Although the teaching roles

at issue are strongly theory based, tailored to student-centred higher education and applicable to many educational settings,²⁰ the teaching framework did not explicitly include the roles of teachers in workplace learning.

The other studies in this thesis focused on teaching in the clinical context. In **Chapters 3 and 4** we explored the validity of the Maastricht Clinical Teaching Questionnaire (MCTQ) for the veterinary clinical curriculum. The study in **Chapter 3** was designed to address the first four sources of validity as proposed by the American Psychological and Education Research Associations:²² (1) Content, (2) Response process, (3) Internal structure and (4) Relations to other variables. The content validity of the MCTQ was assured because it is based on theory on effective apprenticeship learning, the cognitive apprenticeship model²³ and because a wide array of stakeholders in medical education was involved in its development.¹⁴

The MCTQ's sound theoretical basis and development was also reflected in the results of the response process studies described in Chapter 3. Based on students' comments in these pilot studies, only minor alterations were needed to adapt the MCTQ to the veterinary setting. The students mentioned supervision time as a possible source of error due to the often short periods of supervision, which caused reluctance to evaluate some aspects of teaching effectiveness. Also Stalmeijer found that some aspects of the MCTQ can only be properly evaluated after an extended period of supervision.¹⁶ We therefore added a question to the MCTQ to control for the possible effect of supervision time. Surprisingly, no strong relation was found between the duration of supervision and evaluation outcomes. A plausible explanation is that most of the teacher behaviours described in the instrument can be demonstrated in a short period of time.

Another aspect that contributed to the response process was the distribution of the MCTQ using the Internet. The students in the study groups stated that they did not feel comfortable filling out paper forms in the clinic, a method commonly used for evaluations. In their opinion, an Internet-based questionnaire was the only way to guarantee anonymity. After changing the delivery method and adding a question about supervision time to the questionnaire, validity evidence for the response process was considered satisfactory.

The third source of validity evidence studied in Chapter 3, internal structure, was explored by analyzing the psychometric properties of the instrument using confirmatory factor analysis (CFA) and a generalizability study. CFA showed that the original MCTQ model, as developed for the medical context¹⁴ did not fit our data. The original 24-item model comprised seven domains: General Learning

GENERAL DISCUSSION

Climate (GLC), Modelling, Coaching, Scaffolding, Articulation, Reflection and Exploration. The modification indices in the CFA led to adaptation of the model. For the veterinary context a model comprising five domains with three items each sufficed: GLC, Modelling, Coaching, Articulation and Exploration (see also Box 1). In the analysis Coaching and Scaffolding aggregated into one factor as did Exploration and Reflection. We believe that this shift in domains did not violate the original model. After all, Coaching and Scaffolding are both concerned with having students perform a task, helping them when they experience difficulties while performing this task and giving them constructive feedback during and after task performance. Exploration in the original model was about setting and pursuing new learning goals. An integral part of reflection is setting new goals and proposing alternative behaviour after an experience.^{24,25} It is therefore not surprising that these two domains coincided in the CFA. The generalizability study demonstrated that at least six ratings per teacher were required to achieve a reliable overall judgement, while for reliable domain scores ten to twelve student responses were needed. Combined with the reduction in the number of items (24 items in the original model, 15 items after CFA), the modified MCTQ appears to be a practical instrument for use in the clinical context.

As mentioned above, the fourth source of validity evidence, relations to other variables, was investigated by analyzing the correlations between the domain scores and supervision time. In order to determine criterion validity we also investigated correlations between the domain scores and the overall satisfaction score. All correlations proved to be strong. After obtaining validity evidence for this wide range of sources we concluded that the MCTQ was a valid instrument for evaluating veterinary clinical teachers.

In **Chapter 4** we further investigated the validity evidence of the MCTQ in the veterinary context by exploring a commonly neglected aspect of reliability: the degree to which student evaluations are influenced by teachers' behaviours and not by other factors not related to teaching.²⁶ Multilevel analysis, a method especially suitable for nested data,²⁷ was used to quantify how much of the variance in MCTQ student ratings was due to differences between individual students or to the higher level construct that was being measured: teacher effectiveness. Another asset of this study method was that the multilevel analysis allowed us to include other, potentially predictive, characteristics that might explain MCTQ outcomes.

The results of the multilevel analysis showed that differences in teaching behaviour had a substantial effect on student evaluations. This was especially true for ratings on General Learning Climate, Modelling and Coaching. More than 20% of the

variance in these domains resided at the teacher level. In social and behavioural sciences this is considered a considerable portion.²⁷ This means that scores on these three domains are to a large extent due to the behaviours the teachers demonstrate during clinical supervision. These findings support the importance of these domains in clinical teaching and are in line with other studies reporting favourable factors for workplace learning. Boor et al., for example, emphasized the importance of a positive and safe learning climate for learning in the workplace.²⁸ Other studies described as essential aspects being a role model for students and providing them with rich and concrete feedback.^{4,9,29-31}

The results for clinical teachers' influence on the scores in the domains Articulation and Exploration were less clear-cut to some extent. This suggests that differences between students contribute somewhat more to the evaluation of these aspects, which include teaching strategies of which the success also depends on individual students' knowledge, insights and reflective skills. For example, the degree of personal reflection may vary considerably between students.^{32,33}

It is therefore not surprising that ratings on these domains are not exclusively determined by clinical teachers' behaviours. This does not mean that teachers' behaviours do not influence the ratings on these two MCTQ domains, but it suggests that student roles and characteristics should be taken into account when interpreting the feedback. Nevertheless the amount of variance related to differences between teachers in the ratings of the five domains demonstrates the reliability of the MCTQ in the veterinary context.

Another result that supports the reliability of the MCTQ is that both teacher and student characteristics, such as department and gender, had no or very little influence on MCTQ outcomes. The scientific literature is inconclusive on this issue. While some authors state that the interaction between teacher and student characteristics affects student ratings,³⁴⁻³⁶ others suggest the opposite.^{37,38} The fact that we found no teacher or student characteristics that had a substantial influence on MCTQ outcomes indicates that teaching behaviours are the main focus of the instrument, thereby supporting its validity.

Box 1: The modified Maastricht Clinical Teaching Questionnaire

Factor 1: General Learning Climate (GLC)

The clinical teacher...

1. established an environment where I felt free to ask questions or make comments.
2. showed an interest in me as a student.
3. treated me with respect.

Factor 2: Modelling

The clinical teacher...

1. demonstrated how different skills should be performed.
2. explained, while performing a task, which aspects were important and why.
3. was a role model for me.

Factor 3: Coaching

The clinical teacher...

1. provided me with constructive and concrete feedback during or following direct observation.
2. allowed me to perform tasks independently.
3. was supportive when I experienced difficulties with a task.

Factor 4: Articulation

The clinical teacher...

1. asked me to explain my reasoning and actions.
2. alerted me to gaps in my knowledge and skills.
3. asked questions to increase my knowledge and understanding.

Factor 5: Exploration

The clinical teacher...

1. stimulated me to formulate my own goals.
2. stimulated me to become aware of my own strengths and weaknesses.
3. stimulated me to think about how to improve my own strengths and weaknesses.

In **Chapter 5** we explored an entirely different aspect of the validity of the MCTQ in a veterinary context. We investigated the ‘consequences’ of evaluating clinical

teachers with the MCTQ. Facilitation of MCTQ feedback to foster reflection is important, because it is known from the literature that providing teachers with a feedback report based on student ratings does not automatically enhance their teaching performance.³⁹⁻⁴² The research question in this study was formulated as follows: 'What types of reflection do clinical teachers demonstrate in a pre-structured, written reflection report after a feedback facilitation strategy comprising self-assessment and student feedback and does reflection change when a peer reflection meeting is added to the feedback facilitation strategy?' This study also examined how clinical teachers evaluated these two different feedback facilitation strategies.

The results indicated that after receiving MCTQ generated feedback in a facilitated manner, comprising a self-assessment, a feedback report and a pre-structured reflection report, clinical teachers reflected deeply. In most cases teachers' reflection reports went beyond the mere description of the MCTQ feedback to also include alternatives for teaching practice. The latter finding is important because formulating alternative actions is an essential part of reflection.²⁵ These findings resonate with a substantial body of research indicating that student ratings coupled with a systematic intervention are substantially more effective than student ratings alone.⁴³

The results were also quite conclusive with regard to the added value of a peer reflection meeting in a feedback facilitation strategy. The clinical teachers who participated in a reflection meeting showed deeper reflection and formulated more concrete alternatives for teaching practice. When teachers critically reflect upon practice, they are more likely to be able to actively improve it,⁴⁴ so it seems beneficial to add a reflection meeting to a feedback facilitation strategy. This conclusion is in line with other studies on the beneficial effects of peer reflection meetings.^{45,46} Most teachers who participated in peer reflection as described in Chapter 5 indicated that peer reflection was especially useful for translating feedback in concrete alternatives for practice. This is one of the main goals of providing teachers with student feedback.⁴⁷ Although the study in Chapter 5 demonstrated that feedback facilitation stimulated teachers to reflect and generate alternatives for practice, it did not investigate whether teaching effectiveness actually improved as a result of feedback and reflection.

In the final study, reported in Chapter 6, we investigated the effect of facilitated feedback and reflection on clinical teachers' MCTQ scores, an aspect of validity

GENERAL DISCUSSION

evidence that is rarely examined. The results showed a positive effect of the strategy comprising self-assessment, feedback consisting of self-assessment and student ratings and a pre-structured reflection report on overall effectiveness ratings and the scale score of General Learning Climate. Also items in the domains of Modelling and Coaching were significantly positively affected by the feedback strategy. These findings are in line with a review performed by Marsh and Roche,⁴⁸ who reported evidence of the usefulness of student evaluations in improving teaching effectiveness.

The added value of the peer reflection meeting was not replicated in this study. No significant effects were found on the scale and overall scores, although ratings on three items from the Articulation and Exploration domains showed a positive impact. On one hand this result is rather puzzling in light of the findings reported in Chapter 5. After all, the teachers who participated in the peer reflection meetings were better able to reflect and formulate alternatives for their actions. One would therefore expect an improvement in MCTQ ratings after this feedback strategy. On the other hand, however, improvement of MCTQ ratings did occur for domains where differences between teachers are more important than differences between students as demonstrated by the multilevel analysis in Chapter 4. Feedback facilitation with peer reflection showed an effect for Articulation and Exploration, domains that are somewhat more sensitive to differences between students, whereas these domains showed no significant positive effects after feedback facilitation without peer reflection. Thus adding a peer reflection meeting seems to have a beneficial effect on domains that are less dependent on teachers' behaviour and therefore possibly harder for them to improve. Despite the uncertain impact of peer reflection, the results support the general conclusion that providing MCTQ feedback in combination with a feedback facilitation strategy has a positive influence on the teaching effectiveness of veterinary clinical teachers.

GENERAL CONCLUSIONS

Overall the five studies provided new valuable insights with regard to faculty development interventions in veterinary education, especially concerning the use of the MCTQ for providing feedback to clinical teachers. According to teachers' perceptions, faculty development has a positive effect on competences for a wide range of contemporary teaching roles. The modified MCTQ proved to be a valid,

reliable and usable instrument for evaluating the performance of teachers in the clinical environment. A positive finding is that differences between teachers are strong determinants of MCTQ ratings, especially for General Learning Climate, Modelling and Coaching, which is an important outcome considering that the MCTQ was designed to discriminate between the performances of individual teachers. Student and teacher characteristics had a negligible effect on the outcomes of the modified MCTQ. MCTQ feedback, facilitated through self-assessment, a feedback report and a pre-structured reflection report led to deeper reflection and increased formulation of alternatives for teaching practice. Adding peer group reflection to the feedback facilitation enhanced this process. These results are promising for improvements in clinical teaching practice. Besides stimulating reflection, the combination of feedback and feedback facilitation had a positive effect on MCTQ scores compared to the scores of teachers who did not receive feedback. Thus a faculty development approach in which clinical teachers are offered facilitated MCTQ feedback has proved valuable and shows promise for future faculty development.

LIMITATIONS

There are some limitations to the studies in this thesis. Firstly, the beneficial effects of the University Teaching Qualification (UTQ) were based on teachers' self-assessment, which has been shown to be not entirely reliable. It is known that there are groups of teachers whose self-assessments are too low or too high.⁴⁹ The lack of evidence from other sources, like student, expert or peer ratings, limits the evidence of this particular study. This limitation was less applicable to the other studies in this thesis, however, where we used student ratings from the modified Maastricht Clinical Teaching Questionnaire which was validated for the veterinary context to investigate the effects of providing facilitated feedback as a component of a faculty development programme. The impact and value of the MCTQ data might have been enhanced if we had used more data sources. Although student feedback from the MCTQ proved to be valid, reliable and able to stimulate reflection and practice improvement, we did not explore the effects of adding feedback from other sources, such as peers, educational experts or even patient owners. Different viewpoints might have enriched the feedback on some of the MCTQ domains⁵⁰ and helped teachers to better reflect on their feedback. In

GENERAL DISCUSSION

Chapter 5, some clinical teachers doubted the reliability of the feedback reports. Adding more sources could have dealt with this issue.

A second potential limitation is that we used pooled data from fifth- and sixth-year students in the studies in which MCTQ data were used. This is a common limitation among validation studies.^{11,51} Chapter 4 explored if pooling data from fifth- and sixth-year students limited the reliability of the instrument. Although the effect was small, we did find a relationship between study year and MCTQ scores. The sixth-year students gave significantly higher ratings compared to the fifth-year students on the domains of General Learning Climate, Coaching, Articulation and Exploration. There are some possible explanations for this finding. For example, students in the last phase of their study may have different needs regarding certain domains of the MCTQ and therefore value these differently than students who have just begun their clinical training. It is also possible that sixth-year students are better able to put themselves in the position of the teacher and may therefore be less strict in their evaluations. Although it was not the prime purpose of this thesis, obtaining insight into why different groups of students valued some domains of the MCTQ differently would have helped us to better interpret the outcomes of the MCTQ data.

Thirdly, we did not pay much attention to effects of the clinical context on feedback, reflection and actions for improvement. As stated in Chapter 1, teachers face competing demands from teaching, research and patient care.^{2,4,5,13,30,52} Faculty development, including feedback on teaching effectiveness, is intended to help teachers to cope with these challenges. However, we did not explore how the clinical context influenced teaching effectiveness as measured with the MCTQ. Also we did not investigate specifically if providing facilitated MCTQ feedback helped clinical teachers to deal with the conflicting demands in the clinical environment. Information of this kind might help administrators to optimize the clinical learning environment and so improve teaching effectiveness in the clinical setting.⁵³

FUTURE RESEARCH

Although this thesis focused on many aspects of the implementation of the MCTQ in a veterinary context, several questions have remained unanswered. In fact, new questions have arisen from the studies in this thesis.

As proposed in the 'Limitations' section more sources could be tapped to evaluate clinical teachers. Multisource feedback has been used extensively in industry and the health care sector and may also be useful to enhance the reliability of data obtained with instruments like the MCTQ.^{50,54} Although this kind of feedback is no panacea for human judgement error, it may enrich the feedback,⁵⁰ providing the desired behaviours are described clearly.⁵⁴ When transferring the MCTQ to other types of respondents, new validation evidence should be gathered concerning response process but also psychometric properties.⁵⁵⁻⁵⁷ Obvious respondents would be educational experts or peers. But also patient owners or supervisors could prove to be reliable information sources.⁵⁸ As MCTQ feedback from students already stimulates reflection and improvement in teaching effectiveness, the added value of multisource feedback should be demonstrated thoroughly.

A whole other aspect in need of research is how to combine the MCTQ feedback-reflection cycle with other faculty development initiatives. The evaluation of individual clinical teachers with the MCTQ will give rise to new faculty development needs, which should be explored.⁵⁹⁻⁶⁰ As shown in a review by Steinert et al.,¹⁸ there are many types of interventions feasible for enhancing teaching effectiveness. Interventions commonly used are workshops, short courses or fellowships, but also longitudinal programmes, focusing on self-directed learning.^{9,53} Empowering and rewarding teachers and letting faculty learn together in intra- and interprofessional groups are also powerful options for faculty development.⁵³ These kinds of interventions could be included in the University Teaching Qualification programme or in a future didactic programme for lifelong learning. It seems safe to predict that licensing or accreditation of medical teachers are future scenarios in medical education.⁵⁹ Further studies are required to investigate which interventions are most efficient and successful for clinical teachers⁵⁹ and what is needed for sustaining new attitudes and behaviours.⁶¹ In this thesis we followed groups of clinical teachers during one measurement-intervention-measurement cycle. It would be most interesting to investigate the effects of a longitudinal programme of feedback and feedback facilitation, preferably comprising peer reflection meetings. Do clinical teachers actually succeed in improving their teaching effectiveness? Are we going to see an effect on students' learning results? Is there a shift in feedback culture observable and does clinical teaching become more prominent in the career perspectives of clinical staff?

In some cases providing facilitated MCTQ feedback may not be enough for behavioural change. Remediation is needed in those cases. It is therefore

GENERAL DISCUSSION

worthwhile to explore methods of training teachers in the competences for the different domains of the MCTQ. It is obvious that training teachers how to establish a safe learning climate likely differs from training in stimulating students to reflect. Another important topic of future research is how stimulate active engagement in faculty development of veterinary clinical teachers, who are faced with highly demanding service commitments.⁶²

The study in Chapter 3 revealed differences in ratings between fifth- and sixth-year students. Although we now know that sixth-year students give somewhat higher ratings, we still do not know what causes this effect. It may be interpreted as an indication that the required degree of cognitive apprenticeship varies depending on the level of training. More in-depth research, preferably qualitative, is needed to explore the needs, wishes and perceptions of both groups of students concerning the different domains of the MCTQ.

A last and important aspect to investigate is the role of institutional organizations. As proposed in Chapter 1 and the 'limitations' section of this chapter, teachers face high and competing demands in the clinical workplace. Future research should look for ways to deal with this to stimulate the highest quality of clinical teaching possible.⁵³

PRACTICAL IMPLICATIONS

Based on the findings from the studies in this thesis some recommendations for the evaluation of veterinary clinical teachers can be made.

The studies on the validity evidence of the Maastricht Clinical Teaching Questionnaire (MCTQ) in a veterinary curriculum showed that the modified instrument is valid and reliable for the veterinary context, even when supervision periods are rather short. The data indicated that ten to twelve evaluations per individual teacher are sufficient to obtain reliable data per domain. When the MCTQ is used to evaluate teachers this is the number of responses to pursue. Obtaining many evaluations per teacher is even more important when the MCTQ is used for promotion or tenure decisions.^{11,35} In those cases it is also recommended to collect data from more sources than only students. One might for instance use 360° multisource feedback from patient owners, veterinary technicians, peers and supervisors.^{50,58}

In some situations it will be difficult to collect ten to twelve evaluations per individual teacher. This does not have to be a big problem. Of course, high stakes decisions cannot be made based on information from a limited number of sources, but, according to the findings of Chapters 5 and 6, feedback from at least six student ratings is sufficient to stimulate reflection on and improvement of clinical teaching effectiveness. In any case, depending on the number of teachers to be evaluated, a large amount of data has to be processed. We therefore recommend using an Internet-based questionnaire to collect evaluations. When implemented properly, this data collection method is not only safe for users (in this case the students) but also fast and cost-efficient.

If MCTQ results are going to be used for purposes of human resource management, administrators should be trained to read the MCTQ data and to coach teachers in improving their clinical teaching practices. It has to be pointed out that outcomes on the domains of Articulation and Exploration should be interpreted with some caution. Teacher behaviour plays an important role in how these aspects are evaluated, but the influence of student characteristics should not be neglected. Improving teaching effectiveness in these two domains will require change not only of individual teachers but also of students. The latter might be achieved by training students in reflection skills and by stimulating active feedback seeking behaviours.

Some remarks should be made about the training of teachers. The teachers in our dataset who had obtained an UTQ did not perform significantly better on the MCTQ. At the Faculty of Veterinary Medicine of Utrecht University, 'supervising individual students' is one of the training sessions that teachers attend for obtaining the UTQ. It is probably advisable to use the MCTQ domains as the cornerstones of this training. The effect of this change in training may be amplified by also including in the UTQ programme the feedback-reflection cycle with the MCTQ as described in Chapters 5 and 6. As mentioned before, providing clinical teachers with feedback is only useful when it stimulates reflection and improves teaching effectiveness.^{55,63} This implies that it is important to help teachers to translate feedback into a concrete plan of action and to prevent negative emotions that can impede reflection.^{42,64,65} Based on the findings in Chapter 5 we would recommend to combine feedback from student ratings with a self-assessment and pre-structured reflection report. Again we would like to stress the importance of a peer reflection meeting. This type of intervention can be very instructive and supportive,^{45,46} especially for improving teaching effectiveness in Articulation and Exploration, as indicated in Chapter 6.

GENERAL DISCUSSION

In conclusion, taking into account the aspects discussed above, we think we are justified in concluding that the Maastricht Clinical Teaching Questionnaire is a useful, valid, reliable and practical instrument for the evaluation of clinical teachers in a veterinary curriculum. Positive effects on teaching effectiveness in the clinical context can be expected when teachers are given MCTQ feedback in a facilitated manner. It is important to keep collecting longitudinal data on the functioning of the clinical teachers to monitor both the growth and decline of teaching effectiveness. This has added value for future feedback reports compared to only providing the mean scores of other teachers. Many institutes are not successful in collecting and archiving this longitudinal data.⁴³ Based on the findings of this thesis we would recommend putting considerable effort into prolonging the evaluation of clinical staff and helping them to become better teachers.

REFERENCES

1. Spencer J. ABC of learning and teaching in medicine: Learning and teaching in the clinical environment. *Brit Med J* 2003;326:591-4.
2. Lane IF, Strand E. Clinical veterinary education: Insights from faculty and strategies for professional development in clinical teaching. *J Vet Med Educ* 2008;35:397-406.
3. Bolt DM, Witte TH, Lygo-Baker S. The complex role of veterinary clinical teachers: How is their role perceived and what is expected of them? *J Vet Med Educ* 2010;37:388-94.
4. Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, et al. Clinical teaching: Maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000;34:820-6.
5. Hoffman KG, Donaldson JF. Contextual tensions of the clinical environment and their influence on teaching and learning. *Med Educ* 2004;38:448-54.
6. Prideaux D, Spencer J. On theory in medical education. *Med Educ* 2000;34:888-9.
7. Irby DM, Papadakis M. Does good clinical teaching really make a difference? *Am J Med* 2001;110:231-2.
8. Deketelaere A, Kelchtermans G, Struyf E. Disentangling clinical learning experiences: An exploratory study on the dynamic tensions in internship. *Med Educ* 2006;40:908-15.
9. Ramani S, Leinster S. AMEE guide no. 34: Teaching in the clinical environment. *Med Teach* 2008;30:347-64.
10. Ramani S. Twelve tips to promote excellence in medical teaching. *Med Teach* 2006;28:19-23.
11. Beckman TJ, Ghosh AK, Cook DA, Erwin PJ, Mandrekar JN. How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 2004;9:971-7.
12. Fluit C, Bolhuis S, Grol R, Laan R, Wensing M. Assessing the quality of clinical teachers. *J Gen Intern Med* 2010;25:1337-45.
13. Dolmans DHJM, Wolfhagen HAP, Gerver WJ, de Grave W, Scherpbier AJJA. Providing physicians with feedback on how they supervise students during patient contacts. *Med Teach* 2004;26:409-14.
14. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach* 2008;30:272-7.
15. Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: Resnick LB, editor. *Knowing, learning and instruction: Essays in honor of Robert Glaser*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.; 1989. p. 453-94.
16. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Scherpbier AJJA. Cognitive apprenticeship in clinical practice: Can it stimulate learning in the opinion of students? *Adv Health Sci Educ* 2009;14:535-46.
17. Stalmeijer RE, Dolmans DHJM, Wolfhagen HAP, Muijtjens AMM, Scherpbier AJJA. The Maastricht Clinical Teaching Questionnaire (MCTQ) as a valid and reliable instrument for the evaluation of clinical teachers. *Acad Med* 2010 ;85:1732-8.
18. Steinert Y, Mann K, Centeno A, Dolmans DHJM, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME guide no. 8. *Med Teach* 2006;28:497-526.

GENERAL DISCUSSION

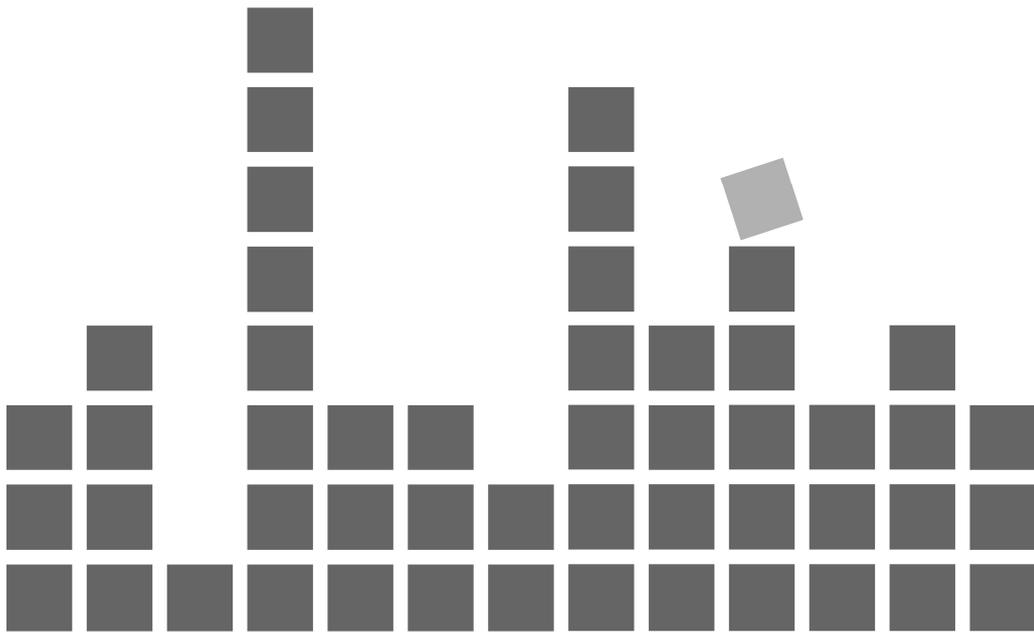
19. Postareff L, Lindblom-Ylänne S, Nevgi A. The effect of pedagogical training on teaching in higher education. *Teach Teach Educ* 2007;23:557-71.
20. Tigelaar DEH, Dolmans DHJM, Wolfhagen HAP, Van der Vleuten CPM. The development and validation of a framework for teaching competencies in higher education. *High Educ* 2004;48:253-68.
21. Elliot DL, Skeff KM, Stratos GA. How do you get to the improvement of teaching? A longitudinal faculty development program for medical educators. *Teach Learn Med* 1999;11:52-7.
22. American Education Research Association and American Psychological Association. Standards for educational and psychological testing. Washington DC: American Education Research Association; 1999.
23. Collins A, Brown JS, Holum A. Cognitive apprenticeship: Making thinking visible. *American Educator* 1991;15:6-11.
24. Hatton N, Smith D. Reflection in teacher education: Towards definition and implementation. *Teach Teach Educ* 1995;11:33-49.
25. Korthagen FAJ, Vasalos A. Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice* 2005;11:47-71.
26. Downing SM. Threats to the validity of clinical teaching assessments: What about rater error? *Med Educ* 2005;39:353-5.
27. Snijders TAB, Bosker RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage;2004.
28. Boor K, Scheele F, Van der Vleuten CPM, Teunissen PW, Den Breejen EME, Scherpbier AJJA. How undergraduate clinical learning climates differ: A multi-method case study. *Med Educ* 2008;42:1029-36.
29. Mullan P, Sullivan D, Dielman T. What are raters rating? Predicting medical student, pediatric resident, and faculty ratings of clinical teachers. *Teach Learn Med* 1993;5:221-6.
30. Dolmans DHJM, Wolfhagen HAP, Essed GGM, Scherpbier AJJA, van der Vleuten CPM. The impacts of supervision, patient mix, and numbers of students on the effectiveness of clinical rotations. *Acad Med* 2002;77:332-5.
31. Elnicki DM, Cooper A. Medical students' perceptions of the elements of effective inpatient teaching by attending physicians and housestaff. *J Gen Intern Med* 2005;20:635-9.
32. Kember D, Leung DYP, Jones A, Loke AY, McKay J, Sinclair K, et al. Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education* 2000;25:381-95.
33. Aukes LC, Geertsma J, Cohen-Schotanus J, Zwierstra RP, Slaets JP. The development of a scale to measure personal reflection in medical practice and education. *Med Teach* 2007;29:177-82.
34. Leone-Perkins M, Schnuth R, Kantner T. Preceptor-student interactions in an ambulatory clerkship: Gender differences in student evaluations of teaching. *Teach Learn Med* 1999;11:164-7.
35. Centra JA, Gaubatz NB. Is there gender bias in student evaluations of teaching? *J Higher Educ* 2000;71:17-33.

36. Beckman TJ, Mandrekar JN. The interpersonal, cognitive and efficiency domains of clinical teaching: Construct validity of a multi-dimensional scale. *Med Educ* 2005;39:1221-9.
37. Ory JC. Faculty thoughts and concerns about student ratings. *New Directions for Teaching and Learning* 2001;2001:3.
38. McOwen KS, Bellini LM, Guerra CE, Shea JA. Evaluation of clinical faculty: Gender and minority implications. *Acad Med* 2007;82:94-6.
39. Atwater LE, Waldman DA, Brett JF. Understanding and optimizing multisource feedback. *Hum Resour Manage* 2002;41:193-208.
40. Seifert CF, Yukl G, McDonald RA. Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *J Appl Psychol* 2003;88:561-9.
41. Smither JW, London M, Reilly RR. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Person Psychol* 2005;58:33-66.
42. Sargeant JM, Mann KV, Sinclair D, Van der Vleuten CPM, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv in Health Sci Educ* 2008;13:275-88.
43. Marsh HW. Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *J Educ Psychol* 2007;99:775-90.
44. Watts M, Lawson M. Using a meta-analysis activity to make critical reflection explicit in teacher education. *Teach Teach Educ* 2009;25:609-16.
45. Meirink JA, Meijer PC, Verloop N. A closer look at teachers' individual learning in collaborative settings. *Teachers & teaching* 2007;13:145-64.
46. Tigelaar DEH, Dolmans DHJM, Meijer PC, de Grave WS, Van der Vleuten CPM. Teachers' interactions and their collaborative reflection processes during peer meetings. *Adv in Health Sci Educ* 2008;13:289-308.
47. Ward JR, McCotter SS. Reflection as a visible outcome for preservice teachers. *Teach Teach Educ* 2004;20:243-57.
48. Marsh HW, Roche LA. Making students' evaluations of teaching effectiveness effective. *Am Psychol* 1997;52:1187-97.
49. Ross JA, Bruce CD. Teacher self-assessment: A mechanism for facilitating professional growth. *Teach Teach Educ* 2007;23:146-59.
50. Berk RA. Using the 360° multisource feedback model to evaluate teaching and professionalism. *Med Teach* 2009;31:1073-80.
51. Beckman TJ, Cook DA, Mandrekar JN. Factor instability of clinical teaching assessment scores among general internists and cardiologists. *Med Educ* 2006;40:1209-16.
52. Irby DM, Bowen JL. Time-efficient strategies for learning and performance. *The Clinical Teacher* 2004;1:23-8.
53. Steinert Y, Mann KV. Faculty development: Principles and practices. *J Vet Med Educ* 2006;33:317-24.
54. Wood L, Hassell A, Whitehouse A. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Med Teach* 2006;28:e185-91.

GENERAL DISCUSSION

55. Snell L, Tallett S, Haist S, Hays R, Norcini J, Prince K, et al. A review of the evaluation of clinical teaching: New perspectives and challenges. *Med Educ* 2000;34:862-70.
56. Beckman TJ, Cook DA, Mandrekar JN. What is the validity evidence for assessments of clinical teaching? *J Gen Intern Med* 2005;20:1159-64.
57. Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Med Educ* 2010;44:165-76.
58. Violato C, Lockyer JM, Fidler H. Changes in performance: A 5-year longitudinal study of participants in a multi-source feedback programme. *Med Educ* 2008;42:1007-13.
59. Benor DE, Mahler S. Teacher training and faculty development in medical education. *Isr J Med Sci*. 1987;23:976-82.
60. McLean M, Cilliers F, Van Wyk JM. Faculty development: Yesterday, today and tomorrow. *Med Teach* 2008;30:555-84.
61. Bowe CM, Lahey L, Armstrong E, Kegan R. Questioning the 'big assumptions' Part I: Addressing personal contradictions that impede professional development. *Med Educ* 2003;37:715-22.
62. Steinert Y, McLeod PJ, Boillat M, Meterissian S, Elizov M, Macdonald ME. Faculty development: A 'field of dreams'? *Med Educ* 2009;43:42-9.
63. Morrison J. ABC of learning and teaching in medicine: Evaluation. *Brit Med J* 2003;326:385-7.
64. DeNisi AS, Kluger AN. Feedback effectiveness: Can 360-degree appraisals be improved? *The Acad Manage Exec* 2000;14:129-39.
65. Overeem K, Wollersheim H, Driessen E, Lombarts K, Van de Ven G, Grol R, et al. Doctors' perceptions of why 360-degree feedback does (not) work: A qualitative study. *Med Educ* 2009;43:874-82.

Summary



SUMMARY

All veterinary and medical students are exposed to the clinical context in patient- and case-based training in the clinical workplace. Many factors influence workplace-based learning, such as the learning climate and the case mix to which students are exposed, but above all it is the clinical teachers that determine students' learning experiences. Exploring and improving the teaching effectiveness of clinical teachers through feedback and reflection is the central topic of this thesis.

Chapter 1, the General Introduction, summarises what is already known about workplace learning, the role of the clinical teacher and how teachers can be supported. The role of clinical education within the curriculum is elaborated. The rise of clinical training worldwide resulted primarily from the fact that patient-based education became a requirement for programme accreditation. Within the clinical context students can integrate insights and skills from theoretical courses with practical skills.

The next section of the General Introduction deals with the role of the clinical teacher. The teaching effectiveness of individual clinical teachers is essential for the success of workplace-based learning. Clinical teachers have to bridge the gap between theory and clinical practice and create learning opportunities for students without compromising the safety of both patients and students. Because of the key role of clinical teachers, faculty development activities for these teachers are receiving a great deal of attention worldwide.

The importance of faculty development and recommendations for research around this topic are further elaborated in this chapter. Feedback for clinical teachers on their teaching performance plays a central role and many evaluation instruments have been developed to collect feedback. However, many instruments are lacking in theoretical background and their validity is not always established. An evaluation instrument with a strong theoretical basis in the cognitive apprenticeship model is introduced: the Maastricht Clinical Teaching Questionnaire (MCTQ). It comprises several teaching strategies for workplace learning in medical education. Although the MCTQ is validated for the medical context, evidence on its validity and reliability in the veterinary context is lacking. When an evaluation instrument is used in a new context a broad range of validity evidence should be collected. A model describing different approaches to validity evidence is introduced.

This problem definition leads to the central research question of this thesis:

What is the validity evidence of the MCTQ in a veterinary context?

Because validity is a broad concept, four sub-questions were formulated to address the main research question.

- 1. Does a faculty development programme improve teachers' perceived competence in different teaching roles? (Chapter 2)**
- 2. What is the validity evidence for the MCTQ in a veterinary curriculum? (Chapters 3 and 4)**
- 3. Does facilitated feedback from the MCTQ stimulate clinical teachers to critically reflect on their teaching behaviour and formulate plans to improve it? (Chapter 5)**
- 4. What is the effect of facilitated feedback on clinical teachers' teaching performance? (Chapter 6)**

In Chapters 2 through 6, the research questions are studied using both quantitative and qualitative methods. Data are obtained from teachers' and students' perceptions, questionnaires and reflection reports written by teachers.

Chapter 2 focuses on the different teaching roles in a contemporary curriculum and the influence of a faculty development programme on teachers' perceived competence in these roles. A framework comprising the seven roles of a teacher in a student-centred curriculum is described as well as the Dutch nationally recognized University Teaching Qualification (UTQ). The programme leading up to this qualification encompasses learning by doing, reflection and training. Pilot studies among five experienced teachers were conducted to investigate the relevance and clarity of the seven teacher roles.

Using an Internet-based questionnaire, the modified framework was presented to all teachers of the Faculty of Veterinary Medicine. The teachers were asked to rate their level of competence for each role. More than 190 teachers responded (77%). Of these teachers 55% had obtained the UTQ. Stepwise linear regression analyses

demonstrated that the UTQ had a positive effect on perceived teaching competence. These findings are supported by similar studies in the medical education context.

These positive results led to the study described in **Chapter 3**, in which the above mentioned model of sources of validity evidence plays a central role. Four aspects of validity are investigated for the original, seven domains of the MCTQ in a veterinary context. The content of the instrument is guaranteed by the theory on workplace learning which underlies the domains. In order to explore the response process, pilot groups of students were asked to complete the questionnaire while explicating their thought processes. The students also discussed the items of the MCTQ and the preferred method of distribution. The instrument was modified based on the results of the pilot groups. At the students' recommendation, an additional item 'supervision time' was added to the questionnaire, because they were hesitant to evaluate supervisors on certain aspects of the MCTQ because their contacts with supervisors were often rather brief. Next, the MCTQ was distributed among all fifth- and sixth-year students as an online questionnaire. The psychometric validity and reliability of the instrument were assessed using confirmatory factor analysis and a generalizability study. The final aspects of validity that were addressed in this study are the relations between the outcomes on the MCTQ domains and the items concerning supervision time and Overall teaching effectiveness.

The results of the psychometric analyses demonstrated the validity of the modified MCTQ comprising fifteen items relating to five domains (General Learning Climate, Modelling, Coaching, Articulation and Exploration). For the five domains, reliable outcomes were obtained with ten to twelve evaluations per teacher. Six to eight evaluations sufficed for a reliable outcome on Overall teaching effectiveness. Contrary to students' expectations in the pilot study, the results showed no large effect of supervision time on the MCTQ domains. These results demonstrated that the MCTQ is a valid and reliable instrument for the evaluation of veterinary clinical teachers, provided that at least six evaluations per teacher are collected.

Although the previous chapter focused on a broad range of validity evidence, several aspects remained underexposed. It remained to be investigated to what

SUMMARY

extent between-teacher differences accounted for the MCTQ scores. In other words, to what extent did the clinical teacher's behaviour influence the outcomes on the MCTQ? Also the effects of certain characteristics (like gender and experience) of teachers, students and their interactions had not been examined. **Chapter 4** describes a study in which these often underexposed aspects were investigated through multilevel analyses.

Data from 1277 completed evaluations of 168 different clinical teachers were divided in two models comprising components related to between-teacher differences and between-student differences. The results indicated that a fairly large amount of the variance in the domain scores on General Learning Climate, Modelling and Coaching resided at the teacher level, while this applied to a lesser extent for the domains of Articulation and Exploration. The inclusion of explanatory variables barely changed the amount of explained variance. This means that the MCTQ outcomes were largely determined by teachers' behaviours, especially for the domains related to Learning Climate, Modelling and Coaching. Differences between students and teacher and student characteristics had no decisive effect on these ratings. This is a positive result because an instrument like the MCTQ should mainly provide information about the effectiveness and behaviours of the teachers.

Two aspects outside teaching behaviours were found to affect the teachers' scores. Sixth-year students generally, but not consistently, rated their teachers higher compared to fifth-year students. This is probably due to the fact that students have different needs and demands with respect to guidance in these two different phases of the curriculum. The results also indicated a small negative relationship between years of teaching experience and MCTQ scores. This decrease in teaching performance, despite increasing experience, may be explained by the fact that there is no continuous training of teachers.

In **Chapter 5** another important aspect of validity evidence is explored: does MCTQ feedback stimulate clinical teachers to reflect on their performance? It is stated in the introduction to this study that in many cases the mere provision of feedback is unlikely to be an incentive to reflect on and improve teaching effectiveness. A feedback facilitation strategy may be able to solve this problem.

A group of 56 veterinary clinical teachers participated in this study. Reliable student ratings on these teachers' teaching performance were collected using the

MCTQ. The teachers were assigned to one of two groups through maximum variation sampling. The first group followed a standard strategy comprising self-assessment, a feedback report and a pre-structured reflection report. The other group received the same strategy but also participated in an additional peer reflection meeting. The written reflection reports served as a qualitative data source and were analysed on the depth of reflection. The concreteness of the improvements proposed by the teachers was also examined. An online questionnaire was used to explore the teachers' opinions on the different steps of the feedback facilitation strategies.

The results showed that the majority of the clinical teachers showed forms of reflection after completing the standard strategy and proposed concrete alternatives for teaching practice. The group that completed the strategy including a peer reflection meeting demonstrated more critical reflection. Their proposed alternatives for teaching practice were also more concrete. This was a positive finding, since concrete alternatives can promote real improvement in educational performance. The positive effect of the peer reflection meeting was partly explained by the teachers' remarks in response to the online questionnaire. According to the teachers, the peer reflection meetings contributed greatly in translating the aspects mentioned in their feedback reports into concrete steps for improvement. Adding a feedback strategy, including a peer reflection meeting, thus appears to be very beneficial for processing feedback generated by the MCTQ. It remained unclear, however, whether the reflections and concrete proposals for improvement would actually lead to more effective teaching.

Chapter 6 aimed to investigate the effect of facilitated feedback on teaching effectiveness as measured by the MCTQ. The introduction of this chapter explains that evaluation of clinical teachers and providing facilitated feedback is important, but that evidence on the effectiveness of these interventions is scarce. This led to the following research questions: Does providing teachers with MCTQ feedback, facilitated through a strategy comprising self-assessment, a feedback report and a pre-structured reflection report, lead to increased teaching effectiveness as measured with the MCTQ? And does teaching performance improve even more when a peer reflection meeting is added to the feedback facilitation strategy?

SUMMARY

Three groups of teachers for whom six or more MCTQ evaluations were available were included in the study. One group was assigned to the facilitation strategy without peer reflection meeting and one group took part in a peer reflection meeting in addition to the standard strategy. A third group served as control group and did not receive any feedback or facilitation. As of February 2010 a one-year post-measurement was performed. Regression analysis was performed to identify the effects of the strategies on the domains and items of the MCTQ.

The differences between the post- and pre-measurement demonstrated that the teachers who participated in a feedback strategy on average scored better on the MCTQ after the intervention. The regression analyses indicated that the standard strategy had a positive effect on the Overall effectiveness score and the domain score on Learning Climate. Significant positive effects were also found for items within the Modelling and Coaching domains. The additional peer reflection meeting had a significant positive effect on a few items in the Articulation and Exploration domains. From these results we can conclude that facilitated feedback has a positive effect on teaching effectiveness. An additional peer reflection leads to further improvement in domains where active participation of students is essential.

Chapter 7 summarises the previous chapters and presents conclusions. Also some limitations of the studies are discussed and directions for future research provided. Finally, practical implications arising from the conclusions of the studies are described.

The five studies in this thesis provided new insights into faculty development, especially concerning generating and providing feedback for veterinary clinical teachers using the Maastricht Clinical Teaching Questionnaire (MCTQ). Faculty development in the form of successful completion of a programme leading to the University Teaching Qualification (UTQ) had a positive effect on teachers' perceived competence in a wide range of teaching roles. The MCTQ proved to be a valid, reliable and practical instrument for the evaluation of teaching roles in the clinical setting. A favourable finding of the multilevel analyses was that differences between teachers had a substantial influence on student evaluations with the MCTQ. Furthermore, characteristics like department and gender had little effect on the outcomes of these evaluations.

Providing and facilitating feedback in a strategy comprising self-assessment, a feedback report and a pre-structured self-reflection assignment promoted critical reflection and the development of alternatives for teaching practice. These reflections and proposed alternatives appeared to become more critical and concrete when teachers also participated in a peer reflection meeting. From the last study we can conclude that these reflections and proposed alternatives also resulted in improved teaching effectiveness. After receiving the feedback strategy teachers had better scores on the MCTQ. Peer feedback appeared to be especially beneficial for scores on the domains where the active participation of students is required.

An analysis of the limitations of the studies gave rise to some recommendations. In the first study we examined teachers' self-perceptions. This makes the outcomes of this study somewhat less reliable. In the remaining studies more objective data sources were used. In this section attention is paid to the effects of study year on MCTQ outcomes, although this effect was limited. This chapter also describes the possibility of using more sources than just student evaluations to generate feedback for teachers. The potential of using other sources has not been explored in this thesis. Another limitation is that little attention was paid to the effects of the clinical context on the educational achievements of clinical teachers.

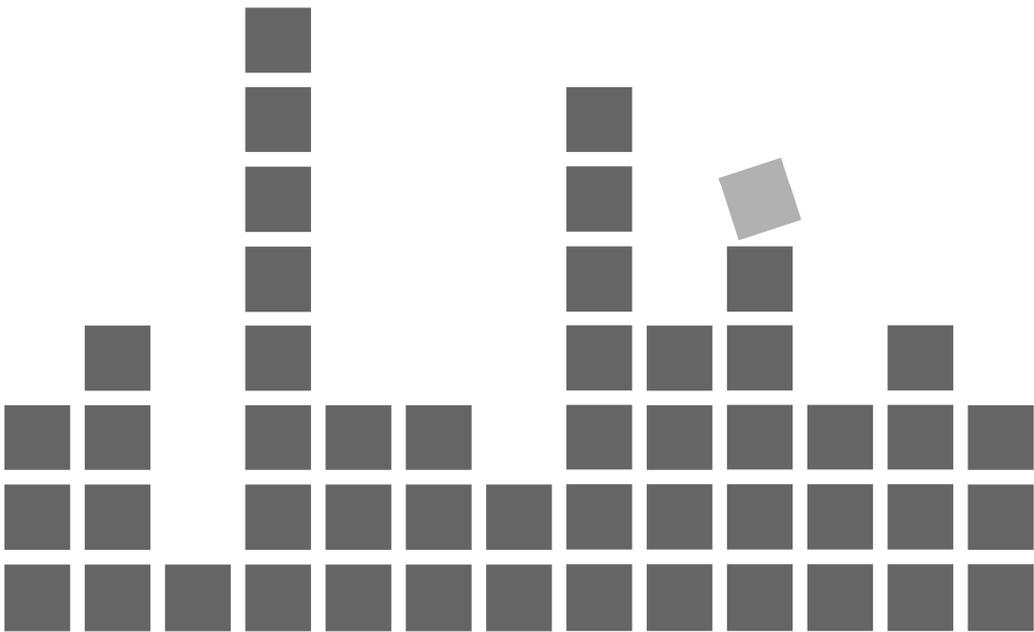
This thesis also gives rise to recommendations for further research. The studies showed that feedback from student evaluations often leads to improved educational performance. It seems useful to investigate if additional feedback from other sources, for example colleagues, will lead to even better results. Another aspect that deserves further investigation is how the MCTQ feedback and reflection cycle could be embedded in other faculty development activities. This thesis provides no insights into the training of teachers to improve performance in relation to different domains of the MCTQ or what is needed to sustain good evaluations on the MCTQ. Future research should also focus on the differences between fifth- and sixth-year students in their needs and appreciation concerning the different MCTQ domains. Such research could help to better interpret the differences in MCTQ evaluations between these groups of students. Further studies should also focus on the clinical context. How could this be organized to optimize the use of the MCTQ teaching strategies?

The studies in this thesis also have educational implications. Applying the MCTQ as an evaluation tool can lead to improvement of teaching effectiveness if a number

SUMMARY

of conditions are met. For reliable feedback at least six or more evaluations per teacher are needed and the feedback should be accompanied by a facilitation strategy, preferably one including a peer reflection meeting. Attention should also be paid to the training of clinical teachers with the various domains of MCTQ as cornerstones. On a final note, it is emphasized that a continued focus on the evaluation of clinical teachers is desirable in order to help clinical staff to become (even) better teachers.

Samenvatting



SAMENVATTING

Alle veterinaire en medische studenten worden tijdens hun opleiding geconfronteerd met werkplekleren. Tijdens dit leren binnen de klinische context volgen zij het zogenaamde patiënt- of casusgebonden onderwijs. Veel factoren zijn belangrijk voor het welslagen van dit specifieke type onderwijs, zoals het leerklimaat en het patiëntenaanbod, maar bovenal spelen de onderwijsprestaties van de individuele docent een belangrijke rol in de effectiviteit van dit deel van de opleiding. Het in kaart brengen en verbeteren van de onderwijsprestaties van klinische docenten middels feedback en reflectie is het centrale onderwerp van dit proefschrift.

Hoofdstuk 1, de introductie, bespreekt hetgeen reeds bekend is omtrent het leren op de werkplek, de rol van de docenten hierin en de manier waarop docenten in deze rol kunnen worden ondersteund. In eerste instantie wordt de rol van het klinisch onderwijs binnen het curriculum uitgewerkt. De opkomst van klinische opleidingsonderdelen in curricula wereldwijd was voornamelijk het gevolg van het feit dat het aanbieden van patiëntgebonden onderwijs een accreditatie-eis werd. Binnen de klinische context kunnen studenten kennis, inzichten en vaardigheden uit de meer theoretische onderdelen van hun studie, integreren met praktische vaardigheden.

Het volgende deel van de introductie behandelt de rol van de docent binnen het klinisch onderwijs. De onderwijsprestaties van individuele docenten zijn essentieel voor de effectiviteit van werkplekleren. Docenten dienen een brug te slaan tussen de theorie en praktijk en te zorgen voor onderwijsmomenten waarbij de student leert van de casus, zonder daarbij de student en de patiënt in gevaar te brengen. Vanwege de sleutelrol die de docent heeft in de klinische onderwijsfase wordt er wereldwijd veel aandacht besteed aan de professionalisering van deze docenten.

Het belang van docentprofessionalisering en aanbevelingen wat betreft onderzoek naar dit onderwerp worden verder uitgewerkt. Een belangrijk aspect van docentprofessionalisering is het docenten voorzien van feedback met betrekking tot hun onderwijsprestaties. Voor dit doel zijn er in de afgelopen decennia vele evaluatie-instrumenten ontwikkeld. Echter de theoretische basis en aspecten van validiteit zijn niet in alle gevallen gewaarborgd. Een voorbeeld van een evaluatie-instrument met een sterke theoretische basis wordt aangehaald. Het betreft de Maastricht Clinical Teaching Questionnaire, kortweg MCTQ. Dit model met doceerstrategieën is toegespitst op de klinische context en is gebaseerd op een leermeester-gezel model met een aanzienlijk cognitieve component. Hoewel het model uitgebreid gevalideerd is voor de humaan medische context, is er voor de veterinaire context nog geen bewijs van validiteit en betrouwbaarheid. Het

SAMENVATTING

verzamelen van validiteitsbewijs is belangrijk zodra men een instrument als de MCTQ introduceert in een nieuwe context. Een model dat verschillende invalshoeken van validiteitsbewijs beschrijft wordt geïntroduceerd.

Deze probleemstelling leidt vervolgens tot de hoofdvraag van dit proefschrift:

Wat is het validiteitsbewijs van de MCTQ in een veterinaire context?

In de verschillende deelonderzoeken wordt ingegaan op de verschillende aspecten van het brede begrip validiteitsbewijs. Om de hoofdvraag te beantwoorden worden vier deelvragen geformuleerd:

- 1. Draagt een docentprofessionaliseringsprogramma bij aan het gepercipieerde gevoel van competentie in verschillende onderwijsrollen? (Hoofdstuk 2)**
- 2. Wat is het validiteitsbewijs van de MCTQ in een veterinair curriculum? (Hoofdstukken 3 en 4)**
- 3. Zet gefaciliteerde feedback, gegenereerd door de MCTQ, docenten aan tot het reflecteren op hun doceergedrag en stimuleert dit hen tot het formuleren van concrete alternatieven voor hun onderwijspraktijk? (Hoofdstuk 5)**
- 4. Wat is het effect van deze gefaciliteerde MCTQ feedback op de onderwijsprestaties van veterinair klinische docenten? (Hoofdstuk 6)**

In de hoofdstukken 2 t/m 6 worden de onderzoeksvragen middels kwantitatieve en kwalitatieve studies onderzocht. Zowel percepties van docenten en studenten, vragenlijsten en reflectieverslagen worden gebruikt als databronnen.

In **hoofdstuk 2** staan de verschillende rollen die een docent heeft in een modern curriculum en de invloed van een docentprofessionaliseringsprogramma op hun competentiegevoel in deze rollen, centraal. In dit hoofdstuk wordt een raamwerk aangehaald waarin de zeven rollen van een docent in een studentgecentreerd curriculum worden beschreven. Ook maken we in dit hoofdstuk kennis met de landelijk erkende Basis Kwalificatie Onderwijs (BKO). Een programma waarin leren van de onderwijspraktijk, reflectie en training centraal staan. De zeven

docentrollen en hun beschrijvingen werden middels pilotstudies met vijf ervaren docenten gecontroleerd op relevantie en eenduidigheid.

Middels een digitale enquête werd het aangepaste raamwerk gepresenteerd aan de docenten van de Faculteit Diergeneeskunde, waarbij hun gevraagd werd om hun competentieniveau aan te geven voor iedere rol. 194 docenten (77%) werkten mee aan dit onderzoek, waarvan 55% in het bezit was van een BKO. Een stapsgewijze lineaire regressieanalyse liet zien dat er een positieve relatie was tussen het behalen van de BKO en het competentiegevoel van de docenten in zes van de zeven rollen. Deze bevindingen duiden er op dat een docentprofessionaliseringsprogramma een positief effect heeft op de attitudes aangaande het eigen competentiegevoel met betrekking tot onderwijs. Dit effect wordt ook door andere studies in de context van het medisch onderwijs onderschreven.

Deze positieve bevinding gaf draagvlak voor het uitvoeren van de studie beschreven in **hoofdstuk 3**. In deze studie speelt het eerder genoemde model van validiteitsbewijs een centrale rol. Vier aspecten van validiteit met betrekking tot de originele, zeven domeinen tellende, MCTQ worden onderzocht in de veterinaire context. De inhoud van het instrument wordt geborgd door de sterke verankering van het instrument in de literatuur over werkplekleren. Het invulproces van het instrument wordt onderzocht door in pilotgroepen studenten het instrument hardop te laten invullen en te laten discussiëren over de verschillende items van de vragenlijst. Ook de wijze van distribueren wordt voorgelegd aan deze studenten. Aan de hand van deze pilotgroepen wordt het instrument aangepast. Op aanraden van de studenten wordt het item 'uren contacttijd' toegevoegd, omdat zij ervaren dat ze sommige docenten te kort zien om betrouwbare uitspraken middels de MCTQ te kunnen doen. Vervolgens wordt de MCTQ onder alle studenten in het vijfde en zesde studiejaar gedistribueerd als digitale enquête. De psychometrische validiteit en betrouwbaarheid worden in kaart gebracht door de data te analyseren in een confirmatorische factoranalyse en een generaliseerbaarheidsstudie. Het laatste aspect van validiteit dat in dit hoofdstuk de revue passeert, is de relatie tussen de scores op de domeinen van de MCTQ en de items die de contacttijd en de satisfactie met betrekking tot de onderwijsprestaties van de individuele docent bevragen.

De resultaten van de psychometrische analyses laten zien dat een aangepaste MCTQ met vijf domeinen (General Learning Climate, Modelling, Coaching,

SAMENVATTING

Articulation en Exploration) en 15 items valide is. Een acceptabele betrouwbaarheid voor de scores op de domeinen werd gevonden bij 10 tot 12 evaluaties per individuele docent. Zes tot acht evaluaties waren nodig voor een betrouwbaar algeheel tevredenheidcijfer. Een sterke positieve relatie werd gevonden tussen de scores op de vijf domeinen en dit algehele tevredenheidcijfer. Ondanks het feit dat studenten 'uren contacttijd' aandroegen als mogelijk storende factor, vonden we geen sterk verband tussen dit item en de scores op de MCTQ. Deze resultaten dragen allen bij aan het bovengenoemde validiteitsbewijs. De bevindingen duiden er op dat de MCTQ valide en betrouwbaar is voor het gebruik als evaluatie-instrument in een veterinaire klinische omgeving, mits er voldoende oordelen, dat wil zeggen tenminste zes per docent, beschikbaar zijn.

Hoewel het vorige hoofdstuk vele bewijzen van validiteit met betrekking tot de MCTQ behandelde, bleven een aantal aspecten onderbelicht. Zo was nog onbekend wat het aandeel was van verschillen tussen docenten met betrekking tot de scores die ze ontvangen op de MCTQ. In andere woorden: in hoeverre bepaalt het gedrag van de docent zijn of haar score op het instrument? Ook was nog niet in kaart gebracht in hoeverre karakteristieken (zoals ervaring, geslacht) van docenten, studenten en hun interacties invloed hadden op de uitkomsten van de MCTQ. **Hoofdstuk 4** beschrijft een studie waarin middels een multilevel analyse deze in medisch onderwijskundig onderzoek veelvuldig onderbelichte aspecten worden onderzocht.

Data van 1277 ingevulde formulieren over 168 verschillende klinische docenten werden in een tweetal modellen ontbonden in componenten gerelateerd aan verschillen tussen docenten en verschillen tussen studenten. De resultaten wezen uit dat een aanzienlijk deel van de variantie in de domeinscores van General Learning Climate, Modelling en Coaching werd verklaard door verschillen tussen docenten. Dit gold in mindere mate voor de domeinen Articulation en Exploration. Het toevoegen van verklarende variabelen, zoals geslacht, behaalde BKO en departement, veranderde nauwelijks iets aan deze hoeveelheid verklaarde variantie. Dat betekent dat het gedrag van docenten sterk bepalend is voor de scores die ze krijgen op de MCTQ en dan met name op het gebied van leerklimaat, voor doen/rolmodel zijn en coachingsvaardigheden. Verschillen tussen studenten en karakteristieken van docenten en studenten hebben maar weinig invloed op deze uitkomsten. Dat is gunstig, want met een dergelijk instrument wil men vooral de prestaties en gedragingen van een docent in kaart brengen.

Twee aspecten hebben wel invloed op de scores van de docent. Studenten in hun zesde jaar beoordeelden over het algemeen, maar niet consequent, hun docenten beter dan studenten uit het vijfde jaar. Waarschijnlijk ligt dit aan het feit dat de wensen en eisen van studenten in deze twee verschillende fases van de studie wat betreft hun begeleiding verschillen. De resultaten gaven ook een beperkte negatieve relatie tussen jaren ervaring van docenten en de score op de MCTQ aan. Een verklaring kan gevonden worden door het feit dat er geen continue training is van docenten. Hierdoor kunnen, ondanks het toenemen van ervaring, onderwijsprestaties afnemen.

In **hoofdstuk 5** wordt een ander belangrijk aspect van validiteitbewijs geëxploreerd: zet het terugkoppelen van MCTQ feedback aan tot reflectie bij veterinaire klinische docenten? De introductie van deze studie beschrijft dat het enkel voorleggen van feedback aan docenten in veel gevallen niet leidt tot reflectie en prestatieverbetering. Het terugkoppelen van deze feedback middels een verwerkingsstrategie werkt mogelijk beter.

Een groep van 56 veterinaire klinische docenten participeerde in dit onderzoek. Over deze docenten was middels de MCTQ een betrouwbaar studentenoordeel over hun doceerprestaties verzameld. Door middel van gepaarde randomisatie werden deze docenten ingedeeld in twee groepen. Eén groep volgde de standaardstrategie met zelfbeoordeling, een feedbackrapport en een voorgestructureerd reflectierapport. De andere groep volgde dezelfde strategie, maar dan aangevuld met een teaminterview. De reflectierapporten dienden als kwalitatieve databron en werden gescoord op diepte van reflectie. Ook werd per rapport bepaald hoe concreet de voorgestelde verbeterplannen waren. Een elektronische vragenlijst werd ingezet om de opinies van de docenten over de verschillende onderdelen van de trajecten te peilen.

De resultaten wezen uit dat het merendeel van de klinische docenten vormen van reflectie lieten zien na het doorlopen van de standaardstrategie. Deze groep docenten ontwikkelde ook alternatieven voor hun onderwijspraktijk. De groep die het traject inclusief teaminterview had doorlopen liet kritischer vormen van reflectie zien. Ook de voorgestelde alternatieven waren concreter van aard. Dit was een gunstige bevinding, omdat concrete alternatieven meer kans bieden op daadwerkelijke verbetering van onderwijsprestaties. Het positieve effect van de teaminterview werd deels verklaard door de opmerkingen in de evaluatie over de strategieën die afgenomen was middels de elektronische enquête. Volgens de

docenten droeg de teamintervisie sterk bij aan het vertalen van de feedback in concrete stappen die ze konden nemen om de in hun feedbackrapport genoemde aspecten te verbeteren. Het toevoegen van een feedbackstrategie, inclusief teamintervisie, blijkt dus zeer gunstig te zijn voor de verwerking van feedback gegenereerd met de MCTQ. Onduidelijk bleef of deze verwerking in reflectie en concrete verbeterpunten ook zouden leiden tot daadwerkelijk betere onderwijsprestaties.

Hoofdstuk 6 had als doel om het effect van gefaciliteerde feedback op de met de MCTQ meetbare onderwijsprestaties in kaart te brengen. De introductie van dit hoofdstuk zet uiteen dat het evalueren van klinische docenten en het gefaciliteerd terugkoppelen van feedback belangrijk is, maar dat bewijs over de effectiviteit van deze interventies op de onderwijsprestaties vaak niet wordt beschreven. Deze lacune in de literatuur gaf aanleiding tot de volgende onderzoeksvragen. Leidt het terugkoppelen van MCTQ feedback middels een strategie met zelftoetsing, een feedbackrapport en een voorgestructureerd reflectierapport tot betere beoordelingen op de MCTQ? En worden deze beoordelingen nog beter wanneer er een teamintervisie aan het traject wordt toegevoegd?

In dit hoofdstuk worden drie groepen met docenten gevolgd waarover tussen maart en december 2009, zes of meer MCTQ evaluaties waren verzameld. Eén groep kreeg de feedbackstrategie zonder teamintervisie aangeboden, terwijl een andere groep als onderdeel van dit traject wel participeerde in een teamintervisie. Een derde groep fungeerde als controlegroep en kreeg geen feedback of facilitering aangeboden. Vanaf februari 2010 werden een jaar lang weer MCTQ evaluaties verzameld. Middels een regressieanalyse werd bepaald op welke domeinen en items van de MCTQ de standaard feedbackstrategie en de teamintervisie afzonderlijk, een positief effect hadden gehad.

De verschillen tussen de voor- en nameting lieten zien dat de docenten die een feedbacktraject aangeboden hadden gekregen gemiddeld beter scoorden na de interventie. De docenten die geen feedback en facilitering hadden gehad, scoorden gemiddeld lager in de nameting. De regressieanalyse gaf aan dat het volgen van een standaardtraject met zelftoetsing, een feedbackrapport en een voorgestructureerd reflectierapport een significant, positief effect had op de algemene tevredenheidsscore en de domeinscore met betrekking tot het leerklimaat. Tevens werden er significant positieve effecten gevonden voor items binnen het Modelling en Coaching domein. De afzonderlijke teamintervisie had een

significant positieve invloed op enkele items in de Articulation en Exploration domeinen. Uit deze resultaten konden we concluderen dat het teruggeven van gefaciliteerde feedback een positief effect heeft op de onderwijsprestaties van klinische docenten. Een teaminterview helpt daarbij het onderwijs te verbeteren in domeinen waarin een actieve inbreng van de student vereist is.

Hoofdstuk 7 vat de bevindingen van de vorige hoofdstukken samen en presenteert op basis daarvan enkele conclusies. Tevens worden de beperkingen van de studies behandeld en aanbevelingen voor toekomstig onderzoek gegeven. Tot slot worden in dit hoofdstuk adviezen gegeven over de praktische implicaties die voortvloeien uit de conclusies van de beschreven studies.

De vijf studies in dit proefschrift gaven de volgende nieuwe inzichten op het gebied van docentprofessionalisering en dan met name op het gebied van het genereren en terugkoppelen van feedback aan veterinaire klinische docenten middels de Maastricht Clinical Teaching Questionnaire (MCTQ). In de perceptie van docenten heeft docentprofessionalisering, in dit geval in de vorm van het behalen van een Basis Kwalificatie Onderwijs (BKO), een positief effect op de competenties in een breed scala van docentrollen. Voor het evalueren van docentrollen in de klinische context blijkt de MCTQ een valide, betrouwbaar en praktisch inzetbaar instrument te zijn. Een gunstige bevinding van de multilevel analyse was dat verschillen tussen docenten een substantiële invloed hebben op hoe zij worden geëvalueerd met de MCTQ en dat karakteristieken zoals departement of geslacht nauwelijks invloed hebben op deze evaluaties.

Het terugkoppelen van de feedback middels een strategie met zelfbeoordeling, een feedbackrapport en een voorgestructureerde zelfreflectieopdracht leidt tot kritische reflectie en het opstellen van alternatieven voor de eigen onderwijspraktijk. Deze reflecties en alternatieven blijken respectievelijk kritischer en concreter te worden zodra klinische docenten ook participeren in een teaminterview als onderdeel van de feedbackstrategie. Dat deze reflecties en opgestelde alternatieven ook leiden tot verbetering in onderwijsprestaties blijkt uit een laatste studie. Docenten die feedback krijgen teruggekoppeld scoren daarna beter op de MCTQ. Het participeren in een teaminterview blijkt vooral gunstig te zijn voor scores op de domeinen waar een actieve inbreng van studenten vereist is.

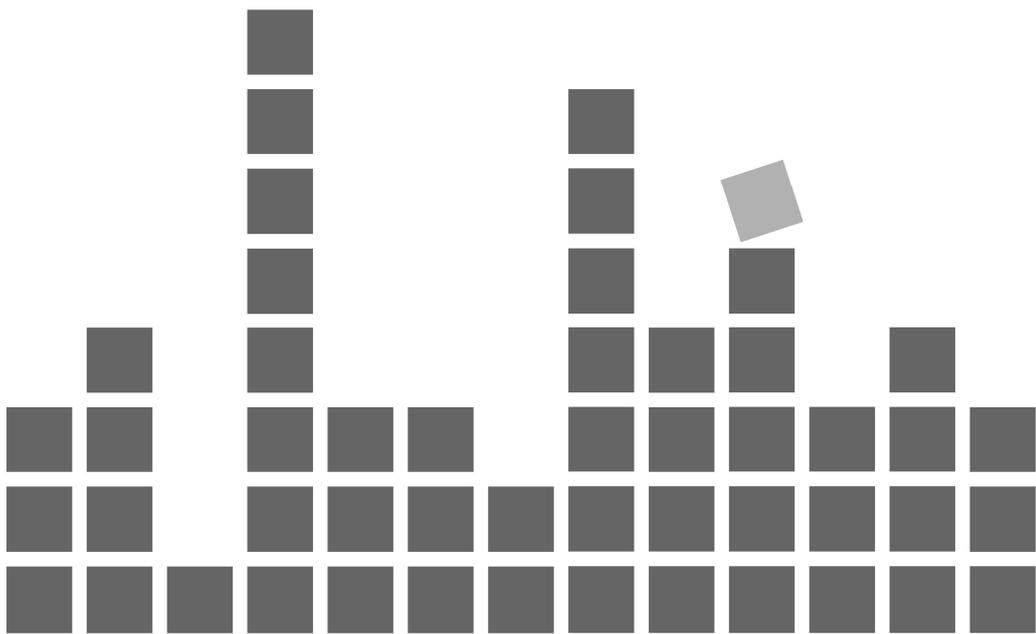
Uit een analyse van de beperkingen van de studies komen een aantal aandachtspunten naar voren. In de eerste studie zijn zelfpercepties van docenten

gebruikt. Dit kan de uitkomsten enigszins vertroebelen. In de overige studies zijn andere, meer objectievere databronnen gebruikt. Daarnaast wordt er in deze paragraaf stilgestaan bij het feit dat in de studies gecombineerde data zijn gebruikt van zowel vijfde- als zesdejaars studenten. Uit de multilevel analyse bleek dat het studiejaar, hetzij beperkt, invloed kon hebben op de MCTQ uitslagen. Ook besteedt dit hoofdstuk aandacht aan de mogelijkheid om meer bronnen dan alleen studentenevaluaties te gebruiken voor het genereren van feedback aan docenten. De potentie van het gebruik van andere bronnen is niet in kaart gebracht binnen dit proefschrift. Een andere beperking van deze studies is dat er weinig aandacht is besteed aan de wijze waarop de klinische context invloed uitoefent op de onderwijsprestaties van klinische docenten.

Dit proefschrift geeft ook aanleiding tot verder onderzoek. Uit de studies blijkt dat het terugkoppelen van studentenevaluaties in veel gevallen leidt tot verbetering van onderwijsprestaties. Het lijkt nuttig om in kaart brengen of het aanvullen van deze feedback met informatie uit andere bronnen, zoals feedback van collega's, aanleiding geeft tot nog betere resultaten. Een ander aspect dat nader onderzoek verdient, is de vraag hoe de MCTQ feedback-reflectiecyclus ingebed dient te worden in andere docentprofessionaliseringsactiviteiten. Het is immers nog onduidelijk hoe de verschillende domeinen van de MCTQ getraind moeten worden en wat er nodig is om goede evaluaties op de MCTQ in stand te houden. Verder onderzoek dient er gedaan te worden naar de verschillen tussen de vijfde- en zesdejaars studenten met betrekking tot hun voorkeuren en waardering voor de verschillende MCTQ domeinen. Dit kan helpen om verschillen in scores van groepen studenten beter te interpreteren. Ook dient er aandacht besteed te worden aan de klinische context. Hoe moet men deze context inrichten zodat de onderwijsstrategieën van de MCTQ zo optimaal mogelijk uitgevoerd kunnen worden?

De uitkomsten van de studies in dit proefschrift hebben ook onderwijskundige implicaties. Het inzetten van de MCTQ als evaluatie-instrument kan leiden tot het verbeteren van onderwijsprestaties als een aantal voorwaarden in acht worden genomen. Zo zijn er voor betrouwbare feedback minimaal zes beoordelingen per docent nodig en dient de feedback teruggekoppeld te worden middels een verwerkingsstrategie. Bij voorkeur aangevuld met een teaminterview. Ook dient er aandacht te worden besteed aan de training van klinische docenten, waarbij de verschillende domeinen van de MCTQ een rol spelen. Als laatste noot wordt de wens uitgesproken dat er in de toekomst aandacht blijft voor het evalueren van klinische docenten zodat zij geholpen worden om (nog) betere docenten te worden.

Dankwoord



DANKWOORD

Dit proefschrift beschrijft een model voor het leren van studenten op de werkplek. Een vorm van cognitive apprenticeship is ook zeker ook van toepassing geweest op mijn eigen aaneenschakeling van leermomenten tijdens dit promotietraject.

Beste Diana, Debbie, Albert en Peter. Als co-promotoren en promotoren hebben jullie altijd gewaakt voor een veilige leeromgeving. Zowel in het bourgondische Maastricht als in het gezellige Utrecht heb ik mij altijd meer dan welkom gevoeld. Nooit ervoer ik een drempel om een vraag te stellen of een stuk voor feedback voor te leggen. Om in de lijn van het model te blijven: jullie waren mijn rolmodel, coach, wetenschappelijk geweten, maar ook diegenen die mij hebben gestimuleerd tot reflectie en verbetering.

Peter, bedankt voor je vertrouwen in dit ongeleide projectiel. Je hebt mij alle ruimte gegeven om te groeien binnen het onderzoek, maar zeker ook als professional. Deuren die op de faculteit in eerste instantie voor mij gesloten waren, wist jij met gezonde pressie toch open te krijgen. Mede hierdoor is het promotietraject zonder enorme tegenslagen of vertragingen tot een goed einde gebracht. Jouw toewijding voor de kwaliteit van het diergeneeskundig onderwijs werkt aanstekelijk en ik geniet van de discussies die we voeren over hoe we het onderwijs nog verder kunnen verbeteren. Jouw steun en betrokkenheid, ook privé, is hartverwarmend geweest en het voelt dan ook als een eer dat ik je tot mijn vriendenkring mag rekenen.

Sjiek! Doen! Prima! Nog nooit heb ik iemand ontmoet die met zulke korte begrippen zoveel kon zeggen. Deze kreten ter goedkeuring gaven mij altijd enorm veel energie. Albert, door jouw expertise, pragmatisme, maar belangrijker nog, enthousiasme, wist je mij telkens weer te inspireren. Ondanks het feit dat je een drukbezet man bent en veel promovendi onder je hoede hebt, nam je altijd veel tijd om mij te helpen. Als een waakhond waakte je over de lijn van het boekje en ik denk dat je daarmee een sleutelrol hebt gespeeld in het welslagen van dit project.

Feedback tot kunst verheffen. Diana is één van de weinige personen die dit kan. Zonder jouw uitvoerige, constructieve en snelle commentaren op mijn werk had ik het nooit gered. Je hebt mij tijdens mijn eerste stapjes in het onderzoek naar medisch onderwijs begeleid en hebt mij, ook toen ik al meer zelfstandig rondstapte in deze wonderlijke wereld, altijd de goede weg gewezen. Je bent een geweldig persoon en ik heb genoten van onze samenwerking.

Debbie, jij hebt een belangrijke rol gespeeld bij een aantal grote gebeurtenissen in mijn leven. Al tijdens mijn onderzoeksstage wist je mij te enthousiasmeren voor de

DANKWOORD

onderwijskunde en dat ben je altijd blijven doen. Van de keuze om te gaan voor dit promotietraject heb ik, mede door jouw fantastische hulp, nooit spijt gehad. Als ervaringsdeskundige wist je mij te behoeden voor de valkuilen van het promoveren. Jouw mening heb ik derhalve altijd zeer belangrijk en waardevol gevonden, evenals onze spar- en brainstormsessies. Uiteraard ben ik je ook enorm dankbaar voor het feit dat je mij op slinkse wijze hebt gekoppeld aan Eva. Dat ik in één jaar zowel mijn huwelijk als promotie mag vieren kun jij absoluut op je conto schrijven.

Weliswaar niet officieel behorend tot de (co-)promotoren, maar minstens net zo belangrijk waren Nicole, Arno en Tim. Ieder op jullie eigen wijze waren jullie mijn psychologisch, statistisch en multilevel geweten en orakel. Ik heb enorm veel van onze samenwerking geleerd en ik wil jullie heel erg bedanken voor jullie bijdragen aan de diverse artikelen in dit proefschrift.

Mereke, door jouw hulp en feedback heb ik in vier jaar tijd mijn 'Stonecole English' kunnen oppoetsen tot een acceptabel niveau. Elke keer weer wist jij mij te verrassen met mooie woorden en fantastische zinsconstructies. Het zijn mooie en krachtige teksten geworden. Daarnaast heb je je ook altijd zeer betrokken getoond. Bedankt!

De onderzoeken in dit proefschrift waren op niets uitgelopen als ik niet de hulp had gehad van alle participerende docenten en studenten. Meermaals heb ik jullie bestookt met pilotgroepen, interviews, e-mails en nieuwsbrieven. Iedere keer was jullie respons overweldigend. Jullie tips en adviezen, ook gegeven buiten de officiële dataverzamelingen om, waren zeer nuttig. De docenten die hebben geparticipeerd in de reflectiestudie ben ik extra dankbaar. Binnen dit onderzoek moesten jullie je zeer kwetsbaar opstellen en dat verdient respect. Een speciaal woord van dank aan de aandachtsveldhouders onderwijs. Jullie steun en hulp hebben zeker bijgedragen aan de totstandkoming van dit proefschrift. Ik hoop dat dit proefschrift op zijn beurt zal bijdragen aan een verdere verbetering van ons klinisch onderwijs, maar zeker ook aan een verhoogde arbeidsvreugde van mijn collega's in de kliniek.

Ook de collega's van de Leerstoel Kwaliteitsbevordering Diergeneeskundig Onderwijs en Onderwijs- en Studentenzaken verdienen een groot woord van dank. Het met jullie mogen samenwerken, heb ik als zeer prettig ervaren. Het is geweldig om onderdeel te zijn van een groep waarin je helemaal jezelf kunt zijn. Waar er ruimte is voor humor, hard werken, mopperen en ontspanning.

Gelukkig was er in onze jonge-honden/AiO kamer ook altijd ruimte voor deze genoegens. Annemarie en Harold, het is geweldig dat jullie naast leuke collega's en 'roomies' ook goede vrienden zijn. Jullie zijn grandioze onderzoekers en ik heb er alle vertrouwen in dat jullie proefschriften fantastisch zullen worden. Ik vind het daarom een grote eer dat jullie mijn paranimfen willen zijn.

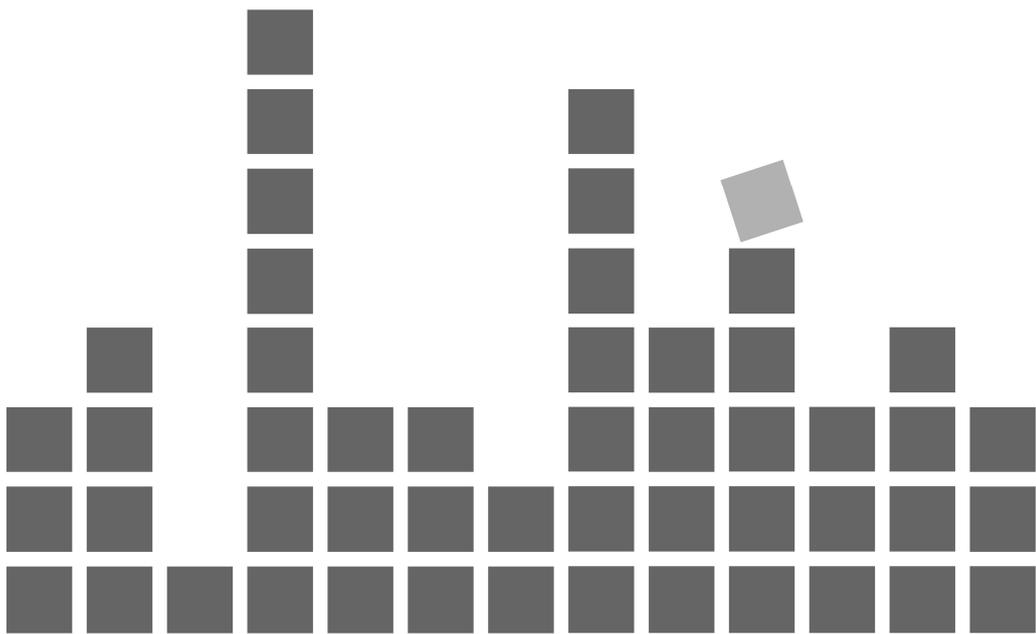
Promoveren is net werken of netwerken. In ieder geval legt men tijdens het traject veel nieuwe en waardevolle contacten. De leden van de journal club van de VU bijvoorbeeld. Ik heb altijd genoten van onze levendige bijeenkomsten en heb veel van jullie mogen leren. Dit geldt ook voor de collega-onderzoekers van het UMCU en O&O Maastricht. Met name Renée, 'the founding mother of cognitive apprenticeship', verdient een extra woord van dank. Beste Renée, bedankt voor al je hulp. Ik hoop dat jouw MCTQ op veel plekken gaat bijdragen aan (nog) beter onderwijs.

Tot slot wil ik mijn familie en vrienden bedanken en dan met name Frank, Berna, opa Bernard, maar ook Onno, Angelique, Dirk en Jaap plus aanhang. Jullie steun, vriendschap en warmte zijn ontzettend belangrijk voor mij. Het was soms moeilijk te duiden waar ik nu precies mee bezig was, maar jullie stonden altijd achter mijn beslissingen. Pa en ma, jullie hebben er tot op de dag van vandaag voor gezorgd dat ik beschikking heb over voldoende basis en mogelijkheden om te bereiken wat ik ambieer. Het maximale uit jezelf halen, hard werken, maar ook genieten. Dat heb ik van jullie geleerd.

Lieve Eva, er zijn niet genoeg woorden om je te bedanken voor jouw onvoorwaardelijke steun en liefde. Ik ben blij dat mijn promotietraject ons indirect bij elkaar heeft gebracht en ik kijk uit naar een lange en gelukkige toekomst samen. Ergens om de hoek wachtte inderdaad geluk.

Tobias

Curriculum Vitae



CURRICULUM VITAE

Tobias Bertus Bernard Boerboom was born on the 20th of January 1981 in Zevenaar, the Netherlands. He attended 'Gymnasium' at the Liemers College in Zevenaar where he graduated in 1999. In the same year he started his veterinary training at Utrecht University, the Netherlands. As a student, Tobias was active in various student committees. In 2004, he won the Intervet-price for his efforts in these committees. It was during his research internship at the 'Chair Quality Improvement in Veterinary Education' where his interest for academic education really started. Tobias graduated 'met genoegen' in December 2005.

Directly after graduation, he started working as a project assistant at the Royal Dutch Veterinary Association (KNMvD). In 2006 he started working as a researcher at the same chair where he conducted his research internship at the Faculty of Veterinary Medicine of the Utrecht University. Tobias combined writing his dissertation with teaching tasks, educational development projects, faculty development initiatives and course coordination. In 2010 he obtained the nationally recognized University Teaching Qualification. He has done a number of consultancies on topics ranging from e-learning to faculty development. Tobias has published papers in various peer reviewed journals and has given presentations at several national and international conferences on (veterinary) medical education.

In his spare time, Tobias likes to design and build websites. He also likes to tinker with his vintage automobile. Tobias lives in Utrecht and is married to Eva.