

# The Maastricht Clinical Teaching Questionnaire (MCTQ) as a Valid and Reliable Instrument for the Evaluation of Clinical Teachers

Renée E. Stalmeijer, MSc, Diana H.J.M. Dolmans, PhD, Ineke H. A. P. Wolfhagen, PhD, Arno M. M. Muijtjens, PhD, and Albert J. J. A. Scherpbier, MD, PhD

## Abstract

### Purpose

Clinical teaching's importance in the medical curriculum has led to increased interest in its evaluation. Instruments for evaluating clinical teaching must be theory based, reliable, and valid. The Maastricht Clinical Teaching Questionnaire (MCTQ), based on the theoretical constructs of cognitive apprenticeship, elicits evaluations of individual clinical teachers' performance at the workplace. The authors investigated its construct validity and reliability, and they used the underlying factors to test a causal model representing effective clinical teaching.

### Method

Between March 2007 and December 2008, the authors asked students who

had completed clerkship rotations in different departments of two teaching hospitals to use the MCTQ to evaluate their clinical teachers. To establish construct validity, the authors performed a confirmatory factor analysis of the evaluation data, and they estimated reliability by calculating the generalizability coefficient and standard error measurement. Finally, to test a model of the factors, they fitted a structural linear model to the data.

### Results

Confirmatory factor analysis yielded a five-factor model which fit the data well. Generalizability studies indicated that 7 to 10 student ratings can produce

reliable ratings of individual teachers. The hypothesized structural linear model underlined the central roles played by modeling and coaching (mediated by articulation).

### Conclusions

The MCTQ is a valid and reliable evaluation instrument, thereby demonstrating the usefulness of the cognitive apprenticeship concept for clinical teaching during clerkships. Furthermore, a valuable model of clinical teaching emerged, highlighting modeling, coaching, and stimulating students' articulation and exploration as crucial to effective teaching at the clinical workplace.

The importance of clerkships in the medical curriculum has given rise to the development of several evaluation instruments to measure the quality of the clinical teaching of medical students at the workplace (i.e., the ambulatory or inpatient clinic). Although most instruments used today have clear strengths, they also have weaknesses. Some lack sound underpinning theories of effective clinical teaching; some developed without the input of crucial stakeholders; and some include items that are too broadly defined, limiting their value for evaluating individual teachers.

Please see the end of this article for information about the authors.

Correspondence should be addressed to Ms. Stalmeijer, Maastricht University, Faculty of Health, Medicine, and Life Sciences, Department of Educational Development and Research, Universiteitssingel 60, PO Box 616, 6200 MD Maastricht, the Netherlands; telephone: (+31) 43-388-57-35; fax: (+31) 43-388-57-79; e-mail: r.stalmeijer@maastrichtuniversity.nl.

*Acad Med.* 2010;85:1732-1738.  
First published online September 28, 2010  
doi: 10.1097/ACM.0b013e3181f554d6

Two of the most cited instruments in the medical education literature, the Stanford List<sup>1</sup> and the Cleveland Clinical Teaching Effectiveness Instrument (CCTEI),<sup>2</sup> have both strengths and weaknesses. The strength of the CCTEI lies in the involvement of stakeholders in its design process; however, the lack of clearly specified theoretical dimensions could hamper feedback effectiveness.<sup>3</sup> The Stanford List, on the other hand, has a clear theoretical basis, but it focuses on a broad collection of teaching arrangements that reflect teaching effectiveness for different teaching settings, including organized small-group sessions.<sup>1</sup> This broad focus makes the instrument less suitable for individualized feedback for physicians teaching at the clinical workplace.

The theoretical constructs of cognitive apprenticeship<sup>4</sup> underpin the Maastricht Clinical Teaching Questionnaire (MCTQ), the aim of which is to provide individual clinical teachers with feedback about their teaching skills with regard to supervising medical students rotating

through clerkships at the workplace. The appeal of cognitive apprenticeship for clinical teaching resides in its aim to teach and make explicit the often-tacit processes involved in experts' handling of complex cognitive tasks. Based on apprentice-type learning and teaching methods, cognitive apprenticeship principles advocate "learning through guided experience."<sup>4</sup> At its center are several teaching methods (modeling, coaching, scaffolding, encouraging articulation, encouraging exploration, and encouraging reflection) that clinical teachers use both to externalize the tacit processes underlying their thinking and actions in practice and to model their expert strategies. In cognitive apprenticeship, teaching starts with modeling by a teacher who explicitly demonstrates a task (how to perform a physical exam, for example) and acts as a role model for students, explaining certain elements of the task. In the next step, coaching, the teacher observes students performing a task and gives them feedback. While modeling and coaching, the teacher should be aware of

the level of knowledge and skills that their students have already attained and, based on this level, decide whether and when to provide additional guidance. This method is known as scaffolding—providing support to the level of the student and gradually fading that support as the student progresses. To access students' problem-solving strategies, the clinical teacher encourages articulation, stimulating students to externalize knowledge and skills. A clinical teacher should also stimulate both reflection, which helps students become aware of their strengths and weaknesses, and exploration, which means that he or she encourages autonomy in students by asking them to formulate and pursue their own personal learning goals.<sup>4</sup> Because of the established importance of generating a safe learning environment to promote clinical teaching and learning,<sup>5</sup> we have added this element to the MCTQ on top of the cognitive apprenticeship teaching methods. In summary, cognitive apprenticeship (with the addition of safe learning environment) distinguishes among teaching methods that are strongly facilitated by teachers (modeling, creating a safe learning environment), that are aimed at stimulating interactions between teacher and student (coaching and scaffolding), and that are aimed at stimulating self-regulated learning by students (articulation, reflection, exploration).

Previous research has already established the value and content validity of the MCTQ for the undergraduate clinical teaching setting. Focus-group research with senior medical students established that the teaching methods were observable and viable during clerkship rotations.<sup>6</sup> Further, stakeholders including physicians, educationalists, and medical students rated the relevance of the MCTQ items as high.<sup>7</sup> In this current study, we will address the MCTQ's construct validity and reliability and investigate how its factors relate to one another.

Additionally, on the basis of research underlining the great importance of (role) modeling,<sup>8</sup> we hypothesized that the first tasks of a good clinical teacher are to provide modeling and a safe learning environment. We further hypothesized that after establishing a secure learning environment and being a good model, a teacher should interact

with students, giving feedback and providing support (coaching and scaffolding), and make sure that students actively engage in clinical practice.<sup>9,10</sup> Finally, we hypothesized that providing an appropriate level of autonomy within the learning environment is beneficial to the students' learning experience<sup>11</sup>; thus, teachers should encourage students to self-regulate their learning through articulation, reflection, and exploration. We tested how well this hypothesized model fit our data.

The current study focuses on three research questions:

1. What is the construct validity of the MCTQ?
2. How many student ratings of one teacher are required for the ratings to be reliable?
3. How are the different factors underlying the MCTQ related to one another and to the overall MCTQ score?

## Method

### Setting

We conducted this study with students undertaking clerkships in outpatient clinics and on the wards of two teaching hospitals affiliated with Maastricht Medical School. Clinical clerkships occur in years 4 and 5 of the six-year curriculum and consist of nine hospital-based rotations of varying duration (5–10 weeks) in internal medicine, surgery, pediatrics, obstetrics–gynecology, neurology, dermatology, ENT, ophthalmology, and psychiatry.

### Participants

Between March 2007 and December 2008, we asked all students engaged in their clerkship rotations at one of the two teaching hospitals to complete, at the end of each rotation, a maximum of three MCTQs for any of the clinical teachers with whom they had had the most contact at the workplace. Participation was voluntary, and students did not receive an incentive for participating. After the initial request at the end of each rotation, we sent no further requests to fill out MCTQ forms. Students needed only about five minutes to fill out an MCTQ for an individual clinical teacher. Students completed the MCTQ anonymously.

### Steps taken to protect human participants

Participating students were all responsible adults who spoke Dutch and rotated through one of the indicated clerkships between March 2007 and December 2008. Likewise, clinical faculty were employed between March 2007 and December 2008 by one of the two teaching hospitals where the indicated clerkships took place, and, as mentioned above, participating students completed forms about faculty members based on the amount of time they spent with each faculty.

Students were recruited noncoercively and participated only after, first, a full explanation of the study goals and procedures and, second, an opportunity to ask questions. Additionally, students responded anonymously to the questionnaires so that neither we the researchers, nor the clinical faculty of the department, knew their identities.

We notified clinical faculty in the department where the study occurred of our intention to conduct the study. We obtained verbal consent from the clinical faculty only after we informed them of the study goals and procedures and gave them an opportunity to ask questions about the study. We did not make any data collected about individual faculty members available to the public.

Participation was voluntary for both faculty and students, and we made it clear that there would be no repercussions for not participating or for withdrawing from the study at any given point. Further, we had no professional or personal relationship with any of the participants.

We saved the collected data on a password-protected computer in a secured data warehouse in our department. Nobody could unintentionally get access to the data. The names of the files did not link them in any way to the participants. Only the research team saw the data on individual clinical faculty members and their performance, and the team maintained absolute confidentiality.

We were well aware of potential risks to both students and clinical faculty and therefore put due procedures in place to reduce these risks to an absolute minimum.

## Instrument

We developed the MCTQ based on the previously described principles of cognitive apprenticeship<sup>4</sup> and on content validity established through focus groups with and/or surveys of three groups of stakeholders (clinical teachers, educationalists, and senior students).<sup>6,7</sup> The 24 items of the first version of the instrument (i.e., the complete version; some items were removed after analysis) represented cognitive apprenticeship (modeling, coaching, scaffolding, stimulating articulation, stimulating reflection, stimulating exploration) and establishing a safe learning environment. The items were statements that students scored on a Likert scale (1 = fully disagree, 5 = fully agree).<sup>7</sup> We also asked students to give an overall judgment of each clinical teacher's individual teaching performance at the workplace (a mark out of 10 where below 6 is insufficient) and to provide written comments on these faculty members' strengths and weaknesses.

## Analysis

Because the MCTQ aims at evaluating the performance of an individual clinical teacher at the workplace, we first aggregated the data for those teachers who had received four or more individual student ratings by computing mean scores across students per individual teacher. We determined the construct validity of the MCTQ by confirmatory factor analysis (CFA) using Amos 7.0 (SPSS, Chicago, Illinois).<sup>12</sup> First, we checked the normality of the distribution by calculating skewness (tilt in the distribution) and kurtosis (peakedness of the distribution). The skew and kurtosis values of all the data we used were smaller than  $\pm 1.5$ , or even  $\pm 1.0$ , implying that they were normally distributed; therefore, we could use a maximum likelihood estimation to conduct the CFA. We used the Amos program to determine whether the data confirmed the theoretical model. We used the following fit indices and criteria:

1.  $\chi^2$  divided by the degrees of freedom (CMIN/df) is  $< 2$ ;
2. the goodness-of-fit index (GFI) is  $> 0.90$ ;
3. the comparative fit index (CFI) is  $> 0.90$ ;
4. the root mean square residual (RMSEA) is  $< 0.1$ ; and
5. the PCLOSE value is  $> 0.5$ .<sup>13</sup>

Table 1

### Results for the Confirmatory Factor Analysis of the Maastricht Clinical Teaching Questionnaire With One-, Two-, Three-, Four-, and Five-Factor Solutions\*

No. of factors	$\chi^2$	df	P	CMIN/df	GFI	CFI	RMSEA	PCLOSE
1	474.6	77	.00	6.16	0.58	0.76	0.20	0.00
2	264.3	76	.00	3.48	0.72	0.89	0.14	0.00
3	221.8	74	.00	3.00	0.78	0.91	0.13	0.00
4	157.6	71	.00	2.22	0.83	0.95	0.10	0.00
5	73.0	67	.30	1.09	0.92	1.00	0.03	0.85

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers. Results indicate that the five-factor solution has the best fit to the data. The following fit indices and criteria were used: (1)  $\chi^2$  divided by the degrees of freedom (CMIN/df) is  $< 2$ ; (2) the goodness-of-fit index (GFI) is  $> 0.90$ ; (3) the comparative fit index (CFI) is  $> 0.90$ ; (4) the root mean square residual (RMSEA) is  $< 0.1$ ; and (5) the PCLOSE value is  $> 0.5$ .

Additionally, we used SPSS 15 to calculate correlations between the factor scores and the overall judgment.

We also used SPSS 15 to determine the generalizability (G-coefficient) of the ratings by estimating the number of student ratings required for a reliable rating per individual teacher. For the subset of 126 teachers who received an overall judgment from four students or more, we calculated G-coefficients for both the overall judgment of teaching and for each factor. This design allowed variance-component estimation of two sources: (1) differences between teachers (T) (object of measurement) and (2) differences between students nested within teachers and general error (S: T, e).<sup>14</sup> For example, students 1 and 2 give teacher A, respectively, a score of 7 and 8 (mean = 7.5); students 3 and 4 give teacher B, respectively, a score of 4 and 6 (mean = 5). So the difference between teachers amounts to 2.5 (7.5 - 5), while the differences between students nested within teachers amount to 7 - 7.5, 8 - 7.5, 6 - 5, and 4 - 5, hence -0.5, +0.5, +1, and -1. For acceptable reliability, a G-coefficient of at least 0.70 is necessary.<sup>15</sup> We calculated Cronbach alphas to indicate the reliability of each scale (internal consistency).

Finally, we tested a model comprising all the factors and the overall judgment by fitting a linear structural model to the data using the structural equation modeling program Amos 7.0, with fit indices and criteria identical to those listed above for the CFA.

## Results

### Response

We collected a total of 1,315 questionnaires on 291 faculty members completed by fourth- and fifth-year medical students. In all, 126 physicians received four or more anonymous ratings.

### Construct validity

Because CFA demonstrated a suboptimal fit of all cognitive apprenticeship principles as described in the Method section, we removed from the instrument items that (1) showed possible overlap in wording and meaning and/or (2) were suggested for removal based on the Amos 7.0-generated modification indices (which indicate the items that could be removed to achieve a better fit of the model). Subsequently, we generated alternative, more parsimonious models representing cognitive apprenticeship, and we subjected those models to stepwise testing. The results (Table 1) demonstrated that a five-factor model with 14 items (Appendix 1) provided an excellent fit with

- CMIN/df 1.09,
- GFI 0.92,
- CFI 1.0,
- RMSEA 0.03, and
- PCLOSE 0.85.

Because the correlations between the factors were quite high, varying between 0.57 and 0.87 (Table 2), we also tested a one-, two-, three-, and four-factor model based on the correlation data and theoretical assumptions. The five-factor

Table 2

**Correlations Between the Five Factor Scores of the Maastricht Clinical Teaching Questionnaire and the Overall Performance Indicator\***

Factor	N	Modeling	Coaching	Articulation	Exploration	Learning environment	Overall judgment
Modeling	126	1.0					
Coaching	126	0.76 <sup>†</sup>	1.0				
Articulation	126	0.66 <sup>†</sup>	0.75 <sup>†</sup>	1.0			
Exploration	126	0.69 <sup>†</sup>	0.61 <sup>†</sup>	0.74 <sup>†</sup>	1.0		
Learning environment	126	0.82 <sup>†</sup>	0.74 <sup>†</sup>	0.57 <sup>†</sup>	0.59 <sup>†</sup>	1.0	
Overall judgment	126	0.87 <sup>†</sup>	0.82 <sup>†</sup>	0.76 <sup>†</sup>	0.72 <sup>†</sup>	0.87 <sup>†</sup>	1.0

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers.

† Correlation is significant at the 0.01 level (two tailed).

model yielded a better fit (Table 1). We cross-validated the five-factor model by dividing the dataset into two random subsets (63 teachers each). The results demonstrated an acceptable fit for both subsets (Table 3).

**Reliability**

The results of the generalizability studies demonstrated that the variance associated with teachers for the overall judgment is 0.33. The variance associated with students within teachers varies per factor between 0.44 and 0.74 (Table 4).

Table 5 provides the G-coefficients per factor as a function of the number of student responses. To obtain a reliable G-coefficient of at least 0.70 or higher, at least seven student responses are necessary for the overall judgment. Four factors (i.e., modeling, coaching, stimulating articulation, and establishing a safe learning environment) required 8 to 10 ratings, and stimulating exploration required 14 ratings (Table 5).

Alpha-coefficients for all factors (0.83–0.96; Appendix 1) indicated high internal consistency of all the factors.

**Linear relationships**

Testing of the proposed five-factor model of clinical teaching, based on the CFA, resulted in a good fit to the data:

- CMIN/df 1.34,
- GFI 0.90,
- CFI 1.0,
- RMSEA 0.05, and
- PCLOSE 0.43.

Path coefficients show that modeling, learning environment, and articulation significantly affect the overall judgment of clinical teaching (Figure 1). Furthermore, modeling plays an important role as it substantially affects coaching, which in turn clearly affects articulation, and articulation substantially impacts exploration. Although learning environment does not significantly affect coaching, it had a direct effect on overall judgment. Coaching and exploration do not seem to directly affect overall judgment, but coaching does have an effect on overall

judgment through the mediating variable articulation.

**Discussion and Conclusions**

First, we tested the construct validity of the MCTQ as an instrument to elicit students' evaluations of the teaching quality of an individual clinical teacher. Based on the teaching methods as suggested by cognitive apprenticeship,<sup>4</sup> and on the concept of a safe learning environment,<sup>5</sup> the CFA yielded a five-factor model with an excellent fit. Moreover, all five factor scores correlate well with the overall judgment (Table 2), which also lends support to the validity of the MCTQ.

Besides validity, the results also confirmed the reliability of the MCTQ. G-coefficients showed that seven ratings suffice for a reliable overall judgment of the clinical teaching performance at the workplace of an individual clinical teacher. As for the reliability of the individual factors, we need 8 to 10 ratings to reach reliability with modeling, coaching, stimulating articulation, and establishing a safe learning environment, but stimulating exploration needs at least 14 ratings. An explanation for this latter finding may be that exploration consists of only two items (Appendix 1). Nevertheless, we value exploration as an indispensable element of the MCTQ, not only because CFA revealed that it is a strong, individual factor but also because the international literature highly commends exploration as a learning activity<sup>16</sup> and because research leading up to the design of the MCTQ showed that three groups of stakeholders deemed the two current exploration items to be highly relevant.<sup>7</sup>

To investigate the mutual impact of the MCTQ factors and their impact on the

Table 3

**Results for the Confirmatory Factor Analysis of the Maastricht Clinical Teaching Questionnaire of the Five-Factor Solution: The Total Dataset and Two Random Subsets\***

Dataset	N	$\chi^2$	df	P	CMIN/df	GFI	CFI	RMSEA	PCLOSE
Total	126	73.0	67	.30	1.09	0.92	1.00	0.03	0.85
Subset 1	63	94.0	67	.02	1.40	0.83	0.97	0.08	0.11
Subset 2	63	89.1	67	.04	1.33	0.83	0.98	0.07	0.18

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers. The authors cross-validated the five-factor model by dividing the dataset into two random subsets (63 teachers each). The results demonstrated an acceptable fit for both subsets. The following fit indices and criteria were used: (1)  $\chi^2$  divided by the degrees of freedom (CMIN/df) is <2; (2) the goodness-of-fit index (GFI) is >0.90; (3) the comparative fit index (CFI) is >0.90; (4) the root mean square residual (RMSEA) is <0.1; and (5) the PCLOSE value is >0.5.

Table 4

**Variance Components of the Maastricht Clinical Teaching Questionnaire for Each Factor and for the Overall Judgment\***

Factor	Teacher (T)	Student: Teacher (S:Te)
Modeling	0.12	0.44
Coaching	0.11	0.44
Articulation	0.12	0.46
Exploration	0.12	0.74
Learning environment	0.14	0.48
Overall judgment	0.33	0.88

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers. The table represents the interaction between the variance associates with the effect of the teacher (T) and the teacher: student interaction (S:Te).

overall judgment of clinical teaching performance at the workplace, we fitted a structural linear model to the data. Our hypotheses were confirmed because modeling and a safe learning environment seemed to be prerequisite for effective clinical teaching. Previous research corroborates the strength and importance of (role) modeling, by showing that the single most powerful predictor of students' satisfaction with clinical teaching was the effort by a clinical teacher to make his/her own clinical reasoning transparent to students.<sup>17</sup> The importance of a safe learning environment and its profound effect on what students learn was previously established by another study.<sup>5</sup> In our model, modeling impacts coaching. Through modeling, a clinical teacher makes the tacit processes underlying this expertise explicit so that in the next step (coaching) students can observe, enact, and practice these processes with guidance from the teacher.<sup>4</sup> These coaching activities

underline the relevance of the process that Leinster<sup>9</sup> described as engagement, which, he claims, fosters a more meaningful learning experience—or, in Leinster's own words, "clinical exposure is necessary for clinical learning, but not enough." Coaching can stimulate students to engage in articulation and exploration. In our model, articulation seems to be crucial for exploration, which encourages learner autonomy. These findings (i.e., the relationships between coaching and articulation, and between coaching and exploration) emphasize the value of feedback and of tailoring teaching to individual students.<sup>18</sup> Finally, modeling, a safe learning environment, and articulation all determine the overall judgment of clinical teaching performance at the workplace of an individual teacher, indicating that, from the students' perspective, these three teaching methods are crucial.

Notably, the majority of the path coefficients among the five factors are

rather high, whereas those explaining the overall judgment are generally low (Figure 1); however, this does not imply that the factors poorly predict the overall judgment. Each of the factors does have quite some power to predict the overall judgment as evidenced by the correlations shown in the last row of Table 2, which range from 0.72 to 0.87. Table 2 also shows that the intercorrelations of the factors generally are somewhat smaller (0.57–0.82), but still of considerable size. Because the predictors of the overall score in Figure 1 strongly correlate, their contributions may show relatively low path coefficients while the total explained variance is still high as indicated by the high correlations in the last row of Table 2. The overall judgment is a holistic rating of the teacher's performance. As such, we would expect it to correlate with each of the factors, but the path coefficients in Figure 1 show how the factors simultaneously contribute to the overall judgment.

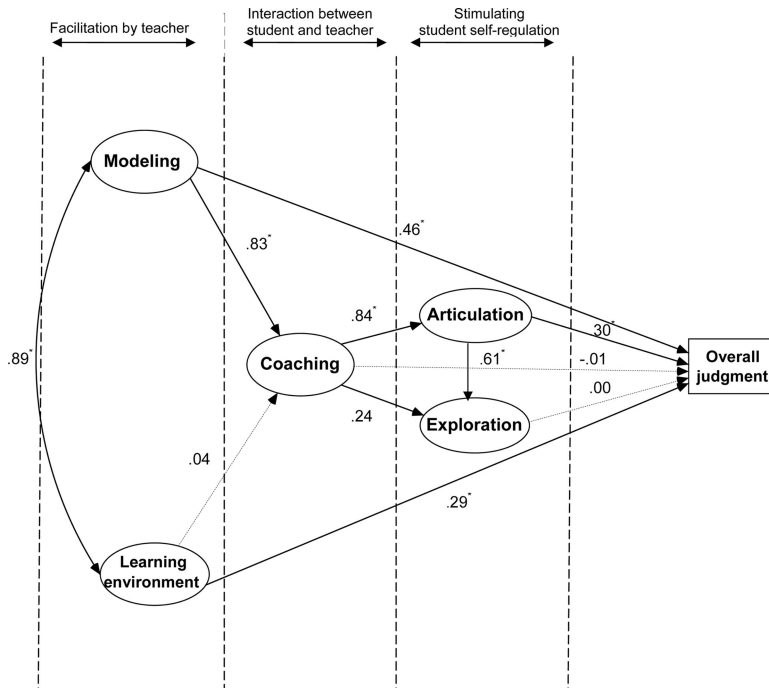
A possible limitation of this study lies in the chosen procedure for data collection. We asked students to complete the MCTQ for a maximum of three clinical teachers with whom their contact had been most extensive. This strategy could have resulted in the evaluation of a select group of physicians—namely, those physicians who more naturally engage in clinical teaching—and the relative neglect of physicians who avoid clinical teaching. Further research could focus on this issue. Furthermore, to investigate whether higher teacher performance scores on the MCTQ result in better student learning would be interesting. Finally, we should mention that the

Table 5

**Generalizability (G-Coefficients) per Number of Student Ratings (N) of the Maastricht Clinical Teaching Questionnaire for Each Factor and for the Overall Judgment\***

No. of ratings	G-coefficients					
	Modeling	Coaching	Articulation	Exploration	Learning environment	Overall judgment
6	0.61	0.59	0.61	0.50	0.64	0.69
8	0.68	0.66	0.67	0.57	0.70	0.75
10	0.73	0.71	0.72	0.62	0.75	0.79
12	0.76	0.75	0.75	0.66	0.78	0.82
14	0.79	0.77	0.78	0.70	0.81	0.84

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers. For acceptable reliability a G-coefficient of at least 0.70 is required. Results indicate that at least 7 student ratings are necessary for a reliable overall judgment and between 8 and 14 ratings for a reliable judgment per factor of the Maastricht Clinical Teaching Questionnaire.



**Figure 1** Linear model of clinical teaching explaining the relationships among the different teaching methods belonging to the Maastricht Clinical Teaching Questionnaire. Path coefficients (standard regression coefficients) show that modeling, learning environment, and articulation significantly affect the overall judgment of clinical teaching. Furthermore, modeling plays an important role as it substantially affects coaching, which in turn clearly affects articulation, and articulation substantially impacts exploration. Although learning environment does not significantly affect coaching, it had a direct effect on overall judgment. Coaching and exploration do not seem to directly affect overall judgment, but coaching does have an effect on overall judgment through the mediating variable articulation.

\*The standard regression coefficient is significant at the .01 level (also indicated by solid lines).

structural model is a simplified and linear model of how clinical teachers and medical students interact with each other; in reality, different variables will influence one another to a greater or lesser extent, and some paths could be recursive. Therefore, we will continue to investigate the model to see how it “behaves in reality” in different clinical workplace settings.

The current study supports the validity and reliability of the MCTQ as an instrument for evaluating the teaching skills of individual clinical teachers, provided that evaluators can obtain judgments from a minimum of 7 to 10 students per individual teacher. In addition, the results of this study seem to present a model that can give direction to effective sequencing of teaching methods. Teaching behaviors aimed at modeling, coaching, and stimulating articulation and exploration in students are crucial to the overall teaching effectiveness of clinical teachers during workplace learning. By presenting this kind of

information, the MCTQ model can provide individual clinical teachers with feedback about their teaching at the workplace during clerkships.

*Acknowledgments:* The authors wish to thank Mereke Gorsira for editing the final version of the manuscript, and Karen Vreeswijk for her assistance with the data collection.

*Funding/Support:* None.

*Other disclosures:* None.

*Ethical approval:* Participating students were all responsible adults who spoke Dutch and rotated through one of the eight participating clerkships at one of two teaching hospitals between March 2007 and December 2008, and clinical faculty taught in one of the clerkships at one of the hospitals during the same time period. Students selected faculty for evaluation based on the amount of contact they had with any one of the participating faculty members.

Students were recruited noncoercively and participated only after, first, a full explanation of the study goals and procedures and, second, an opportunity to ask questions. Additionally, students responded anonymously to the questionnaires, so neither the researchers nor the department’s clinical

faculty knew their identities. Likewise, the researchers notified clinical faculty in the departments where the study occurred of their intention to conduct the study, and they obtained verbal consent from clinical faculty only after both explaining study goals and procedures and providing an opportunity for faculty members to ask questions. Participation for both students and clinical faculty was voluntary, and students and faculty members understood that there would be no repercussions if they chose not to participate or if they decided to withdraw from the study at any given point. The researchers made none of the data collected about faculty available publicly. They maintained the data on a password-protected computer, stored on a secured data warehouse in their department. Nobody could unintentionally get access to the data. The names of the files did not link them in any way to the participants. Only the researchers, who maintained absolute confidentiality, saw the data on the performance ratings of clinical faculty.

*Previous presentations:* A shorter version of this manuscript has been presented at the American Educational Research Association conference in Denver, Colorado, April 30 to May 4, 2010.

**Ms. Stalmeijer** is researcher and educationalist, Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands.

**Dr. Dolmans** is associate professor and educational psychologist, Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands.

**Dr. Wolfhagen** is associate professor and educational psychologist, Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands.

**Dr. Muijtjens** is associate professor, statistician and methodologist, Department of Educational Development and Research, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands.

**Dr. Scherpbier** is professor of medical education and scientific director, Institute for Medical Education, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, the Netherlands.

## References

- 1 Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med.* 1998;73:688–695.
- 2 Copeland HL, Hewson MG. Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical centre. *Acad Med.* 2000;75:161–166.
- 3 Bowden J, Marton F. Quality and qualities. In: Bowden J, Marton F, eds. *The University of Learning: Beyond Quality and Competence in Higher Education.* London, UK: Kogan Page Limited; 1998:211–245.
- 4 Collins A, Brown JS, Newman SE. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In: Resnick LB, ed. *Knowing, Learning, and*

- Instruction: Essays in Honor of Robert Glaser. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1989:453–494.
- 5 Kilminster SM, Jolly BC. Effective supervision in clinical practice settings: A literature review. *Med Educ.* 2000;34:827–840.
  - 6 Stalmeijer RE, Dolmans DH, Wolfhagen IH, Scherpbier AJ. Cognitive apprenticeship in clinical practice: Can it stimulate learning in the opinion of students? *Adv Health Sci Educ Theory Pract.* 2009;14:535–546.
  - 7 Stalmeijer RE, Dolmans DH, Wolfhagen IH, Muijtjens AM, Scherpbier AJ. The development of an instrument for evaluating clinical teachers: Involving stakeholders to determine content validity. *Med Teach.* 2008;30:e272–e277.
  - 8 Elzubeir MA, Rizk DE. Identifying characteristics that students, interns and residents look for in their role models. *Med Educ.* 2001;35:272–277.
  - 9 Leinster S. Learning in the clinical environment. *Med Teach.* 2009;31:79–81.
  - 10 Dornan T, Boshuizen H, King N, Scherpbier A. Experience-based learning: A model linking the processes and outcomes of medical students' workplace learning. *Med Educ.* 2007;41:84–91.
  - 11 Beckman TJ, Lee MC. Proposal for a collaborative approach to clinical teaching. *Mayo Clin Proc.* 2009;84:339–344.
  - 12 Arbuckle JL. *Amos 7.0 User's Guide.* Chicago, Ill: SPSS Inc.; 2006.
  - 13 Byrne BM. *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming.* Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
  - 14 Shavelson RJ, Webb NM. *Generalizability Theory: A Primer.* London, UK: Sage; 1991.
  - 15 Grönlund NE. *How to Construct Achievement Tests.* 4th ed. Englewood Cliffs, NJ: Prentice-Hall; 1988.
  - 16 Kaufman DM. ABC of learning and teaching in medicine: Applying educational theory in practice. *BMJ.* 2003;326:213–216.
  - 17 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;19:766–771.
  - 18 Branch WT Jr, Paranjape A. Feedback and reflection: Teaching methods for clinical settings. *Acad Med.* 2002;77:1185–1188.

## Appendix 1

### Mean Score (Scale 1 = Fully Disagree, 5 = Fully Agree) and Corresponding Standard Deviation (SD) per Item of the Maastricht Clinical Teaching Questionnaire and for the Overall Judgment (Scale 1–10)\*

Questionnaire item—The clinical teacher ...	Mean (1–5)	SD	Alpha-coefficient ( $\alpha$ )
<b>Modeling</b>			<b>.86</b>
Consistently demonstrated how to perform clinical skills.	3.9	0.45	
Created sufficient opportunities for me to observe him/her.	4.1	0.45	
Served as a role model as to the kind of doctor I would like to become.	3.9	0.57	
<b>Coaching</b>			<b>.83</b>
Gave useful feedback during or immediately after direct observation of my patient encounters.	3.8	0.56	
Adjusted his/her teaching activities to my level of experience.	3.9	0.42	
Offered me sufficient opportunities to perform activities independently.	4.2	0.50	
<b>Articulation</b>			<b>.89</b>
Asked me to provide a rationale for my actions.	3.8	0.47	
Asked me questions aimed at increasing my understanding.	4.0	0.50	
Stimulated me to explore my strengths and weaknesses.	3.7	0.46	
<b>Exploration</b>			<b>.94</b>
Encouraged me to formulate learning goals.	3.5	0.52	
Encouraged me to pursue my learning goals.	3.5	0.54	
<b>Safe learning environment</b>			<b>.96</b>
Created a safe learning environment.	4.3	0.46	
Was genuinely interested in me as a student.	4.3	0.56	
Showed that he/she respected me.	4.4	0.46	
<b>Overall judgment of clinical teaching (scale 1–10)</b>	<b>7.8</b>	<b>0.71</b>	

\* The dataset consisted of four or more anonymous ratings for 126 individual clinical teachers.