*Artikel*

*Högre utbildning*
Vol. 1, Nr. 2 December 2011, 139-150

# Uniform Evaluation of Clinical Teaching - an Instrument for Specific Feedback and Cross Comparison Between Departments

Pernilla Kihlberg[a], Mattias Perzon[a], Rolf Gedeborg[b], Per Blomqvist[a] and Jakob Johansson[c,*]
[a]*Medical Student Council, Uppsala University;* [b]*Uppsala Clinical Research Center, Uppsala University;* [c]*Department of Surgical Sciences - Anaesthesiology & Intensive Care, Uppsala University*

The quality of clinical teaching in Swedish medical schools has recently been questioned by student representatives at the medical schools in Sweden. The clinical teaching at Uppsala University Medical School has previously only been sparsely evaluated and with varying instruments. Our aim was to develop a simple assessment instrument and implement it uniformly, enabling cross comparisons between different departments. Further we wanted to investigate students' response rate and analyze their response pattern.

A questionnaire inspired by a concept from Stanford University was created and introduced at hospital departments performing clinical teaching of medical students. Ten questions reflecting different aspects of clinical teaching were used. The outcome was used for specific feedback to the clinical teachers, cross comparison and official ranking of the departments.

The evaluation instrument has been successfully developed and uniformly implemented throughout the departments of the hospital. The students have with a relatively high response rate graded the items with a great deal of variation between and within the clinical rotations. Altogether, this has provided specific feedback to the clinical teachers and enabled cross comparison between departments. Our belief is that the instrument also applies to other educations involving supervision.

*Keywords:* Education, medical, undergraduate, questionnaires, evaluation, clinical teaching

## INTRODUCTION

Student internship constitutes a major part of many different educations. In medical schools in Sweden, the students spend the majority of their three final years of undergraduate medical education in several different clinical rotations. Medical students are supposed to be involved in the daily practice of the clinical working environment where they should be able to train clinical skills. Successful learning depends on active involvement of students and it has been demonstrated that good supervision is the key to effective learning in clinical practice (Dolmans, Wolfhagen, Essed, Scherpbier & van der Vleuten, 2002).

During recent years, the number of admitted students to medical schools in Sweden has increased, which puts even higher demands on the learning environment in the hospitals than before. In addition, health care often has limited resources, doctors experience increasing time pressure and a decreasing number of patients are admitted to hospitals (Spencer, 2003). All this may have a negative influence on the learning environment. The quality of the clinical

* Författarkontakt: jakob.johansson@surgsci.uu.se

ISSN 2000-7558

http://www.hogreutbildning.se

Tidskriften tillämpar kollegial granskning för bidrag av typen "Artikel". Övriga bidrag granskas redaktionellt. För mer information hänvisas till http://hogreutbildning.se/page/om-tidskriften

teaching has recently been questioned by student representatives at the medical schools in Sweden (Fredriksson et al., 2008). Similar concerns about variable standards in training of medical students have been raised in both the United Kingdom and in the United States (Chitsabesan, Corbett, Walker, Spencer & Barton, 2006).

According to the pedagogical program of Uppsala University (2008), all teaching has to be evaluated and analyzed with regard to quality. At the Medical School, this is the responsibility of the Program Committee, which has collaborated in this matter with student representatives and the Unit for Medical Education. Earlier evaluations have primarily been focusing on teaching events such as lectures, labs and small group events. The results from these evaluations have been compiled by student representatives and been reported back to the Program Committee, the responsible teachers and the concerned students. In contrast, the clinical teaching has previously only been sparsely evaluated and with varying instruments among the different clinical rotations. Also, earlier assessments of the clinical teaching have mainly featured high-inference questions such as global ratings of the quality of the clinical teaching. Furthermore, a system to aggregate and report the results back to the clinical teachers and students has been missing. The importance of getting to know the results from earlier evaluations has been emphasized both by teachers and students, but it is also known that this feedback too seldom is provided (Hedin, 2006).

Performed evaluations can be of great importance when working with educational development. One fundament is to not only ascertain that there is a problem, but to thoroughly analyze what the problem is (Elmgren & Henriksson, 2010). When designing an evaluation, a common way is to focus on three major aspects: presage, process and outcome. Presage is e.g. the students' previous knowledge and expectations before a course or a clinical rotation, while the process can be the teachers' performance, the content of the course, the learning climate and the examination. The outcome can be measured in different ways, e.g. the reaching of goals, change in interest and motivation for the field concerned (Elmgren & Henriksson, 2010). The literature does not support the superiority of one existing evaluation instrument over any other (Beckman, Ghosh, Cook, Erwin & Mandrekar, 2004b). Yet, the instrument used to perform the evaluation has to be considered. In order to reflect many aspects of the clinical setting, assessment instruments should cover several domains of clinical teaching (Beckman et al., 2004b). Different institutions have their own cultures of teaching, and evaluations should be consistent with the philosophy of the institutions in which they are used (Snell et al., 2000). The evaluation instrument should also have a solid foundation in theories of effective learning (Beckman et al., 2004b). The teaching at Uppsala University Medical School reflects the view of constructivism through the use of problem-based learning throughout the whole curriculum and the great amount of learning in the context of the clinical working environment. Constructivism is a learning theory that implies that students are encouraged to construct their own knowledge in realistic situations together with others (Elmgren & Henriksson, 2010). Further, Uppsala University Hospital uses the Stanford Faculty Development Program (Skeff, Stratos & Bergen, 1992) to develop the teaching skills of the clinical teachers. The domains in this program are consistent with and have been inspired by several learning theories, among others the ones of Gagne (Gagne, 1985) and Rogers (Rogers, 1951).

Regarding the subject of evaluation, several described evaluations focus on the individual teacher (McOwen, Bellini, Morrisson, & Shea, 2009; Stalmeijer, Dolmans, Wolfhagen, Muijtjens & Scherpbier, 2008; Williams, Litzelman, Babbott, Lubitz & Hofer, 2002). It has been estimated that at least eight evaluations are needed to produce a reproducible score (Hayward, Williams, Gruppen & Rosenbaum, 1995). At the Medical School at Uppsala University the clinical teachers

seldom supervise enough students to achieve such a number of evaluations, which in turn may render more ambiguous results. In view of this, one might consider focusing the evaluation on a whole department's teaching performances. Such an approach would also allow cross comparison of different departments, making it easier to learn from each other and share experiences, a fundament in educational development (Gibbs, 1992).

Our aim was to develop a new assessment instrument for evaluation of the clinical teaching, inspired by the well-recognized Stanford Faculty Development Program, and reflecting several different aspects of clinical teaching. The outcome was assessed by the students' response rate and their response pattern for the evaluated items. The purpose of this paper is to describe that process, the implementation of the new instrument and to present the outcome.

## METHOD

### *Setting*

Uppsala University Medical School in Sweden currently admits approximately 100 new students each semester. The curriculum comprises eleven semesters and is since 2006 based on problem-based learning. From the fifth semester, students spend a great majority of their time in clinical rotations at Uppsala University Hospital or at one of the affiliated hospitals in the nearby region. During one semester a student completes between two and five different rotations. The clinical teachers on the clinical rotations are the regular staff of doctors working in the concerned department, which in addition to their normal clinical duties is responsible for the clinical teaching of the students scheduled at their department.

### *Developing the evaluation instrument*

In order to improve the clinical teaching, the Program Committee of the Medical School formed a working group in October 2008. The group consisted of the Dean of Medical Education together with representatives from the hospital board, university teachers, clinical teachers and medical students. In December 2008, the group decided to develop and implement a new evaluation instrument. They reached consensus that

- The instrument should be based on a valid educational theory
- All items assessed should be on a low-inference level
- The instrument should be concise
- The evaluation should be uniformly implemented at all clinical rotations, allowing cross comparison between departments
- The outcome of the evaluation should be official and available both to teachers and students

A questionnaire was subsequently created in line with these ambitions. It comprised ten items to be rated on a six point Likert-type scale reflecting different aspects of clinical teaching. The grading included behavioral anchors representing the lowest and highest score on the Likert scale (Table 1). It also contained an open-ended question at the end where students in their own words could comment on the clinical rotation. The ten questions were inspired by a validated evaluation tool with high internal consistency from the Stanford Faculty Development Program Litzelman, Stratos, Marriot & Skeff, 1998) (Table 1). This program has been thoroughly described elsewhere (Skeff et al., 1992) and has been successfully transferred

to Uppsala University (Johansson, Skeff, & Stratos, 2009). Since 2005, the Stanford program constitutes an essential part of the program for faculty development at Uppsala University hospital. Between two and four courses are given annually to the clinical teachers. In brief, this program is built on the following domains of teaching: learning climate, control of session, communication of goals, promotion of understanding and retention, evaluation, feedback and promotion of self-directed learning. These domains also reflect the three important aspects of evaluations mentioned earlier: presage, process and outcome (Elmgren & Henriksson, 2010). The ten questions derived out of these domains were then thoroughly reviewed by all relevant stakeholders: the Program Committee of the Medical School, representatives from the hospital board and the Unit for Medical Education. Finally, medical students from different clinical rotations reviewed the questionnaire in order to detect ambiguity of the intention of the questions.

Table 1. The ten questions and the behavioral anchors in the evaluation instrument. The right column indicates which teaching domains of the Stanford Faculty Development Program the question relates to.

| | Questions included in the evaluation questionnaire | Behavioral anchors (scale 1-6) | Teaching domains* |
|---|---|---|---|
| 1 | What was the introduction and the welcoming at the department like? | 1 = Poor<br>6 = Very good | LC |
| 2 | To what extent did you know what you were expected to learn during your rotation (both theoretical and practical skills)? | 1 = I had no idea<br>6 = I knew exactly | CG |
| 3 | Were your supervisors familiar with the course curri culum, concerning theory and practical skills? | 1 = They had no idea<br>6 = They knew | CG |
| 4 | To what extent was your rotation well planned (tasks during the day, schedule, framing, etc.)? | 1 = Not at all<br>6 = To great extent | CS |
| 5 | Did your supervisor/s take an interest in you, assessing your knowledge and skills? | 1 = No, not at all<br>6 = Yes, indeed | LC, EV |
| 6 | Did the rotation stimulate to further learning in the field concerned? | 1 = No, not at all<br>6 = Yes, indeed | SDL |
| 7 | Did you handle patients on your own (e.g. by leading rounds, suggesting further investigations/diagnoses/ treatments)? | 1 = Not at all<br>6 = To great extent | CS, UR |
| 8 | What was the teaching of practical skills like? | 1 = Very poor[1]<br>6 = Very good[2] | UR |
| 9 | How often did you get feedback during your rotation (e.g. patient record notes, practical skills, etc.)? | 1 = Never<br>2 = Always | FB |
| 10 | Did staff members, other than doctors, take an interest in your training? | 1 = No, not at all<br>6 = Yes, indeed | LC, UR |

* Teaching domains: LC: Learning Climate; CG: Communication of Goals; CS Control of Session; EV: Evaluation; UR: Understanding and Retention; SDL: Self Directed Learning; FB: Feedback

1: I have not been supervised at all/have not been allowed to try

2: My supervisor/s has/have shown and instructed me and I have been allowed to try by my self

The web-based questionnaire was distributed through a system designed by Uppsala University. Students had access to the questionnaire from the beginning of the semester in September 2009 and the questionnaire could be submitted anytime until one week after the semester was finished in January 2010.

## IMPLEMENTATION

In order to specifically supervise the implementation of the new curriculum, which is based on problem-based learning and was introduced in 2006, only students studying according to the new curriculum were included in this project. During the fall semester of 2009, the new curriculum covered four classes of medical students having reached their fifth, sixth, seventh, and eighth semester in the program. These four classes encountered 21 different departments during their clinical rotations scheduled for the study period which all accepted to participate in the evaluation. To stimulate the students' participation, information was spread in multiple ways. Firstly, representatives from the working group visited class lectures for all the students included, personally informing them about the evaluation and thereby encouraging participation. The main message of this information was to emphasize the importance of providing feedback in order to influence the clinical teaching henceforth. Secondly, advertisements were placarded on strategical sites around campus. Thirdly, after a completed clinical rotation the students of that rotation received an e-mail reminding them of the evaluation. As an additional encouragement for completing an evaluation questionnaire, the students had the opportunity to voluntarily participate in a lottery. Students could among other things win an auscultation at the intensive care helicopter of the hospital.

In order to enthuse not only the students but also the clinical teachers and to further emphasize the importance of clinical teaching, an official competition among the participating departments was launched. To be included in the competition, a department had to acquire a minimum of ten submitted questionnaires in one semester. The results of the questionnaire were used to rank the different departments based on the mean rating of the ten items. The ranking list and the ratings of the separate items were published on the hospital website shortly after the end of the semester. The winning department was awarded by the hospital director at an official ceremony and a picture of the event was published on the hospital website. The directors of the different departments and the teachers involved received an e-mail with their overall result, the separate ratings of each of the ten items and the specific comments from the open-ended question. Each department also got the overall result plus the rating for each item for all the other departments, in order to enable comparison of their relative strengths and weaknesses. After the results had been published and delivered as mentioned above, student representatives from the working group visited the participating departments to discuss their results in order to initiate dialogues about clinical teaching and possible ways of improving it.

### *Statistics*

The response rate was calculated on the basis of submitted questionnaires (numerator) and the students' estimated total number of rotations (denominator). Differences in response rate between the four classes included were compared using the Chi-squared test for trend in proportions. The average score of the ten items for the participating departments are presented with mean and range and the overall ratings for the different semesters are presented with mean (± SD). The distribution of the average score for the ten items within each clinical rotation is presented

as mean (range). The difference between the highest and lowest rating was calculated for each admitted questionnaire (i.e. values between 0 and 5) and is presented as mean (± SD). One-way analysis of variance (not assuming equal variances) was performed to detect differences between means of each class.

*Results*

Of 876 estimated possible questionnaires 577 were submitted, which gave an overall response rate of 66 %. Analyzing the classes separately, decreasing response rates were observed with the number of completed semesters ($p < 0.001$) (Table 2). The median percentage of completed questionnaires for each department was 71 % (interquartile range 57 % - 79 %). In 58 % of the questionnaires the students made comments on the open-ended question (further analyses of these comments were not performed within this study). Out of 21 participating departments, four that had less than ten submitted questionnaires were included only in calculating response rate but were excluded from further analyses.

Table 2. The response rate and mean rating (±SD) for each class.

| Class | Response rate | Mean rating | SD |
|---|---|---|---|
| 5 | 146/169 (86 %) | 3.91 | 1.02 |
| 6 | 169/236 (72 %) | 4.35 | 0.97 |
| 7 | 113/178 (63 %) | 4.15 | 1.07 |
| 8 | 149/293 (51 %) | 4.61 | 1.10 |

The mean ratings (±SD) are calculated on all submitted questionnaires within each class for departments with ≥ 10 submitted questionnaires (in total n=17). The response rates are calculated for all participating departments (in total n=21). The students rated the items on a six-point scale.

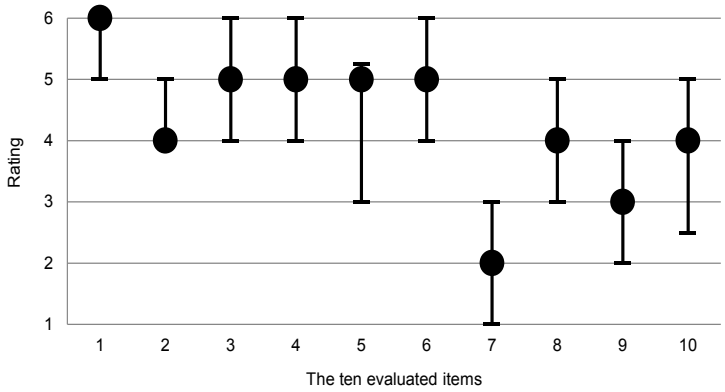


*Figure 1. An example of the variation of ratings for the different items for one department. The circle represents the median and the whiskers represent the first and third quartiles. The item number refers to the ten questions in table 1.*

Among the remaining 17 departments, there was a substantial difference between the departments with regards to their overall result. The overall mean rating on all the ten items for the departments ranged between 3.36 and 5.54 (mean 4.23). There was some variation between

departments (Table 3), but more apparently quite substantial variation between the ratings of each of the ten questions within each department was noticed (Figure 1). The mean difference between the highest and lowest rated item for each department was 1.71 (range 0.79 - 3.13). Introduction and welcoming was generally rated high, while the item regarding handling patients independently was generally rated low (Table 3).

In order to investigate the students' response pattern regarding the use of the six-graded scale, analysis of the variation within each individual evaluation was performed. It showed that the students varied their graded response within the same questionnaire to a great extent. The mean difference between the highest and lowest rated item within all the submitted questionnaires was 2.94 (± 1.28). When analyzing the classes separately, there was a slight difference between the four classes ranging between 2.75 to 3.23 ($p < 0.01$).

Table 3. The overall mean ratings and range for the evaulated items of the participating departments.

| | Questions included in the evaluation questionnaire | Mean rating | Range |
|---|---|---|---|
| 1 | What was the introduction and the welcoming at the department like? | 4.93 | 4.09 – 5.81 |
| 2 | To what extent did you know what you were expected to learn during your rotation (both theoretical and practical skills)? | 4.26 | 3.35 – 5.18 |
| 3 | Were your supervisors familiar with the course curri culum, concerning theory and practical skills? | 4.12 | 2.82 – 5.21 |
| 4 | To what extent was your rotation well planned (tasks during the day, schedule, framing, etc.)? | 4.38 | 2.82 – 5.55 |
| 5 | Did your supervisor/s take an interest in you, assessing your knowledge and skills? | 4.29 | 3.34 – 5.63 |
| 6 | Did the rotation stimulate to further learning in the field concerned? | 4.68 | 3.54 – 5.79 |
| 7 | Did you handle patients on your own (e.g. by leading rounds, suggesting further investigations/diagnoses/ treatments)? | 3.71 | 2.21 – 5.83 |
| 8 | What was the teaching of practical skills like? | 4.17 | 2.96 – 5.50 |
| 9 | How often did you get feedback during your rotation (e.g. patient record notes, practical skills, etc.)? | 4.01 | 2.65 – 5.71 |
| 10 | Did staff members, other than doctors, take an interest in your training? | 3.74 | 2.82 – 5.45 |

The range represents the highest and the lowest rated department for each question.

## DISCUSSION

We developed a new assessment instrument, inspired by the Stanford Faculty Development Program, in order to evaluate the clinical teaching. This instrument was well received when implemented and has consequently been adopted by all clinical rotations at Uppsala University. The satisfying response rate achieved indicates that the instrument has rendered acceptance also among students. The response pattern indicates that the instrument can reflect different

components of quality in clinical teaching between and within the clinical rotations. Altogether, this has provided detailed and constructive feedback to the clinical teachers and enabled cross comparison between departments.

In general, the results showed that introduction and welcoming was rated highest among the ten questions. Stimulation to further learning was also rated high, which constitutes an important outcome parameter in educational evaluations (Elmgren & Henriksson, 2010). The item related to independent handling of patients was generally rated low, which is consistent with earlier findings that students spend most time during clinical rotation observing doctors (Hell, Kuks & Cohen-Schotanus, 2009) and that medical students not often enough get the opportunity to take a leading role in patient care (The survey of Medicine Studerandes Förbund, 2008). Feedback was also rated relatively low in spite of it being recognized as important by both teachers and student. This has also been well described previously (Hedin, 2006).

A noteworthy source of error when performing evaluations is the "halo effect", which is the tendency to give similar ratings across different items, rather than distinguishing among the items. This phenomenon has recently been demonstrated in evaluation of clerkship rotations (McLaughlin, Vitale, Coderre, Violato & Wright, 2009). In our study, the variation in mean ratings for the ten separate items within the same department indicates that students have assessed the items fairly separately. Thus, even if a department rated high in one item, students have rated other items low for the same department. The broad dispersion within the individual questionnaires also demonstrates that the students have graded the different items with a great deal of variation. The low-inference level items assessed in the instrument together with the wide range of ratings rendered specific and diversified feedback. This identifies potential teaching strengths and weaknesses of the different departments and can assist in the making of decisions concerning clinical teaching strategies.

When analyzing the different semesters separately we observed similar dispersion for all semesters. No substantial halo effect was detected in the student ratings. Neither did we observe any obvious differences between the mean overall ratings for the four different classes. Previously this has not been thoroughly investigated on undergraduate level. Studies performed at graduate level have shown that less advanced learners give generally higher mean ratings when assessing clinical teaching (Beckman, Lee & Mandrekar, 2004a).

We consider that we achieved a relatively high response rate, but in view of the extensive information campaign targeted toward the students we had expected it to be even higher. A fact that lowered our response rate was the four excluded departments, which represented the last rotations for the students on the eighth semester. These rotations were part of a course covering two semesters. This might have lead to a misunderstanding among the students, who may have thought that these clinical rotations were supposed to be evaluated the following semester, rendering a low response rate for these rotations. Another possible explanation to the obtained response rate could be that students experience a burden of numerous evaluations (Morrisson, 2003), sometimes in our case adding up to five clinical evaluations during one semester in addition to the evaluations for the theoretical parts. Another important aspect of the students' motivation to evaluate is that the information rendered from evaluations must be valued and used (Elmgren & Henriksson, 2010). The results from earlier evaluations and the intervention based on them may not have been fully communicated back to the students, which in turn may have contributed to an evaluation fatigue. We observed that the response rate decreased with the number of completed semesters, which is consistent with earlier findings where seniors are the

least likely to respond (Layne, DeCristoforo & McGinty, 1999). Studies have also shown that the introduction of web-based evaluation methods has led to lower response rates. However the lower response rates did not appear to affect mean evaluation scores (Avery, Bryant, Mathios, Kang & Bell, 2006). A higher response rate might therefore not have substantially influenced our evaluation ratings.

The main purpose of evaluating the clinical teaching was to provide specific feedback to the teachers in order to initiate a process for improvement of the clinical teaching (Elmgren & Henriksson, 2010). Exchange of ideas and experience between colleagues is an important part of all educational development (Gibbs, 1992). A possibility to cross compare between departments is one way to initiate such a dialogue. In addition, the competition between the departments was launched with the intention to create a greater focus on the clinical teaching among teachers. Some authors suggest that reported student ratings of teachers have little impact, but do offer a point of reference and help define standards (Stalmeijer, Dolmans, Wolfhagen, Peters, van Coppenolle & Scherpbier, 2009). At the same time, it has been demonstrated that when evaluation results are made public, the faculty members with the lowest ratings show the biggest improvements (Maker, Lewis & Donnelly, 2005). The use of an official ranking list together with a competition may be controversial. Our evaluation covers many types of departments with different clinical activities, thus presenting teachers with quite different challenges for how to provide good clinical teaching. This might give some departments a head start to achieve high ratings. Thus, the cross comparison between departments can be perceived by some as unfair. However, the responses from the departments have been very positive and to our knowledge, only one department has questioned the value of calling it a competition. Thus, we still believe that such comparisons and competition on a departmental level mainly have positive effects by increasing attention to clinical teaching, providing an additional incentive to improve teaching quality. However, although the departments have been very enthusiastic we cannot at this point confirm that specific interventions were undertaken to improve the clinical teaching.

*Limitations*

When performing an evaluation, one can question the content and structure of the instrument used. As discussed earlier, there is a variety of factors to consider when selecting an assessment tool. We chose to develop a new instrument inspired by the Stanford Faculty Development Program in accordance with our local conditions. However, although our instrument emanates from a validated evaluation tool with high internal consistency, it has not yet been formally validated, which may be a limitation.

Another limitation in this evaluation is the calculation of the response rate. Since the questionnaire was not directly addressed to each individual student, the total number of possible clinical rotations of all students had to be surveyed in order to calculate the response rate. This was possible to a great extent, but a few minor approximations had to be done.

An additional limitation might be that students have been able to evaluate their rotations throughout the whole semester, meaning that some students have evaluated shortly after a finished rotation and some after several weeks. However, the time between when an event occurs and when it is evaluated only negligibly impacts the final outcome (McOwen et al., 2008). We therefore assume that the variation in elapsed time between the teaching event and the evaluation most likely has no substantial effect on our outcome.

*Future perspectives*

After the first semester we can only conclude that the evaluation instrument seems to be a valuable diagnostic tool for measurement of the quality of the present clinical teaching. We cannot yet draw any conclusions regarding its applicability as a therapeutic tool, i.e. whether it can induce any improvements regarding the quality of the clinical teaching. However, supervision of students has been demonstrated to be improved when clinical teachers are evaluated and provided with feedback (Dolmans et al., 2002). With continuous application of this tool, teachers' initiatives to improve specific teaching aspects can be evaluated. To facilitate this process, the Program Committee has appointed a Director of Studies to perform systematical follow-ups targeting the departments after completed evaluation periods and to be a support in the further development of methods for clinical teaching. To ensure a high response rate we now aim to give students the opportunity to complete the evaluation questionnaire as an integral part of the clinical rotation. Additionally, feedback from already performed evaluations will be more extensively provided to increase student motivation (Hedin, 2006).

With time, the new problem-based curriculum from 2006 will cover additional departments and thereby allow extended cross comparison. Our plan is to include our affiliated hospitals in our nearby region where medical students from Uppsala University do clinical rotations. We have also recently presented this assessment instrument to the heads of the other six medical schools in Sweden. Discussions are ongoing to implement the evaluation instrument at other schools, which eventually may enable uniform comparisons of the clinical teaching between Swedish medical schools.

The medical student of today is the clinical teacher of tomorrow. Yet, the current curriculum of Uppsala University Medical School does not include teaching in pedagogy. However, our assessment instrument encourages the students to reflect over their teachers' skills and to evaluate what characterizes good clinical teaching. We hope that this will contribute to a greater awareness when they step into the mission of becoming clinical teachers themselves.

## CONCLUSION

A simple evaluation instrument to assess clinical teaching during internship for students was successfully developed and uniformly implemented. Additionally, the students have with a relatively high response rate graded the different items with a great deal of variation, differentiating the quality of the aspects involved between and within the clinical rotations. The results provided specific feedback to the clinical teachers and enabled cross comparison between different departments. Our belief is that the instrument applies not only to medical educations, but also to other educations involving supervision during different forms of internship.

## ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Rebecca Ahlstrand for linguistic revision.

## REFERENCES

Avery, R.J., Bryant, W.K., Mathios, A., Kang, H. & Bell D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Economic Education,* 37(1), 21-37.

Beckman, T.J., Lee, M.C. & Mandrekar, J.N. (2004a). A comparison of clinical teaching evaluations by resident and peer physicians. *Medical Teacher,* 26(4), 321–325.

Beckman, T.J., Ghosh, A.K., Cook, D.A., Erwin, P.J. & Mandrekar, J.N. (2004b). How reliable are assessment of clinical teaching? *Journal of General Internal Medicine,* 19(9), 971-977.

Chitsabesan, P, Corbett, S, Walker, L, Spencer, J & Barton, J.R. (2006). Describing clinical teachers' characteristics and behaviours using critical incidents and repertory grids. *Medical Education*, 40(7), 645-653.

Dolmans, D.H.J.M., Wolfhagen, H.A.P., Essed, G.G.M., Scherpbier, A.J.J.A. & van der Vleuten C.P.M. (2002). Students' perceptions of relationships between some educational variables in the out-patient setting. *Medical Education,* 36(8), 735-741.

Elmgren, M. & Henriksson, A-S. (2010). *Universitetspedagogik.* Stockholm: Norstedts.

Fredriksson, J.J., Tyson, Y, Persson, L, Eriksson, J, Moëll, A, Malmin, E, Diallo, A & Wennström, L. (2008). Rusta upp den kliniska handledningen! *Läkartidningen,* 38, 2542.

Gagne, R. (1985). *The conditions of learning* (4th ed.). New York: Holt Rinehart & Winston.

Gibbs, G. (1992). *Improving the quality of student learning: Based on the Improving Student Learning Project funded by the Council for National Academic Awards.* Bristol: Technical and Education Services.

Hayward, R.A., Williams, B.C., Gruppen, L.D. & Rosenbaum, D. (1995). Measuring attending physician performance in a general medicine outpatient clinic. *Journal of General Internal Medicine*, 10(9), 504-510.

Hedin, A. (2006). *Lärande på hög nivå: Idéer från studenter, lärare och pedagogisk forskning som stöd för utveckling av universitetsundervisning.* Uppsala: Uppsala Universitet.

Hell, E.A., Kuks, J.B.M. & Cohen-Schotanus, J. (2009). Time spent on clerkship activities by students in relation to their perceptions of learning environment quality. *Medical Education,* 43(7), 674-679.

Johansson, J, Skeff, K & Stratos, G. (2009). Clinical teaching improvement: The transportability of the Stanford Faculty Development Program. *Medical Teacher,* 31(8), e377-e382.

Layne, B.H., DeCristoforo, J.R. & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education,* 40(2), 221-232.

Litzelman, D.K., Stratos, G,A,. Marriot, D.J., & Skeff, K.M. (1998). Factorial Validation of a widely disseminated educational framework for evaluating clinical teachers. *Academic Medicine*, 73(6), 688-695.

Maker, V.K., Lewis, M.J. & Donnelly, M.B. (2005). Ongoing faculty evaluations: Developmental gain or just more pain? *Current Surgery,* 63(1), 80-84.

McLaughlin, K, Vitale, G, Coderre, S, Violato, C & Wright, B. (2009). Clerkship evaluation – what are we measuring? *Medical Teacher,* 31(2), e36-e39.

McOwen, K.S., Bellini, L.M., Morrisson, G & Shea, J.D. (2009). The development and implementation of a health-system-wide evaluation system for education activities: Build it and they will come. *Academic Medicine,* 84(10), 1352-1359.

Medicine Studerande Förbund. (2008). *Handledarenkät.* Stockholm: Sveriges läkarförbund.

Morrisson, J. (2003). ABC of learning and teaching in medicine. Evaluation. *British Medical Journal*, 326(7385), 385-387.

Uppsala Universitet. (2008). *Pedagogiskt program.* Uppsala: Uppsala Universitet.

Rogers, C.R. (1951). *Client-centered therapy: Its current practice, implications and theory.* Boston: Houghton Mifflin.

Skeff, K, Stratos, G & Bergen, M. (1992). Evalution of a medical faculty development program: A comparison of traditional pre/post and retrospective pre/post self-assessment ratings. *Evaluation & the Health Professions,* 15(3), 350-366.

Snell, L, Tallett, S, Haist, S, Hays, R, Norcini, J, Prince, K, Rothman, A & Rowe, R. (2000). A review of the evaluation of clinical teaching: new perspectives and challenges. *Medical Education,* 34(10), 862-870.

Spencer, J. (2003). ABC of learning and teaching in medicine. Learning and teaching in the clinical environment. *British Medical Journal,* 326(7389), 591-594.

Stalmeijer, R.E., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P., Muijtjens, A.M.M. & Scherpbier, A.J.J.A. (2008). The development of an instrument for evaluating clinical teachers: involving stakeholders to determine content validity. *Medical Teacher,* 30, e272-e277.

Stalmeijer, R.E., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P., Peters, W.G., van Coppenolle, L, Scherpbier, A.J.J.A. (2009). Combined student ratings and self-assessment provide useful feedback for clinical teachers. *Advances in health sciences education*, 15(3), 315-328.

Williams, B.C., Litzelman, D.K., Babbott, S. F., Lubitz, R.M. & Hofer, T.P. (2002). Validation of a global measure of faculty's clinical teaching performance. *Academic Medicine*, 77(2), 177-180.