
ASSESSMENT OF MEDICAL COURSES IN BRAZIL USING STUDENT-COMPLETED QUESTIONNAIRES. IS IT RELIABLE?

Marco Aurélio Monteiro Peluso, Hermano Tavares and Gilberto D'Elia

RHCFAP/3003

PELUSO M A M et al. - Assessment of medical courses in Brazil using student-completed questionnaires. Is it reliable? **Rev. Hosp. Clín. Fac. Med. S. Paulo** 55 (2):55-60, 2000.

SUMMARY: Introduction: Debates about the quality of medical education have become more evident in the recent past, and as a result several different assessment methods have been refined for that purpose. The use of questionnaires filled out by medical students to assess the quality of lectures is one of the most common methods employed in our milieu. However, the reliability of this investigation method has not yet been systematically tested. The authors present the reliability of a specific form applied to the fourth grade medical students during the clinical psychiatry course.

Method: Eighty-one fourth grade medical students were instructed to complete a form immediately after each clinical psychiatry lecture. Thirty-four students (42%) failed to turn in the forms after the final lecture. These students were given an identical form to assess the lectures in a retrospective fashion. The grades given by both groups of students for each performed lecture and the number of students who have graded an unperformed lecture were compared. Statistical significance for both groups was determined by means of the chi-square test ($p < 0.05$).

Results: Eighteen out of the 34 students who filled out the forms retrospectively (53%) rated the unperformed lecture, whereas only 5 out of the 47 students who filled out the forms during the course (11%) did so. This is statistically significant ($p < 0.05$). There was no statistical difference for the grades given to the lectures that were actually performed.

Discussion: The authors concluded the low reliability rate of the retrospective evaluation warrant a continuous assessment method during the course.

DESCRIPTORS: Medical education. Students. Assessment. Quality. Reliability.

A growing interest in education quality in Brazil has become evident since the Ministry of Education implemented the national test a few years ago. The national test aims at assessing the quality of high school and college education in this country. Important issues such as how to improve the education quality as well as the usefulness of the national test have been examined. The public believes that these assessment methods are highly necessary to improve the quality of education. However, the outcome of a test applied to the students has a question-

able value in the assessment of the education rendered to learners. Some authors advocate that the entire education process should be assessed as opposed to appraising its final outcome only. A few large-scale endeavors like the CINAEM – “Comissão Interinstitucional Nacional de Avaliação das Escolas Médicas” have been made to date. On the other hand, smaller en-

deavors evaluating the quality of lectures in a given discipline within an educational institution is a current practice.

The literature lists several investigations in which the quality of the education is assessed by means of direct inquiry's to the students. Teaching methods^{11,12}, skills and attitude of faculty members^{3,4,6,8} and teaching settings^{1,5}, in addition to other issues^{2,7,9,10} have been appraised in these investigations. A number of those investigations report on qualitative evaluations. However, other investigations employ re-

From the Department of Psychiatry, Hospital das Clínicas, Faculty of Medicine, University of São Paulo.

search instruments on which reliability and validity of the outcomes are the main concerns^{1,3,4,6,7,9,12}. These research instruments are too specific (instruments for assessing learning environment^{1,5}, a measure of medical instructional quality in ambulatory settings⁷, a measure of the faculty staff attitude before students creativity⁶, etc...) and therefore, none of them is adequate for general use.

The most commonly used method in our milieu consists of forms with questions that are answered by the students at the end of a specific course of study. These questions gather the students' opinions about the quality of the lectures rendered to them. However, despite the widespread use of these questionnaires, it is not really known how thoughtfully the students answer the questions, thus jeopardizing the original purpose. Additionally, there are no investigations reporting on the reliability of these methods.

The authors report on the reliability of a specific form applied to medical students during a clinical psychiatry course.

MATERIALS AND METHODS

Clinical psychiatry is a discipline taught to fourth grade medical students over 7 weeks. Tutoring is rendered twice a week, one morning from 8 to 12 AM and one afternoon from 2 to 6 PM. Three theoretical lectures a week are given, and the remaining 2-hour period is used for seeing in-patients on the psychiatric floor.

Eighty-one subjects participated in this investigation during the first semester of 1995. Appraisal of the lectures was performed by means of a form handed out to the students during the opening lecture. Students were asked to fill out the assessment form soon after each performed lecture.

Forms that were filled out were handed back to one of the authors soon after the last lecture before the final test. The students who happened to be without a personal copy were asked to fill out a supplementary copy at that time.

The forms (Fig. 1) portrayed each lecture by the title and the lecturer's name (in figure 1 the lecturer's name is not mentioned). There were 5 boxes reading *very good*, *good*, *regular*, *bad*, and *no grade*. It was said to the students that the *no grade* box should be used whenever one of the following was the case: the specific lecture did not occur, the student did not attend the lecture or the student did not want to issue an opinion about the given lecture.

The lecture entitled "Normal Emotional Development in Childhood and Adolescence" was canceled.

The data collected from forms that were filled out prospectively were compared to the forms filled out retrospectively after the last lecture. Individual lecture grading and the number of respondents in each group who rated the unperformed lecture were analyzed.

RESULTS

A total of 81 students participated in the study. Forty-seven students (58%) delivered the filled out forms as initially indicated. Thirty-four students (42 %) filled out a new copy of the forms retrospectively.

Only 5 students who filled out the forms during the course (11%) rated the unperformed lecture, whereas 18 students (53%) who filled out the forms retrospectively did so (Fig. 2). This is statistically significant ($p < 0.05$) according to the chi-square test. There was no statistical difference for the grades given to the lectures that were actually performed.

DISCUSSION

There are several methods for assessing the quality of teaching. One of the most effective methods is when the tutor discusses with the students every step of the teaching-learning process. This process is not feasible in a teaching setting where more than one faculty member teaches or when there are a large number of participant students. A second alternative is to ask the students to record in writing their opinion about the course. This usually results in an undesirable number of blank sheets of paper and just a few sharp points about the course itself. Therefore, the use of structured questionnaires for the purpose of assessing teaching quality seems reasonable.

However, a complete questionnaire-based evaluation given by the students does not necessarily correlate with effectiveness. Most of these questionnaires have never been tested for reliability.

Taking into consideration that the method is devised to evaluate the outcome of a course, a grade given to an unperformed lecture can be considered as a "false-positive". It is well known that a high rate of false-positives reflects low specificity. Therefore, the testing method in question (if its false-negative rate is low) should only be used as a screening method to further assist what subjects should undergo further testing.

Twenty-three out of 81 students (29%) rated 1 unperformed lecture (29% "false-positive" rate). Yet, to make this matter more complex, there is not a more specific test to apply. Should it have been because lack of attention, or motivation, or any other reason, it is possible these 23 students answered the remaining of the questions in a unheeding fashion, thus jeopardizing the results of the method.

It is also possible to extend this hy-

	Very Good	Good	Regular	Bad	No Grade
Psychopathology I					
Psychopathology II					
Psychopathic personalities					
Schizophrenia (clinical picture)					
Schizophrenia (diagnosis and differential diagnosis)					
Suicide					
Alcohol dependence					
Drug addiction					
Delirium					
Dementia					
Mania					
Depression					
Anxiety disorders I					
Anxiety disorders II					
Eating disorders					
Somatoform disorders					
Psychotherapy					
Sexual disorders					
Chronic delusions					
Mental retardation					
Normal and abnormal emotional development in infancy and adolescence					
Normal and abnormal behavior in infancy and adolescence					
Neurosis of infancy and adolescence					
Psychosis of infancy and adolescence					

Figure 1 - Lectures assessment form.

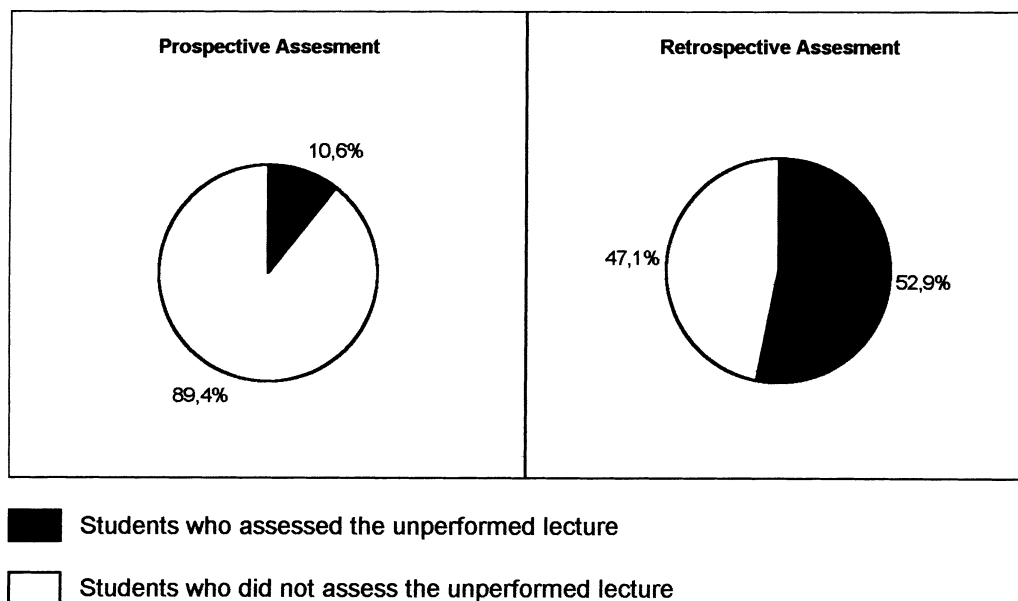


Figure 2 - Assessment rate of unperformed lecture.

pothesis to any course evaluation where a similar method is used. It is likely that reliability affects the validity of the assessment method and if the method aims at reflecting the real quality of the course ministered to the students, the assessing method has to be improved to a higher degree of reliability and validity.

On the other hand, there is a statistically significant difference ($p < 0.05$) between the prospective and retrospective data. The prospective respondents produced a 10% rate of wrong answers whereas the retrospective respondents displayed a more than 50% rate.

Taking into consideration the significance of nearly one-third false-positives, one should question the reliability of the method in reflecting the actual students' opinion. The fact that more than 50% of the students who filled out forms retrospectively rated an unperformed lecture shows a significant problem with retrospective evaluations.

On the other hand, a small percent-

age of students who completed the questionnaire prospectively (11%) also rated the unperformed lecture, showing the prospective evaluations are also not 100% reliable. Nevertheless, the results from prospective evaluations were significantly better than those for the retrospective evaluations.

It is clear to us that this type of evaluation, with this type of form, must be performed in a prospective fashion rather than retrospectively to avoid the risk of results that do not reflect the actual situation.

CONCLUSION

The quest for pedagogical improvement aiming at increasing the efficacy of the learning process calls for a constant evaluation of the teaching methods; therefore, the impressions of the students should definitely be taken into account. However, presented with the opportunity to bring forward their im-

pressions, a considerable number of students do not become involved in this process, denying the tutors access to valuable information for improvement.

Structured questionnaires for assessing a course of study are a viable solution for providing the necessary information. However, if a large percentage of the respondents, either by means of lack of interest or attention, produce low quality information, the final outcome of the assessment effort will be compromised.

This study shows that prospective evaluations are better in quality than the evaluations performed in a retrospective fashion. Therefore, we suggest that an assessment system should be continuously administered during the courses to avoid retrospective data collection. Whenever possible, individual lectures should be assessed by attending students immediately after termination of the lecture, assuring that only the individuals qualified to give an opinion will participate.

RESUMO

RHCFAF/3003

PELUSO M A M e col. – Avaliação da disciplina de psiquiatria clínica na graduação médica: confiabilidade de questionários. **Rev. Hosp. Clín. Fac. Med. S. Paulo** 55 (2):55-60, 2000.

Introdução: Discussões sobre qualidade de ensino têm se tornado cada vez mais freqüentes em nosso meio e métodos variados de avaliação têm sido pesquisados. O uso de questionários, preenchidos por alunos, avaliando a qualidade de aulas ministradas está entre os métodos mais utilizados em nosso meio, no entanto sua confiabilidade não tem sido testada. Os autores apresentam a avaliação da confiabilidade de um destes questionários, o qual foi desenvolvido para um curso de psiqui-

atria clínica ministrada no quarto ano de graduação em medicina.

Método: Os 81 alunos avaliados receberam o questionário no início do curso com a orientação de preenchê-lo após cada aula, no entanto 34 alunos (41,9%) não o devolveram no último dia de aula. Receberam então um novo formulário (idêntico ao primeiro) para que o preenchessem retrospectivamente. Foi comparada, entre os dois grupos, a distribuição das notas para cada aula e a quantidade de alunos que deram nota para a aula não realizada por ausência de professor. A análise estatística utilizou-se do teste de Qui-quadrado (nível de significância $p < 0.05$) para comparar os dois grupos.

Resultados: Entre os questionários preenchidos retrospectivamente a aula

não realizada foi avaliada (com se tivesse ocorrido) por 18 dos 34 alunos (52,9%), enquanto, entre os preenchidos ao longo do curso, 5 dos 47 alunos fizeram o mesmo (10,6%). Esta diferença é estatisticamente significativa. Para as aulas efetivamente realizadas não houve diferença significativa entre as avaliações dos dois grupos.

Discussão: Os autores concluem que a baixa confiabilidade das avaliações feitas respectivamente apontam para a necessidade de um sistema continuado de avaliação ao longo do curso.

DESCRITORES: Ensino médico. Estudantes. Avaliação. Qualidade. Confiabilidade.

REFERENCES

1. CAVANAUGH S & SIMMONS P - Evaluation of a school climate instrument for assessing affective objectives in health professional education. **Eval Health Prof** 1997;20(4):455-478.
2. CHANDRA D, KULSHRESTHA S & CHANDRA M - Weightage of formative examinations in certifying examination of pharmacology: an opinion poll and relative performance of learners. **Med Teach** 1992;14(2-3):197-200.
3. DE GRAVE WS, DOLMANS DH & VAN DER VLEUTEN CP - Tutor intervention profile: reliability and validity. **Med Educ** 1998;32(3):262-268.
4. DOLMANS DH, WOLFHAGEN IH, SCHMIDT HG et al. - A rating scale for tutor evaluation in a problem-based curriculum: validity and reliability. **Med Educ** 1994;28(6):550-558.
5. DUNN SV & BURNETT P - The development of a clinical learning environment scale. **J Adv Nurs** 1995;22(6):1166-1173.
6. GRAVELLE D, LAZURE H & REIDY M - Mesure de l'attitude des professeurs a l'égard de la creativite de l'etudiante-infirmiere (MAPCE). **Can J Nurs Res.** 1991;23(1):41-55.
7. JAMES PA & OSBORNE JW - A measure of medical instructional quality in ambulatory settings: the MedIQ. **Fam Med** 1999; 31(4):263-269.
8. METCALFE DH & MATHARU M - Students' perception of good and bad teaching: report of a critical incident study. **Med Educ** 1995;29(3):193-197.

9. NGANASURIAN WE - Evaluating learning opportunities offered to mental health nursing students. **J Psychiatr Ment Health Nurs** 1998;**5**(5):393-402.
10. PEIRCE AG - Preceptorial students' view of their clinical experience. **J Nurs Educ** 1991;**30**(6):244-250.
11. WAUGH RA, MAYER JW, EWY GA et al. - Multimedia computer-assisted instruction in cardiology. **Arch Intern Med** 1995; **155**(2):197-203.
12. WELK DE - Baccalaureate educator use of environment-based strategies to teach diagnostic patterns. **J Prof Nurs** 1993; **9**(4):228-232.

Received for publication on the 09/02/00