DOCUMENT RESUME

ED 362 523 TM 020 241

AUTHOR Dolmans, Diana; And Others
TITLE Validation of a Rating Scale for Tutor Evaluation in a Problem-Based Medical Curriculum.
PUB DATE Apr 93
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation Methods; Factor Structure; Foreign Countries; Higher Education; *Likert Scales; Medical Education; *Medical Students; Models; Pilot Projects; *Questionnaires; *Student Evaluation of Teacher Performance; Teacher Effectiveness; Test Construction; *Test Validity; *Tutors
IDENTIFIERS Confirmatory Factor Analysis; Netherlands; *Problem Based Learning

ABSTRACT

The construction and validation of a tutor evaluation questionnaire are described. The importance of the tutor in problem-based learning requires assessment of staff conducting this role. After a pilot study in 1990-91 with 100 students and 150 tutors who identified important tutor skills, a questionnaire was prepared. This questionnaire was tested with six courses in the 1991-92 school year, and improvements were made based on the results. The pilot studies resulted in a list of 13 Likert scale statements used in an instrument administered to about 1,079 students in 8 courses (142 tutorial groups) at the Maastricht Medical School (Netherlands) in the 1992-93 academic year. Students indicated whether their tutors demonstrated the behavior described in each statement. Confirmatory factor analysis shows that a 3-factor model comprising 13 items does not fit the data, although skipping 2 items improved the model's fit. Nevertheless, these two items were considered important for content validity. Alpha coefficients indicate that questionnaire reliability is high. Findings otherwise indicate that the students' global and analytical ratings about the functioning of the tutor correspond highly with each other. Some problems involved in administering the questionnaire are discussed. Two tables present study findings. (Contains 11 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
D.H.J.M. DOLMANS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 362 523

Validation of a Rating Scale for Tutor Evaluation in a Problem-Based Medical Curriculum[1]

Diana Dolmans, Ineke Wolfhagen, and Henk Schmidt
Department of Educational Development and Research,
University of Limburg,
Maastricht, The Netherlands

[1] Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, April 1993.
Correspondence concerning this paper should be addressed to Diana Dolmans, Department of Educational Research, University of Limburg, PO Box 616, 6200 MD Maastricht, The Netherlands, Phone: 0-11-31-43-881107, E-mail: Dolmans@Educ.RuLimburg.NL

Introduction

The tutor plays a key role in problem-based curricula. This notion is confirmed by research conducted by Gijselaers and Schmidt (1990). These authors postulated a causal model of problem-based learning (PBL) to identify and measure the effects of important variables, such as group functioning, tutor functioning, quality of the problems presented and time spent on self-study. One of their findings was that tutor functioning has a direct causal influence on the functioning of small-group tutorials which in turn influences students' interest in the subject matter. In addition, it has an indirect causal effect on student achievement. These results indicate the importance of tutors' abilities to guide tutorial groups in an adequate way.

The importance of the tutor in problem-based curricula requires assessment of staff conducting this role. This implies that instruments are required to collect information about the functioning of the tutor. This would enable the school to provide tutors with feedback. In addition, training and remedial teaching should be provided to (prospective) tutors in line with the shortcomings pointed out by the evaluation. At the Medical Faculty of the University of Limburg, Maastricht, The Netherlands, such an attempt to implement a series of those measures has been made. In this paper the construction and validation of a tutor evaluation questionnaire will be described.

Since a tutor evaluation questionnaire should provide teachers with feedback, items should reflect key features of the tutor role. This implies that the questionnaire needs to be based on the tasks set for the tutor at the Medical Faculty of Maastricht, as well as on theoretical conceptions about the tutor role, as described in the literature (Barrows, 1988). The role of a tutor in a problem-based curriculum mainly consists of guiding the tutorial group. The tasks of the tutor are to guide students through all steps of the learning process, to encourage students to attain a deeper level of understanding, to ensure that all students are involved in the group process, to monitor progress of individual students, to motivate students and to help the student group to deal with their own problems of interpersonal dynamics (Barrows, 1988). Consequently, the tutor is a guide and a facilitator and as such an important key aspect of problem-based learning.

Whereas, much descriptive literature is available about the skills a tutor should have (Barrows & Tamblyn, 1980; Barrows, 1988), only few studies have been conducted to identify important features of the tutor role. Most of these studies were concerned with tutor's expertise on subject matter under discussion (Davis, Nairn, Paine, Anderson & Oh, 1992; De Volder, 1982; Feletti, Doyle, Petrovic & Sanson-Fisher, 1982; Moust & Schmidt, 1992). De Volder (1982), for instance, found that tutor's functioning as judged by students was positively related to expertise on subject matter and experience. Feletti et al. (1982) revealed that good tutors were perceived by students as having a thorough up-to-date knowledge of the particular problem being studied and as encouraging them to review their academic progress. Wilker-

son (1992) concluded that two factors described the skills seen as most helpful by both tutors and students: maintaining positive interactions within the group and providing assistance in getting the work of the group accomplished.

Method

Subjects. Data were collected during eight six-week courses in the academic year 1992-1993: two first year courses, two courses in the second year, two courses in the third year and two courses in the fourth year. About 1079 students participated in these eight courses. Both students and tutors were randomly assigned to the tutorial groups. Data were aggregated at the tutorial group level, since the purpose of the study was to develop an instrument to provide tutors with feedback about their functioning. The data included 142 tutorial groups.

Instruments. The construction of the questionnaire consisted of several steps. The first pilot-study was aimed at identifying important tutor skills. In the academic year 1990-1991, 100 students and 150 tutors, divided among the first four study years, received a list of 16 items specifying behavioral characteristics of the tutor. Both students and tutors were asked to rate whether each item was assumed to be an important indicator of a tutor's functioning. Furthermore, they were asked to indicate whether each item was clearly stated. These 16 items were based on descriptions of the tutor role at the Maastricht Medical School, on theoretical conceptions about the role of the tutor (Barrows, 1988) and on program evaluation instruments used at the different faculties of the University of Limburg during the last few years. This pilot-study resulted in skipping five items and rewording six items. As a consequence, the questionnaire contained eleven items reflecting tutors' functioning. In a second pilot-study, in the academic year 1991-1992, this questionnaire was used during six courses. The aim of this pilot-study was to acquire information about the way in which tutors should receive individual feedback. Students were asked to rate the functioning of the tutor on a Likert-scale. Furthermore, students were asked to give an overall judgement (a score ranging between 1-10, 6 was "sufficient") for the functioning of the tutor. It was assumed that these eleven observed variables are affected by three common factors. These factors were related: (1) guiding students through the learning process, (2) content knowledge input and (3) commitment to the group's learning. The results of this second pilot-study were reported to individual tutors. These tutors were asked to comment on the clarity of the items. In general, these tutors considered the feedback as valuable. On the basis of these comments and high percentages of students using the "not applicable" option for a few items, the questionnaire was improved. These pilot-studies resulted in a list of 13 statements, which are affected by the three common factors cited above. The factors with their related items are shown in Table 1.

Table 1 The observed variables and their common factors

Factor 1	Guiding students through the learning process
1	The tutor demonstrates to be well-informed about the process of problem-based learning
2	The tutor stimulates all students to participate actively in the tutorial group process
3	The tutor stimulates a careful analysis of the problems
4	The tutor stimulates the generation of specific learning issues useful for self-study
5	The tutor stimulates an extensive reporting on information collected during self-study
6	The tutor stimulates evaluation of the tutorial group process
Factor 2	Content knowledge input
7	The tutor has an understanding of the subject matter covered in the course
8	The tutor assists students in distinguishing main issues from minor issues
9	The tutor uses his or her expert knowledge appropriately
10	The tutor contributes towards a better understanding of the subject matter
Factor 3	Commitment to the group's learning
11	The tutor gives an impression of being motivated
12	The tutor shows interest in our learning activities during the course
13	The tutor shows commitment with respect to group functioning

Students were asked to indicate on a three-point scale whether their tutor demonstrated the behavior described in each statement: (1) insufficiently, (2) neutral, or (3) sufficiently. Besides, a "not applicable" response option was added. This option could be selected if students themselves initiated the activity described by the statement and tutor intervention in this respect was not necessary. In addition, students were asked to give an overall judgement (ranging from 1-10, 6 was "sufficient") of the performance of the tutor.

Statistical analysis. For each item, mean scores, standard deviation and the number of missing cases, were computed. The percentage of missing cases for each item were lower than 10 percent. As already mentioned above, the data were aggregated at the tutorial group level.

A confirmatory factor analysis was carried out to assess the adequacy of the tutor performance model outlined above. In the confirmatory factor model, as specified in this study, all common factors were correlated, observed variables 1 through 6 were affected by the first common factor, observed variables 7 to 10 were affected by the second common factor, variables 11, 12 and 13 by the third common factor. Furthermore, all observed variables were assumed to be affected by a unique factor (error in each

variable), and no pairs of unique factors were correlated. The Lisrel VII program (Jöreskog & Sörbom, 1990) was used to determine whether the data confirmed this model. In addition, a one-factor model was tested.

Results

The number of students rating each tutor varied between 3 and 11. The mean score on each item for the 142 tutorial groups varied between 2.18 (*SD*=.56) and 2.82 (*SD*=.29), shown in Table 2. The low mean score for item 6 may reflect tutor's behavior of failing to provide adequate feedback. A study of Wilkerson (1992) identifying important skills for the problem-based tutor also showed tutors' failure to provide adequate feedback.

The coefficient alpha for the tutor evaluation questionnaire was .95 (13 items). The coefficient alpha for factor 1 was .90 (6 items), for factor 2 .94 (4 items), and for factor 3 .96 (3 items). These results, in general, demonstrate that the questionnaire is reliable.

The validity of the questionnaire was estimated by means of confirmatory factor analyses, as already described above. The correlations between the observed variables were the input for the confirmatory factor analyses (Lisrel VII). The correlation coefficients between the observed variables varied between .53 and .90 (N=142), except for item 6 'the tutor stimulates the evaluation of the group process' which revealed correlations varying between 0.16 and 0.57. The correlations between the common factors varied between .67 and .87. Although these coefficients are relatively high, a one-factor model did not fit the data. A model is assumed to fit the data if three conditions are met: (1) The chi-square divided by the degrees of freedom should be lower than 2, a p-value that differs from zero; (2) the root mean square residual should be lower than .07; and (3) the goodness-of-fit-index and the adjusted goodness-of-fit-index, which takes into account the number of degrees of freedom, should be higher than .80 (Saris & Stronkhorst, 1984). A three-factor model in which observed variables 1 through 6 are affected by the first common factor, observed variables 7 to 10 are affected by the second common factor, and variables 11, 12 and 13 are affected by the third common factor differed significantly from the data (chi-square [62 *df*] = 160.34, p = .000). The root mean square residual was .057, the goodness-of-fit index was .846, and the adjusted goodness-of-fit index was .774. These results suggest that the first condition specified by Saris and Stronkhorst (1984) is not satisfied, whereas both other conditions are satisfied. Skipping item 2 and item 6 resulted in a three-factor solution that did not differ significantly from the data (chi-square [41 *df*] = 74.91, p = .001). The root mean square residual of .024, goodness-of-fit index of .906, and adjusted goodness-of-fit index of .849, suggest that the model fits the data. A four-factor model in which the fourth factor consists of item 2 and 6, however, did not fit the data. In order to further cross-validate the proposed models, the data-set was split up in two sets. Set 1 consisted of the first courses in study year 1 to 4 and set 2 consisted of the second courses in study year 1 to 4.

With regard to the first set, both models did not fit the data; i.e., the first condition specified by Saris and Stronkhorst (1984) was not satisfied, whereas both other conditions are satisfied. With respect to the second set, on the other hand, both models fitted the data. The 11-item model produced the following results: chi-square [41 *df*] = 56.37, $p = .055$, a root mean square residual of .031, a goodness-of-fit index of .871, and adjusted goodness-of-fit index of .793. The 13-item model produced the following results: chi-square [62 *df*] = 104.57, $p = .001$), a root mean square residual of .050, goodness-of-fit index of .812, and adjusted goodness-of-fit index of .724. Since all three conditions were satisfied, both models show a reasonable fit with the data.

Table 2 The observed variables (mean (scale 1-3), standard deviation (SD), N=142) and their common factors

		Mean	SD
Factor 1	Guiding students through the learning process		
1	The tutor demonstrates to be well-informed about the process of problem-based learning	2.82	.29
2	The tutor stimulates all students to participate actively in the tutorial group process	2.39	.48
3	The tutor stimulates a careful analysis of the problems	2.62	.42
4	The tutor stimulates the generation of specific learning issues useful for self-study	2.57	.36
5	The tutor stimulates an extensive reporting on information collected during self-study	2.50	.42
6	The tutor stimulates evaluation of the tutorial group process	2.18	.56
Factor 2	Content knowledge input		
7	The tutor has an understanding of the subject matter covered in the course	2.70	.39
8	The tutor assists students in distinguishing main issues from minor issues	2.60	.38
9	The tutor uses his or her expert knowledge appropriately	2.60	.46
10	The tutor contributes towards a better understanding of the subject matter	2.57	.45
Factor 3	Commitment to the group's learning		
11	The tutor gives an impression of being motivated	2.67	.43
12	The tutor shows interest in our learning activities during the course	2.56	.43
13	The tutor shows commitment with respect to group functioning	2.67	.42

The overall judgement about the functioning of the tutor (a score ranging from 1-10, 6 was "sufficient") correlates highly with all 13 items (all were above .73, $p < .001$), with the exception of item 6 "the tutor stimulates evaluation of the group process" which correlated .40 ($p < .001$).

Since it is assumed that tutor functioning is a key-aspect of a problem-based curriculum, a relationship between a tutor's functioning during a course and student achievement on a corresponding end-of-course examination would be expected. The correlation coefficient between the average score on the 13 items of the tutor evaluation questionnaire and the average percentage of test items correctly answered, however, was very low .08 (n.s., N=142). For separate courses, this correlation coefficient varied between -.19 (n.s., n=18) and .46 (n.s., n=18).

Discussion

The purpose of this study was to investigate the validity of the tutor evaluation questionnaire. The results of the confirmatory factor analyses showed that a three-factor-model comprising 13 items does not fit the data, since one out of three conditions was not satisfied. Applying the same model to a smaller data set, however, revealed that the model fitted the data. Although skipping item 2 and 6, respectively "the tutor stimulates all students to participate actively in the tutorial group process" and "the tutor stimulates evaluation of the tutorial group process" improved the model's fit, the 13 item model was nevertheless assumed to be the best model. Item 2 and 6 reflect important skills a tutor should possess. Skipping these items would be detrimental to the content validity of the tutor evaluation questionnaire.

Alpha coefficients indicated that the reliability of the questionnaire was high. Nevertheless, further analyses should be conducted to assess the instrument's reliability. The different sources of error caused by students and items and the reproducibility of the scores as a function of the number of student responses should be estimated.

An overall judgement about the functioning of the tutor (expressed as a score from 1-10) correlated highly with all 13 items, with the exception of item 6 "the tutor stimulates evaluation of the group process". These findings indicate that students' global and analytical ratings about the functioning of their tutor highly correspond with each other.

Since it is assumed that tutor functioning is a key-aspect of a problem-based curriculum, the relationship between a tutor's functioning during a course and student achievement on a corresponding end-of-course examination was assessed. The correlation coefficient between the average score on the 13 items of the tutor evaluation questionnaire and the average percentage of test items correctly answered was low. A possible explanation for this low correlation could be that the tutor has no direct influence on achievement. Student achievement seems to be highly influenced by the extent to which the subject matter covered in a course links up with students' prior knowledge and time spent on self-study (Gijselaers & Schmidt, 1990).

Whereas, most tutors at the Maastricht Medical School highly appreciated the practical usefulness of the questionnaire, the implementation of the questionnaire showed several bottle-necks. Most problems encountered were related to the reliability of the data collected. For example, 16 tutors, 11 percent, were rated by less than six students out of a group of eight to eleven students. Taken into account that some tutors are rated by only a few students, caution is needed in interpreting the results. Especially, if student ratings of teaching effectiveness represent a source of information used in promotion, tenure and salary decisions. Another problem encountered when implementing the tutor evaluation questionnaire was that some students rated their tutor although they attended only a few tutorial group meetings. It was decided that the ratings of students who visited less than 80 percent of the tutorial group meetings were removed.

Implementing a tutor evaluation questionnaire requires an organizational context in which there are opportunities to communicate directly about the results. At the Maastricht Medical School the results of the tutor evaluation questionnaire are reported to individual tutors, the course-coordinator, and to the Educational Coordination Committee. The course-coordinator is responsible for the management of the educational design, implementation and evaluation of a course. If a tutor scores low, the course-coordinator discusses these results with the individual tutor. The results are also discussed during the meetings of the Educational Coordination Committee, responsible for appointing faculty members to educational roles. If a tutor scores low during three subsequent periods of tutoring, the chair of the department in which the tutor is working will be informed about the results. Such a tutor will be disqualified for several roles, such as the tutor role, the role of member of the planning group, the role of course-coordinator. In addition, this negative judgement may influence promotion and tenure decisions. The chair of a department will also be informed about consistently "good" functioning tutors. This judgement may positively influence promotion decisions. The direct communication and reward system should guarantee that the results are indeed used to improve the functioning of tutors. A review of the literature by Leviton & Hughes (1981) demonstrated five general factors influencing the use of evaluation results, such as the relevance of the results to the needs of the user, close and direct communication among evaluator and consumers, clear presentation of results, credibility of the results which is narrowly related to users' preconceptions about the quality of the results and the consensus between research findings and information from other sources. During pilot-studies, as described earlier, a direct communication between evaluator and consumers took place, in which the relevance of the results was discussed as well as the clarity of the presentation of the individual feedback each tutor receives.

Providing tutors with feedback about their functioning should of course imply that training programs are offered to tutors scoring very low, since improvement of tutor behavior is the ultimate purpose of collecting these data. At the Maastricht Medical Faculty all tutors are obliged to attend a tutor

training before they may guide a tutorial group. During two days, tutors are prepared for the tutor role and opportunities are offered to practice the required skills. However, remedial teaching activities providing practice and feedback should also be offered, since a one-time tutor training program will be not be sufficient for some tutors. These activities could, for example, include observing a good functioning tutor in a tutorial group. Besides, training should be offered in which tutors are observed in simulated groups and receive direct feedback about their functioning. Presently, initiatives are undertaken to realize these activities.

References

Barrows, H. S., (1988). *The tutorial process.* Illinois: Southern Illinois University School of Medicine.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education.* New York: Springer.

Davis, W. K., Nairn, R., Paine, M. E., Anderson, R. M., & Oh, M. S. (1992, April). *Must Small-Group Facilitators Be Content Experts?* Paper presented at the American Educational Research Association, San Francisco, CA.

De Volder, M. L. (1982). Discussion groups and their tutors: Relationship between tutor characteristics and tutor functioning. *Higher Education, 11,* 269-272.

Feletti, G. I., Doyle, E., Petrovic A., & Sanson-Fisher, R. (1982). Medical students' evaluation of tutors in a group-learning curriculum. *Medical Education, 16,* 319-325.

Gijselaers, W. H., & Schmidt, H. G. (1990). Development and evaluation of a causal model of problem-based learning. In: A. M. Nooman, H. G. Schmidt & E. S. Ezzat (Eds.). *Innovation in Medical Education: An evaluation of its present status* (pp. 95-113). New York: Springer.

Jöreskog, K. G., & Sörbom, D. (1990). *Lisrel VII. User's Guide.* Chicago: National Educational Resources.

Leviton , L.C., & Hughes, E.F.X. (1981). Research on the utilization of evaluations. A Review and Synthesis. *Evaluation Review, 5, 4,* 525-548.

Moust, J. C. & Schmidt H. G., (1992, April). *Undergraduate students as tutors: Are they as effective as faculty in conducting small-group tutorials?* Paper presented at the American Educational Research Association, San Francisco, CA.

Saris, W. & Stronkhorst, H. (1984). *Causal modelling in nonexperimental research.* Amsterdam, The Netherlands: Sociometric Research Foundation.

Wilkerson, L. A., (1992, April). *Identification of skills for the problem-based tutor: Student and faculty perspectives.* Paper presented at the American Educational Research Association, San Francisco, CA.