

Tutor intervention profile: reliability and validity

Willem S De Grave, Diana H J M Dolmans & Cees P M van der Vleuten

Department of Educational Development and Research, University of Maastricht, The Netherlands

SUMMARY

An instrument has been developed to provide tutors with feedback about their performance in problem-based tutorial groups. This questionnaire consists of 33 statements reflecting the behaviour of the tutor in the tutorial group when students analyse the case and generate learning issues and the behaviour of the tutor when students report back to the tutorial group about their self-study. The statements are grouped on theoretical dimensions found in the literature about co-operative learning and problem-based learning. This study focuses on the validity and the reliability of this instrument. Confirmatory factor analysis showed that a four-factor model showed a weak fit, whereas each separate factor fitted the data reasonably well. The four factors are: (1) elaboration; (2) directing the learning process; (3) integration of knowledge and (4) stimulating interaction and individual accountability. These factors were strongly correlated with the overall judgement of the tutor at the end of the unit, so the concurrent validity of this instrument was high. Generalizability studies indicated that the rating scales provide reliable information with a sample size of student responses falling within the range of real tutorial group sizes. The results of this study show that the instrument for providing tutors with feedback is valid and reliable. This questionnaire can be used in staff development activities and also provides needs assessment for faculty development.

Keywords

Education, medical, undergraduate, *methods; *problem-based learning; *programme evaluation; *teaching

INTRODUCTION

The tutor role is perceived as being important in problem-based learning (PBL). In causal modelling approaches to learning in PBL (Gijsselaers & Schmidt 1990) a tutor's behaviour is one of three factors affecting the way in which small-group tutorials function

(the other two being the students' prior knowledge and the quality of the problems handled).

In a problem-based curriculum the role of the tutor is different from that of teachers in a more traditional training programme. Because the required tutor activities in PBL are unfamiliar to conventional teachers, there is often confusion about the tutor's role (Rangachari & Crankshaw 1996). When institutions implement PBL there is, as a consequence, a need for faculty development. These faculty development programmes for PBL have been described and analysed by Nayer (1995) and Irby (1996). Nayer concluded that few faculty development programmes have conducted comprehensive evaluations and many programmes lack clear objectives and criteria to define good teaching and tutoring. In order to ensure the effectiveness of faculty development programmes, they should be based on a theory of effective tutoring and there should be instruments to give tutors feedback about their performance. Dolmans *et al.* (1994, p. 550) suggest that such an instrument should be based on the tasks set for the tutor at the school in which the instrument will be used, as well as on theoretical conceptions about the tutor role.

Research on effective tutor behaviour has focused particularly on the expertise of the tutor and its effects on student achievement (for a summary see Schmidt 1994). Only a few studies have identified important dimensions of tutor performance which stimulate student learning (Wilkerson 1992; Moust 1993; Dolmans *et al.* 1994; Schmidt & Moust 1995). Moust (1993) and Schmidt & Moust (1995) emphasize social congruence and cognitive congruence as important factors stimulating student learning. Cognitive congruence assumes sensitivity of the tutor concerning the difficulties experienced by students with a problem or the subject-matter covered with the problem. It is theorized that both subject-matter expertise and interpersonal qualities (social congruence) are necessary conditions for cognitive congruence to occur. Schmidt & Moust (1995) found that a tutor's level of expertise and social congruence is not only directly affected by the level of cognitive congruence but is also affected by other

elements. The level of social congruence influenced group functioning in a direct fashion, while the use of content expertise had a slightly negative effect on the students' level of self-study time and a slightly positive effect on level of achievement. As hypothesized, the level of cognitive congruence influenced tutorial-group functioning, which subsequently affected self-study time and intrinsic interest. Finally, time spent on self-study influenced level of achievement. Wilkerson (1992) identified two factors: maintaining positive interactions within the group and providing assistance in getting the work of the group accomplished. Dolmans *et al.* (1994) found three factors: guiding students through the learning process, content knowledge input and commitment to the group's learning.

An interesting phenomenon in research on effective tutor behaviour is that no references are made to the literature about co-operative learning and the different theories that explain why and under which conditions co-operative learning is effective. Slavin (1996) and Webb & Palincsar (1996) have summarized the theoretical base of co-operative learning. Next to cognitive-based theoretical approaches to co-operative learning there are motivational approaches such as social cohesion, group goals and individual accountability. The social cohesion approach holds that the effects of co-operative learning on achievement are strongly mediated by the cohesiveness of the group. An important characteristic of this approach is an emphasis on team-building activities and feedback or group self-evaluation during and after group activities. The group goals and individual accountability approach of co-operative learning holds that 'use of group goals or group rewards enhances the achievement outcomes of co-operative learning if and only if the group rewards are based on the individual learning of all members' (Slavin 1996, p. 45). The importance of group goals and individual accountability lies in providing students with an incentive to help each other and to encourage each other to put forth maximum effort. These two additional approaches might provide dimensions of tutor performance. In particular, the group goals and individual accountability approach is missing in research on effective tutor performance. Slavin (1996) has shown substantial empirical support for this factor.

Furthermore, these questionnaires should be based on theoretical notions about the process of problem-based learning (Schmidt 1993). Schmidt (1993) emphasized the cognitive process of elaboration during the process of problem-based learning, both before and after self-study. Elaboration refers to the cognitive processes of linking and integrating new information with prior knowledge. It emphasizes the constructive

nature of the learning process in problem-based learning. The positive effects of elaboration (especially self-generated elaborations) on learning are summarized by Hamilton (1989, p. 2): 'Elaborations may increase the redundancy of stored information, impose an organizational scheme on stored information, increase the number of contextual elements that will overlap between encoding context and/or increase the distinctiveness of stored information.' De Grave *et al.* (1996) analysed the cognitive and metacognitive processes during problem analysis (constructing explanations) and they also emphasized the cognitive process of elaboration. The tutor can stimulate the generation of elaboration in the tutorial group both before and after self-study.

Although the studies described above contributed towards better insight into effective tutor behaviour, instruments to give tutors feedback about their performance are only poorly based on these research findings. The instruments developed merely focus upon very global tutor interventions (Wilkerson 1992; Moust 1993; Dolmans *et al.* 1994; Schmidt & Moust 1995). For example, the concept of cognitive congruence, as outlined before, is hardly operationalized in specific tutor-interventions. In addition, the instruments developed do not contain specific items for the discussion preceding the generation of learning issues and for the reporting of the learning issues, whereas different tutor interventions are required for both phases of the learning process.

These considerations have led to the construction of a questionnaire with four theoretically relevant dimensions of tutor performance. Two dimensions deal with tutor interventions in the tutorial group before students begin studying, as follows: (1) stimulating elaboration and (2) directing the learning process (according to the learning issues planned by the planning-group of the unit). The two dimensions for tutor interventions during the report phase are: (3) stimulating integration and (4) stimulating interaction and individual accountability. The dimensions of elaboration and integration might be considered as operationalization of cognitive congruence. Stimulating interaction and individual accountability refer to the theoretical dimensions of social cohesion and group goals/individual accountability. The questionnaire provides an assessment of the profile of possible tutor interventions. It has been named the Tutor Intervention Profile (TIP).

In this study aspects of reliability and validity of this questionnaire are investigated. The construct validity will be assessed by testing the theoretical dimensions to empirical data in a confirmatory factor analysis. The concurrent validity of the instrument is investigated by

the correlation of the dimensions with the overall quality rating by the student of the tutor. Furthermore, generalizability studies are undertaken to estimate the reliability of the instrument. Potential sources of error variance were analysed and used to estimate the number of required student responses to obtain reliable information for a single tutor.

METHOD

Subjects

The study participants consisted of 67 tutors, who ran 67 tutorial groups in three units in the academic year 1996-97 at the medical school of the University of Maastricht; one first-year unit and two second-year units. The number of students involved in each tutorial group varied between eight and 10.

Instruments

Studies investigating effective tutor behaviour were consulted. Based on the findings of these studies four dimensions were considered essential: (1) stimulating elaboration, (2) directing the learning process, (3) stimulating integration and (4) stimulating interaction and individual accountability. In a pilot study, 20 well-performing tutors and 20 randomly selected students were interviewed in order to find out which specific behaviours an effective tutor is assumed to demonstrate for each of the four dimensions. Based on these interviews, a questionnaire was developed that was tested in a trial study. This trial study finally resulted in a list of 33 statements. In order to validate and determine the instrument's reliability, at the end of three units students were asked to rate the performance of their tutor by filling out the TIP. Students were asked to rate the behaviour of the tutor on a five-point Likert scale ranging from 1, 'strongly disagree' to 5, 'strongly agree'. A 'not applicable' response option was added which could be selected if, for instance, students themselves initiated the activity described by the statement and tutor intervention in this respect was not necessary. Four factors were assumed to underlie the 33 statements of the TIP. Nine statements were related to the degree to which a tutor stimulates elaboration, eight statements dealt with directing the learning process, eight statements with stimulating integration and eight statements with stimulating interaction and individual accountability. Table 1 contains two examples of items included in each factor. A full version of the questionnaire is available from the first author.

At the end of each unit, students routinely completed a programme-evaluation questionnaire. One of the

Table 1 Examples of items included in the TIP and the four factors

Factor 1. Stimulating elaboration
• Stimulates a more in-depth brainstorm by, for example, asking questions, asking for clarification and stimulating relations
• Stimulates identification of gaps in student's prior knowledge
Factor 2. Directing the learning process
• Stimulates generation of learning issues with sufficient depth and width
• Draws attention to students' gaps in prior knowledge while generating learning issues
Factor 3. Stimulating integration
• Stimulates students integration of new acquired knowledge with knowledge acquired with previous cases within the same unit
• Stimulates students to apply the knowledge gained during self-study to explain the phenomena described in the case
Factor 4. Stimulating interaction and individual accountability
• Stimulates students to make an inventory of the learning resources consulted during self-study
• Stimulates students to report in their own words rather than reading from notes or photocopies

items included in this questionnaire asked students to give an overall judgement (ranging from 1 to 10; 6 was 'sufficient') of the performance of their tutor. The relationship between this overall judgement of a tutor and the scores on the TIP provided information about the concurrent validity of the TIP.

Statistical analysis

To achieve a fully balanced design convenient for statistical analysis, a random sample of five students was selected from the total number of students rating that tutor. If a tutor was judged by less than five students, that tutor was excluded from the analysis. Initially 67 tutors were involved. In the dataset used for the validity and generalizability analysis 61 tutors were included (six tutors were excluded due to balancing). In total, 61 tutors were each rated by five students. Thus, 305 student judgements (61×5) were available. Both students and tutors were randomly assigned to the tutorial groups.

A confirmatory factor analysis was carried out to assess the adequacy of the theoretical tutor performance model outlined above (construct validity). For the confirmatory factor analysis, data were aggregated at the tutorial group level by computing average scores across students for each tutor. Missing scores were replaced by the average tutorial's group score. In total 61 tutors or cases were involved in the confirmatory factor

Table 2 Common factors, number of statements (*N* items) mean scores (scale 1-5) and standard deviations (SD), *N* = 61

	<i>N</i> items	Mean	SD
Factor 1 Stimulating elaboration	9	3.3	0.5
Factor 2 Directing the learning process	8	3.5	0.4
Factor 3 Stimulating integration	8	3.5	0.5
Factor 4 Stimulating interaction and individual accountability	8	3.0	0.4

analysis. In the confirmatory factor model, as specified in this study, all common factors were correlated, observed variables 1-9 were affected by the first common factor, observed variables 10-17 were affected by the second common factor, observed variables 18-25 were affected by the third common factor and observed variables 26-33 were affected by the fourth common factor. Table 2 contains the mean scores and standard deviations for each 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 AMOS 3.6 program (Arbuckle 1997) was used to determine whether the data confirmed this model.

To test the TIP's concurrent validity, for each tutor an average score for each factor was computed which was subsequently correlated with the overall judgement of the tutor on the programme evaluation questionnaire.

Generalizability studies were conducted to estimate the reliability of the four factors (Brennan & Kane 1979; Crick & Brennan 1983). The analyses were conducted at the level of individual students. In total 61 tutors were involved; each were judged by five students. The variance components that were included in these analyses were: (1) difference in tutors (T) (object of measurement), (2) differences in items (I), (3) differ-

Table 3 Results of confirmatory factor analysis for the four factors and for each separate factor: chi-square (χ^2), degrees of freedom (d.f.), *P* value, root mean square residual (RMR), goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI)

Model	χ^2	d.f.	<i>P</i> value	RMR	GFI	AGFI
Four factors	788.65	489	0.00	0.03	0.58	0.52
Factor 1	34.17	27	0.16	0.01	0.89	0.82
Factor 2	22.99	20	0.29	0.02	0.91	0.84
Factor 3	30.60	20	0.06	0.02	0.89	0.80
Factor 4	23.14	20	0.28	0.02	0.91	0.84

ences between students nested within tutors (S:T), (4) interaction between tutors and items (TI) and (5) interaction between items and students nested within tutors and general error (IS:T, e) (Shavelson & Webb 1991). Although it is disputable that items represent a fixed facet, an all-random model was used which will give the most conservative reliability estimations.

RESULTS

Validity studies

The correlation coefficients between the observed variables varied between 20 and 81 (*N* = 61). The correlation between common factors 1 and 2 was 0.92; between 1 and 3, 0.95; between 1 and 4, 0.82; between 2 and 3, 0.98; between 2 and 4, 0.89; and between 3 and 4, 0.87. Thus, all four factors were highly correlated.

A model is assumed to fit the data if three conditions are met: (1) the χ^2 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 0.07; 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 0.80 (Saris & Stronkhorst 1984).

The results of the four-factor model described above showed that only the second condition specified by Saris & Stronkhorst was satisfied (χ^2 , [d.f. = 489] = 788.65, *P* = 0.00, a root mean square residual of 0.03, a goodness-of-fit index of 0.58, and adjusted goodness-of-fit index of 0.52). In addition, each factor was tested separately in order to establish whether the items represent the underlying concept of each factor. The results are presented in Table 3. As can be seen in this table, for each separate factor, all conditions specified by Saris & Stronkhorst (1984) are satisfied. In general the results of the confirmatory factor analyses indicate that each separate factor seems to fit the data reasonably well, whereas the four-factor model shows a weak fit.

The relationship between the overall judgement of a tutor on the programme evaluation questionnaire and the scores on the TIP also provide information about the concurrent validity of the questionnaire. Overall judgement correlates highly with each factor. The correlations for each factor are, respectively, 0.67, 0.60, 0.70 and 0.51 (*P* < 0.001).

Generalizability studies

In Table 4, the sources of variability and the corresponding estimated variance components are summarized

Table 4 Estimated variance components of the four factors (scale 1-5), corresponding degrees of freedom (d.f.), standard error and variance components expressed as the percentage of total variance

Source	d.f.	Estimated variance component	Standard error	Percentage of total variance
<i>Factor 1</i>				
Tutors (T)	60	0.1957	0.0399	18.6
Items (I)	8	0.0003	0.0022	0.0
Students within tutors (S:T)	244	0.0000	0.0058	0.0
Tutors by items (TI)	480	0.1306	0.0183	12.4
IS:T, e	1952	0.7248	0.0232	68.9
<i>Factor 2</i>				
Tutors (T)	60	0.1736	0.0362	16.5
Items (I)	7	0.0000	0.0018	0.0
Students within tutors (S:T)	244	0.0000	0.0078	0.0
Tutors by items (TI)	420	0.0959	0.0182	9.1
IS:T, e	1708	0.7841	0.0268	74.4
<i>Factor 3</i>				
Tutors (T)	60	0.2167	0.0451	19.9
Items (I)	7	0.0071	0.0058	0.7
Students within tutors (S:T)	244	0.0000	0.0057	0.0
Tutors by items (TI)	420	0.1841	0.0226	16.9
IS:T, e	1708	0.6833	0.0234	62.6
<i>Factor 4</i>				
Tutors (T)	60	0.1389	0.0302	10.9
Items (I)	7	0.0032	0.0038	0.3
Students within tutors (S:T)	244	0.0000	0.0094	0.0
Tutors by items (TI)	420	0.0798	0.0212	6.3
IS:T, e	1708	1.0475	0.0358	82.5

for the four factors. The percentage of true or score variance associated with tutors for factor 1 was 18.6, for factor 2 16.5, for factor 3, 19.9 and for factor 4 10.9 of the total variance. These percentages represent the true score variance or the variance of interest and apparently, based on the TIP, it is possible to an extent to discriminate tutor behaviour.

The estimated variance components presented in Table 4 were used to estimate reliability indices. The

reliability coefficient indicates how many students are required to obtain a generalizability coefficient of a given magnitude. For factors 1-4 these reliability coefficients were 0.86, 0.84, 0.84 and 0.78, respectively, with five student judgements. The standard error of measurement (SEM) also provides relevant information with regard to the reliability of the tutor rating scale. The SEM can be used to estimate confidence intervals for individual scores. For example, the 95%

Table 5 Generalizability coefficients (*G*) and standard errors of measurement (SEM), as a function of the number of student ratings for the four factors (scale 1-5)

Student responses	Factor 1		Factor 2		Factor 3		Factor 4	
	G	SEM	G	SEM	G	SEM	G	SEM
2	0.78	0.23	0.74	0.25	0.76	0.26	0.63	0.28
3	0.82	0.20	0.79	0.21	0.80	0.23	0.71	0.23
4	0.85	0.19	0.82	0.19	0.82	0.21	0.75	0.21
5	0.86	0.18	0.84	0.18	0.84	0.20	0.78	0.19
10	0.89	0.15	0.89	0.15	0.87	0.18	0.84	0.15
15	0.91	0.14	0.90	0.14	0.88	0.17	0.87	0.14

confidence interval of a score can be estimated by multiplying the SEM by 1.96 (Ferguson 1981). Taking the arbitrary standpoint that a difference of at least 1 point is required on a five-point scale to obtain reliable results, the SEM should be lower than or equal to 0.26 (0.5/1.96) at the level of 95% confidence. In Table 4 the reliability coefficients of the four factors are presented as a function of the number of student responses and the standard error of measurement (SEM). Taking into account the practical significance level of 0.26 and a generalizability coefficient of 0.80, Table 5 demonstrates that only *two* student responses are required to obtain reliable results for factors 1 and 3, *three* students are required to obtain reliable results for factor 2 and *five* student responses are required to obtain reliable results for factor 4. In order to construct a profile of the tutor's performance for the four factors included in the TIP, the responses of at least *five* students are required.

CONCLUSION

The purpose of this study was to investigate the validity and reliability of the TIP. The results of the confirmatory factor analyses indicated that a four-factor model comprising 33 items fits the data only weakly, whereas each separate factor indicates a reasonable fit, since all three statistical conditions specified by Saris & Stronkhorst (1984) were at least satisfied. However, all four factors are highly correlated with each other. This suggests that the factor scores provide limited unique information across the four areas. The high correlations between each factor and the overall judgement indicated that the concurrent validity of the TIP was good.

Generalizability analyses indicated that the tutor evaluation questionnaire is a reliable instrument. The reliability coefficients for each factor varied between 0.86 and 0.78, based on five student ratings. The number of students required to obtain reliable results were, for factors 1-4, *two* students, *three* students, *two* students and *five* students, respectively. This implies that tutor evaluations with the TIP provide reliable information in most problem-based settings where group sizes as used in this study are regular.

This questionnaire provides specific feedback for tutors, both on an item level as well as on a factor level. Tutors can compare their own profile with those of other tutors. This questionnaire has, mainly, a formative function. Weimer (1990) emphasizes the use of such questionnaires for making teachers aware of and reflect on their approach to teaching. Questionnaires with global items are more suitable for summative functions: 'The summative assessment provides the overview, the global perspective on in-

structional effectiveness. Formative activities focus on the details.'

The questionnaire may be used during initial tutor training, where it both serves as an observation-checklist as well as providing a standard for what is expected of this new role for teachers. It may also be used for new tutors who acquire feedback on their performance given by students and expert tutors who observe the interventions of the tutor. Experienced tutors use this instrument for self-evaluation and also use the feedback of students to discuss their approach of the tutor role with other tutors during in-service tutor training. Evaluation studies done with this instrument may provide needs assessment for faculty development activities.

REFERENCES

- Arbuckle J L (1997) *Amos Users' Guide Version 3.6*. SmallWaters Corporation, Chicago.
- Brennan R L & Kane M T (1979) Generalizability theory: a review. *New Directions for Testing and Measurement* 4, 33-51.
- Crick J E & Brennan R L (1983) *Manual for Genova: A Generalized Analysis of Variance System*. American College Testing Program, Iowa.
- De Grave W S, Boshuizen H P A & Schmidt H G (1996) Problem-based learning: cognitive and metacognitive processes during problem analysis. *Instructional Science* 24, 321-41.
- Dolmans D H J M, Wolfhagen H A P, Schmidt H G, van der Vleuten & C P M (1994) A rating scale for tutor evaluation in a problem-based curriculum: validity and reliability. *Medical Education* 28, 550-8.
- Ferguson G A (1981) *Statistical Analysis in Psychology and Education*, 5th edn. McGraw-Hill, Auckland.
- Gijselaers W H & Schmidt H G (1990) Development and evaluation of a causal model of problem-based learning. In: *Innovation in Medical Education: An Evaluation of Its Present Status* (eds A M Nooman, H G Schmidt & E S Ezzat), pp. 95-113. Springer, New York.
- Hamilton R (1989) The effects of learner-generated elaborations on concept learning from prose. *Journal of Experimental Education*, 57, 205-216.
- Irby D M (1996) Models of faculty development for problem-based learning. *Advances in Health Sciences Education* 1, 69-81.
- Moust J C (1993) *De rol van tutoren in probleemgestuurd onderwijs. Contrasten tussen student- en docenttutoren [The role of tutors in problem-based learning. Contrasts between student- and staff-tutors]*. Thesis, University of Limburg, Maastricht, The Netherlands.
- Nayer M (1995) Faculty development for problem-based learning programs. *Teaching and Learning in Medicine* 7, 138-48.
- Rangachari P K & Crankshaw D J (1996) Beyond facilitation: the active tutor in a problem-based course. *Biochemical Education* 24, 192-5.
- Saris W & Stronkhorst H (1984) *Causal Modelling in Nonexperimental Research*. Sociometric Research Foundation, Amsterdam, The Netherlands.

- Schmidt H G (1993) Foundations of problem-based learning: some explanatory notes. *Medical Education* 27, 422-32.
- Schmidt H G (1994) Resolving inconsistencies in tutor expertise research: does lack of structure cause students to seek tutor guidance. *Academic Medicine* 69, 656-62.
- Schmidt H G & Moust J H C (1995) What makes a tutor effective? A structural-equations modeling approach to learning in problem-based curricula. *Academic Medicine* 70, 708-14.
- Shavelson R J & Webb N M (1991) *Generalizability Theory. A Primer*. Sage, London.
- Slavin R E (1996) Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology* 21, 43-69.
- Webb N M & Palincsar S A (1996) Group processes in the classroom. In: *Handbook of Educational Psychology* (eds D C Berliner & R C Calfee), pp. 841-73. Simon & Schuster Macmillan, New York.
- Weimer M (1990) *Improving College Teaching*. Jossey Bass, San Francisco.
- Wilkerson L A (1992) *Identification of Skills for the Problem-based Tutor: Student and Faculty Perspectives*. Paper presented at the American Educational Research Association, San Francisco, CA.

Received 21 July 1997; accepted for publication 20 January 1998